

23-356

**IN THE UNITED STATES COURT OF APPEALS
FOR THE SECOND CIRCUIT**

EUGENE VOLOKH, LOCALS TECHNOLOGY INC.,
RUMBLE CANADA INC.,

PLAINTIFFS-APPELLEES,

v.

LETITIA JAMES, IN HER OFFICIAL CAPACITY
AS ATTORNEY GENERAL OF NEW YORK,

DEFENDANT-APPELLANT.

On Appeal from the United States District Court
for the Southern District of New York
Case No. 1: 22-cv-10195

The Honorable Andrew L. Carter, United States District Court Judge

**BRIEF OF *AMICI CURIAE* ELECTRONIC FRONTIER FOUNDATION
AND AMERICAN CIVIL LIBERTIES UNION IN SUPPORT OF
PLAINTIFFS-APPELLEES AND AFFIRMANCE**

BRIAN HAUSS
AMERICAN CIVIL LIBERTIES UNION
FOUNDATION
125 Broad St., 18th Fl.
New York, NY 10004
bhauss@aclu.org
(212) 549-2500

DAVID GREENE
ELECTRONIC FRONTIER FOUNDATION
815 Eddy Street
San Francisco, CA 94109
davidg@eff.org
(415) 436-9333

Counsel for Amici Curiae

CORPORATE DISCLOSURE STATEMENT

Pursuant to Rule 26.1 of the Federal Rules of Appellate Procedure, *Amici Curiae* Electronic Frontier Foundation (EFF) and American Civil Liberties Union (ACLU) state that they do not have a parent corporation and that no publicly held corporation owns 10% or more of their stock.

Dated: September 26, 2023

By: /s/ David Greene
David Greene

TABLE OF CONTENTS

CORPORATE DISCLOSURE STATEMENT	i
TABLE OF AUTHORITIES.....	iv
STATEMENT OF INTEREST	1
INTRODUCTION.....	3
ARGUMENT	3
I. The Content Moderation Systems Targeted by the Law Are Editorial Processes Protected by the First Amendment.	3
A. Content Moderation Is an Inherently Expressive Editorial Process.....	4
1. Content Moderation Is an Historic and Widely Employed Practice.	4
2. Social Media Platforms Have Rules, Standards and Guidelines about What Content They Want and Don't Want on Their Sites.	7
3. Content Moderation Has Long Been and Remains a Fraught and Controversial Process.....	8
B. Content Moderation is Protected by the First Amendment.	14
1. The First Amendment protects the right to speak by curating the speech of others.....	14
2. Social media and other interactive websites have a constitutional right to adopt, define, and implement their own editorial policies.....	17
II. Government Involvement in Content Moderation Raises Serious First Amendment Concerns.	21
III. The New York Law Unconstitutionally Coerces Interactive Websites to Adopt the State's Hate Speech Philosophy at the Expense of Their Own Editorial Policies.....	22

CONCLUSION25
CERTIFICATE OF COMPLIANCE27
CERTIFICATE OF SERVICE.....28

TABLE OF AUTHORITIES

Cases

303 Creative LLC v. Elenis,
600 U.S. ___, 143 S. Ct. 2298 (2023)19

Amer. Freedom Defense Initiative v. Lynch,
217 F. Supp. 3d 100 (D.D.C. 2016)18

Application of Consumers Union of U. S., Inc.,
495 F. Supp. 582 (S.D.N.Y. 1980)15

Backpage.com v. Dart,
807 F.3d 229 (7th Cir. 2015).....23, 24

Bantam Books, Inc. v Sullivan.
372 U.S. 58 (1963)22, 23

Burse v. United States,
466 F.2d 1059 (9th Cir. 1972).....14

Children's Health Def. v. Facebook Inc.,
546 F. Supp. 3d 909 (N.D. Cal. 2021).....18

Herbert v. Lando,
441 U.S. 153 (1974)15

Huber v. Biden,
No. 21-CV-06580-EMC, 2022 WL 827248
(N.D. Cal. Mar. 18, 2022)18

Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Boston,
515 U.S. 557 (1995)18

In re Consumers Union of the U.S., Inc.,
32 Fed. R. Serv. 2d 1373 (S.D.N.Y. 1981)15

*Janus v. American Federation of State, County, & Municipal Employees,
Council 31*,
138 S. Ct. 2448 (2018)17

Kenneally v. Suzuki Motor Co., Ltd.,
 No. M-8-85, 1994 WL 48840 (S.D.N.Y. Feb. 10, 1994)15

Kennedy v. Warren,
 66 F.4th 1199 (9th Cir. 2023).....24

La’Tiejira v. Facebook, Inc.,
 272 F. Supp. 3d 981 (S.D. Tex. 2017).....18

Langdon v. Google, Inc.,
 474 F. Supp. 2d 622 (D. Del. 2007)18

Los Angeles v. Preferred Communications, Inc.,
 476 U.S. 488 (1986)14

Manhattan Community Access Corp. v. Halleck,
 139 S. Ct. 1921 (2019)14

Masterpiece Cakeshop, Ltd. v. Colorado Civil Rights Comm’n,
 138 S. Ct. 1719 (2018)17

Miami Herald Publishing Co. v. Tornillo,
 418 U.S. 241 (1974)passim

Miller v. California,
 413 U.S. 15 (1973)7

Nat’l Inst. of Family & Life Advocates v. Becerra,
 138 S. Ct. 2361 (2018)17

National Rifle Ass’n v. Vullo,
 49 F.4th 700 (2d Cir. 2022).....24

NetChoice LLC v. Attorney General, Florida,
 34 F.4th 1196 (11th Cir. 2022).....18, 19

NetChoice, LLC v Paxton,
 49 F.4th 439 (5th Cir. 2022).....19, 20

New York Times Co. v. Sullivan,
 376 U.S. 254 (1964)16

O’Handley v. Padilla,
 No. 21-CV-07063- CRB, 2022 WL 93625
 (N.D. Cal. Jan. 10, 2022).....18

Okwedy v. Molinari,
 333 F.3d 339 (2d Cir. 2003).....24

Pacific Gas & Elec. Co. v. Pub. Utilities Comm’n of California,
 475 U.S. 1 (1986)18

Prager Univ. v. Google LLC,
 951 F.3d 991 (9th Cir. 2020).....19

PruneYard Shopping Center v. Robins,
 447 U.S. 74 (1980)19

Rumsfeld v. Forum for Academic & Institutional Rights, Inc.,
 547 U.S. 47 (2006)19

State of Missouri v. Biden,
 No. 23-30445, 2023 WL 5821788 (5th Cir., Sept. 8, 2023).....24

Stossel v. Meta Platforms, Inc.,
 634 F. Supp. 3d 743 (N.D. Cal. 2022).....18

Turner Broadcasting Sys., Inc. v. F.C.C.,
 512 U.S. 622 (1994)18

Woodhull Freedom Foundation v. United States,
 72 F.4th 1286 (D.C. Cir. 2023)10

Zhang v. Baidu.com Inc.,
 10 F. Supp. 3d 433 (S.D.N.Y. 2014);.....18

Statutes

N.Y. Gen. Bus. Law § 394-ccc.....3, 20, 22, 23

Other Authorities

A New Policy Against Self-Harm Blogs, Tumblr (Feb. 23, 2012)6

Adult Nudity and Sexual Activity, Facebook7

Article 19, *Sheikh Jarrah: Facebook and Twitter Silencing Protestors, Deleting Evidence* (May 10, 2021)12

Bennett Cyphers, Cory Doctorow, *The New ACCESS Act Is a Good Start. Here’s How to Make Sure It Delivers*, EFF21

Comment on Evaluating the Competitive Effects of Corporate Acquisitions and Mergers, EFF.....21

Community Guidelines, Instagram7

Community Standards Enforcement Report: Bullying and Harassment, Meta.....13

Community Standards Enforcement Report: Hate Speech, Meta13

Danielle Blunt et al., *Posting Into The Void*, Hacking//Hustling (2020)10

Defining the Public Interest, X Help Center6

Digital Well-being, TikTok.....6

Eric Goldman, *Content Moderation Remedies*, 28 Mich. Tech. L. Rev. 1 (2021)6, 8

Facebook Treats Punk Rockers Like Crazy Conspiracy Theorists, Kicks Them Offline, EFF.....11

Gettr – Terms of Use, Gettr (May 17, 2023).....8

Hannah Bloch-Wehba, *Automation in Moderation*, 53 Cornell Int’l L.J. 41 (2020) 4

Hannah Denham, *Another Fake Video of Pelosi Goes Viral on Facebook*, Wash. Post (Aug. 3, 2020).....7

Instagram Stands Against Online Bullying, Instagram6

Jack Shafer, *The Op-Ed Page’s Back Pages: A Press Scholar Explains How the New York Times Op-Ed Page Got Started*, Slate (Sept. 27, 2010)16

James Grimmelman, *The Virtues of Moderation*, 17 Yale J. of Law & Tech 42 (2015)5

Jillian C. York & David Greene, *How to Put COVID-19 Content Moderation Into Context*, Brookings’ TechStream (May 21, 2020)4

Jillian C. York, *Silicon Values: The Future of Free Speech Under Surveillance Capitalism* (Verso 2021)9

Kate Crawford & Tarleton Gillespie, *What Is a Flag For? Social Media Reporting Tools and the Vocabulary of Complaint*, 18 *New Media & Soc’y* 410 (2014)...13

Kate Klonick, *The New Governors: The People, Rules, And Processes Governing Online Speech*, 131 *Harv. L. Rev.* 1598 (2018)5, 7, 8

Kevin Anderson, *YouTube Suspends Egyptian Blog Activist’s Account*, *The Guardian* (Nov. 28, 2007).....9, 11

Malachy Browne, *YouTube Removes Videos Showing Atrocities in Syria*, *N.Y. Times* (Aug. 22, 2017) 11

Megan Farokhmanesh, *YouTube Is Still Restricting and Demonetizing LGBT Videos—and Adding Anti-LGBT Ads to Some*, *The Verge* (June 4, 2018) 11

Michael J. Socolow, *A Profitable Public Sphere: The Creation of the New York Times Op-Ed Page*, *Comm’n & Journalism Fac. Scholarship* (2010)..... 16

Mike Masnick, *Content Moderation At Scale Is Impossible: Recent Examples Of Misunderstanding Context*, *TechDirt* (Feb. 26, 2021) 10

Moderator Code of Conduct, *Reddit*..... 12

Op-Ed, *Wikipedia*..... 16

Pennysaver, *Wikipedia*..... 17

Policy: Terms of Use, *Wikimedia* 12

Promoting Hate Based on Identity or Vulnerability, *Reddit*.....8

Reddiquette, *Reddit* 13

Report Content on Facebook, *Facebook* 13

Role of Administrators and Moderators on Discord, *Discord* 12

Rules Enforcement, *Twitter* (July 28, 2022)..... 14

Ryan Mac, *Instagram Censored Posts About One of Islam’s Holiest Mosques, Drawing Employee Ire*, *BuzzFeed News* (May 12, 2021) 11

Santa Clara Principles21

Sensitive Media Policy, Twitter (March 2023)7

Seny Kamara et al., *Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems*, Ctr. for Democracy & Tech. (2021)5

Taylor Wofford, *Twitter Was Flagging Tweets Including the Word “Queer” as Potentially “Offensive Content”*, Mic (June 22, 2017).....12

Twitter Suspends Accounts of Palestinian Quds News Network, Al Jazeera (Nov. 2, 2019)12

Violent and Graphic Content, Meta Transparency Center.....6

YouTube Community Guidelines Enforcement, Google.....13

STATEMENT OF INTEREST¹

Amici curiae submit this brief to provide additional information and argument on two issues raised by this appeal that are within *amici*'s field of expertise: the nature of content moderation as practiced by social media and other interactive websites as constitutionally protected editorial and curatorial expression; and, the weighty First Amendment concerns raised by governmental interference with such curatorial expression. *Amici curiae* do so from the perspective of the Internet users who read and contribute to such sites, rather than the sites themselves as represented by plaintiffs.

The Electronic Frontier Foundation (EFF) is a member-supported, nonprofit civil liberties organization that has worked for over 30 years to protect free speech, privacy, security, and innovation in the digital world. EFF, with approximately 30,000 members, represents the interests of technology users in court cases and broader policy debates surrounding the application of law to the Internet and other technologies. EFF frequently files briefs in cases addressing online intermediary content moderation, and studies and writes extensively on the issue.

The American Civil Liberties Union (ACLU) is a nationwide, non-partisan,

¹ No counsel for a party authored this brief in whole or in part, and no person other than *amici* or their counsel has made any monetary contributions intended to fund the preparation or submission of this brief. The parties have consented to the filing of this brief.

non-profit organization. The organization is dedicated to defending the principles embodied in the Constitution and our nation's civil rights laws and, for over a century, has been at the forefront of efforts nationwide to protect the full array of civil rights and liberties, including freedom of speech and freedom of the press online. The ACLU has frequently appeared before courts throughout the country in First Amendment cases, both as direct counsel and as *amicus curiae*.

INTRODUCTION

Section 394-ccc is an unconstitutional intrusion into the editorial freedom that all publishers, including social media platforms and other websites, enjoy. The law's clear purpose is to change the editorial policies and editorial methods of these publishers to accord with the state's definition of hateful speech and what the state perceives to be its attendant harms. While the law purports to empower some internet users, it also aims to silence others and deny still others access to information.

Content moderation by the online intermediaries is an already fraught process, and government interjection of itself into that process in any form raises serious First Amendment, and broader human rights, concerns. Courts must generally scrutinize such interventions. Laws such as section 394-ccc that seek to coopt a portion of the process must survive First Amendment scrutiny. Section 394-ccc does not. This Court should affirm the district court's order enjoining the law's enforcement.

ARGUMENT

I. The Content Moderation Systems Targeted by the Law Are Editorial Processes Protected by the First Amendment.

As a threshold matter, this Court should join the other courts that recognize that the social media platforms and other websites subject to section 394-ccc have the First Amendment right to curate the speech of others that is published on their

sites, regardless of whether they curate a lot or a little, and regardless of whether their editorial philosophy is readily discernible or consistently applied.

A. Content Moderation Is an Inherently Expressive Editorial Process.

1. Content Moderation Is an Historic and Widely Employed Practice.

Social media platforms, at least from their point of mass adoption, have rarely published all legal speech submitted to their sites. Instead, they engage in content moderation: the use of policies, systems, and tools to decide what user-generated content or accounts to publish, remove, amplify, or manage.² Large-scale, outsourced content moderation first emerged in the early 2000s.³

Platforms practice content moderation in phases: they define permissible and impermissible content; detect content that may violate their policies or the law; evaluate that content to determine whether it in fact violates their policies or the law; take an enforcement action against violative content; allow users to appeal or otherwise seek review of content moderation decisions that they believe are erroneous; and educate users about content moderation policies and their

² See Hannah Bloch-Wehba, *Automation in Moderation*, 53 Cornell Int'l L.J. 41, 42, 48 (2020).

³ Jillian C. York & David Greene, *How to Put COVID-19 Content Moderation Into Context*, Brookings' TechStream (May 21, 2020), <https://www.brookings.edu/articles/how-to-put-covid-19-content-moderation-into-context/>.

enforcement.⁴ In each phase, platforms make editorial judgments about what content they wish to allow or forbid on their services, or how to display or arrange it.

For example, during the definitional phase, some platforms develop a content policy, *i.e.*, a set of rules about what content is and is not allowed on their platforms.⁵ Platforms may engage in significant internal discussion and debate, conduct internal and external research, and write multiple drafts before determining their content policies.⁶ Smaller platforms might not adopt formal policies at all—particularly, for example, platforms like blogs, where the small community size could make it manageable for a single blogger to moderate the platform via ad hoc decision-making.⁷ Many platforms publish their content policies, but others do not, in order to maintain flexibility and for other reasons.

⁴ Seny Kamara et al., *Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems*, Ctr. for Democracy & Tech. 9–11 (2021), <https://cdt.org/wp-content/uploads/2021/08/CDT-Outside-Looking-In-Approaches-to-Content-Moderation-in-End-to-End-Encrypted-Systems-updated-20220113.pdf>.

⁵ *Id.* at 9.

⁶ See Kate Klonick, *The New Governors: The People, Rules, And Processes Governing Online Speech*, 131 Harv. L. Rev. 1598, 1631-35 (2018).

⁷ See James Grimmelman, *The Virtues of Moderation*, 17 Yale J. of Law & Tech 42, 73 (2015) (“[T]he larger a community is, the better it is at competing with external alternatives, but the more internal moderation it requires As a community grows, it becomes easier for individuals and groups to resist a norm. This breakdown makes it harder to use social norms to moderate large communities. A group of twenty can operate by unspoken consensus in a way that a group of twenty thousand cannot.”).

Once a platform has decided, perhaps after deliberation and debate, during the evaluation phase that particular content violates its policies, the platform must decide what action to take in the enforcement phase. That is not a binary decision about whether to take down content or allow it to remain on a service, but also includes whether to change the manner or place in which content is displayed or to add the platform's own affirmative speech.⁸ For example, depending on the nature of the violative content, a platform might choose to add its own content warning, public service announcement, health or safety resources, or "fact-checking" information to provide context and support for its users.⁹

⁸ See Eric Goldman, *Content Moderation Remedies*, 28 Mich. Tech. L. Rev. 1, 23–39 (2021) (describing various enforcement options).

⁹ *Id.* at 26-27, 30-31. See also, e.g., *Instagram Stands Against Online Bullying*, Instagram, <https://about.instagram.com/community/anti-bullying> (last visited Sept. 22, 2023) (warnings before posting "potentially offensive" comments); *Digital Well-being*, TikTok, <https://www.tiktok.com/safety/en-gb/well-being/> (last visited Sept. 22, 2023) (digital well-being and media literacy resource guide); *Violent and Graphic Content*, Meta Transparency Center, <https://transparency.fb.com/policies/community-standards/violent-graphic-content/> (last visited Sept. 22, 2023) (warning labels for graphic content); *Defining the Public Interest*, X Help Center, <https://help.twitter.com/en/rules-and-policies/public-interest> (last visited Sept. 21, 2023) (public click-through notice added to posts by government official that violate content policies and would otherwise be taken down, but are left up under X's public interest exception); *A New Policy Against Self-Harm Blogs*, Tumblr (Feb. 23, 2012), <https://staff.tumblr.com/post/18132624829/self-harm-blogs> (proposing adding PSAs to self-harm-related search results).

2. Social Media Platforms Have Rules, Standards and Guidelines about What Content They Want and Don't Want on Their Sites.

Social media platforms' content policies commonly prohibit users from posting speech that a platform believes is detrimental to its users and the public, its business interests, its editorial preferences, or all of these, even if that speech is legal. For example, many platforms ban legal, non-obscene sexual content, even though such speech enjoys First Amendment protection, *see Miller v. California*, 413 U.S. 15 (1973).¹⁰

Content moderation differs from platform to platform.¹¹ Some platforms detect potentially violating content only after it is posted; others screen some or all content *ex ante*.¹² Platforms make different judgment calls about whether particular content violates their content policies, even if those policies are similar.¹³ They use

¹⁰ *See, e.g., Adult Nudity and Sexual Activity*, Facebook, <https://transparency.fb.com/policies/community-standards/adult-nudity-sexual-activity/> (last visited Sept. 21, 2023).

¹¹ *Compare Community Guidelines*, Instagram, <https://help.instagram.com/477434105621119> (last visited Sept. 21, 2023) (prohibiting nudity except in the context of breastfeeding, birth-related moments, health-related situations, in paintings or sculptures, or as an act of protest), *with Sensitive Media Policy*, Twitter (March 2023), <https://help.twitter.com/en/rules-and-policies/media-policy> (permitting “consensually produced adult nudity”).

¹² Klonick, *supra* n.6, at 1635.

¹³ *See, e.g., Hannah Denham, Another Fake Video of Pelosi Goes Viral on Facebook*, Wash. Post (Aug. 3, 2020), <https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/> (reporting that TikTok, Twitter and YouTube removed a doctored video of Rep. Nancy Pelosi, while Facebook allowed it to remain with a label).

different methods to enforce their content policies, such as labeling content, placing interstitial warnings over it, or removing the ability to make money from it.¹⁴ Some platforms allow users to appeal content moderation decisions, while others do not.¹⁵

Although many platforms choose to prohibit speech expressing hatred based on race, ethnicity, religion, and other characteristics, they do so in divergent manners. For example, social media platform Gettr explains that it “holds freedom of speech as its core value and does not wish to censor your opinions,” while at the same time reserving the right to “address” content that attacks any religion or race.¹⁶ Reddit’s content policy prohibits content that promotes “hate based on identity or vulnerability,” including race, religion, and national origin.¹⁷

3. Content Moderation Has Long Been and Remains a Fraught and Controversial Process.

Content moderation controversies are not a new problem.

In 2007, YouTube, only two years old at the time, shut down the account of Egyptian human rights activist Wael Abbas after receiving multiple reports that the

¹⁴ Goldman, *supra* n.8, at 23–39.

¹⁵ Klonick, *supra* n.6, at 1648.

¹⁶ *Gettr – Terms of Use*, Gettr (May 17, 2023), <https://gettr.com/terms>.

¹⁷ *Promoting Hate Based on Identity or Vulnerability*, Reddit, <https://www.reddithelp.com/hc/en-us/articles/360045715951> (last visited Sept. 21, 2023).

account featured graphic videos of police brutality and torture.¹⁸ YouTube’s community standards at the time stated that “[g]raphic or gratuitous violence is not allowed.”¹⁹ Just one year before, Abbas became the first blogger to receive the Knight International Journalism Award.²⁰

And government’s attempts to influence content moderation date back just as far: Abbas’s account was restored only after the U.S. State Department communicated with YouTube’s new owner, Google.²¹

Content moderation remains a difficult and often fraught process that even the largest and best-resourced social media companies struggle with, often to the frustration of users. Even when using a set of precise rules or carefully articulated “community standards,” moderated platforms often struggle to draw workable lines between permitted and forbidden speech. Every online forum for user speech, not just the dominant social media platforms, struggles with this problem.

Platforms’ content moderation decisions are thus sometimes inconsistent or seemingly contrary to their own policies. Some of that is inevitable. Given the

¹⁸ Kevin Anderson, *YouTube Suspends Egyptian Blog Activist’s Account*, Guardian (Nov. 28, 2007), <https://www.theguardian.com/news/blog/2007/nov/28/youtubesuspendegyptianblog>.

¹⁹ *Id.*

²⁰ Jillian C. York, *Silicon Values: The Future of Free Speech Under Surveillance Capitalism* 25-27 (Verso 2021).

²¹ *Id.*

staggering amounts of content posted on platforms every day and the subjective judgment calls that some content moderation decisions require, platforms make mistakes in either moderating or failing to moderate content.²²

Beyond mistakes, platforms have often aggressively removed content that is not prohibited by their content policies, especially when attempting to minimize legal or reputational risks arising from government regulation or criticism. For example, many platforms responded to the enactment of the Allow States and Victims to Fight Online Sex Trafficking Act/Stop Enabling Sex Traffickers Act (“FOSTA”) by removing content by sex workers and sex worker advocates that is not actually prohibited by FOSTA.²³ *See Woodhull Freedom Foundation v. United States*, 72 F.4th 1286, 1299-1305 (D.C. Cir. 2023) (describing FOSTA’s very limited prohibitions). Government pressure to remove terrorist content from platforms has also led to over-removals of speech. For instance, in 2021, Instagram removed posts about one of Islam’s holiest mosques, Al-Aqsa, because its name is contained within the name of an organization the company had designated as a

²² See Mike Masnick, *Content Moderation At Scale Is Impossible: Recent Examples Of Misunderstanding Context*, TechDirt (Feb. 26, 2021), <https://www.techdirt.com/2021/02/26/content-moderation-scale-is-impossible-recent-examples-misunderstanding-context/>.

²³ See Danielle Blunt et al., *Posting Into The Void*, Hacking//Hustling (2020), <https://hackinghustling.org/wp-content/uploads/2020/09/Posting-Into-the-Void.pdf>.

terrorist group.²⁴

Platforms make a variety of contentious and seemingly erroneous decisions every day, and the impact of those decisions is felt across the political spectrum. In January 2021, Facebook’s updated policy to remove “harmful conspiracy theories” resulted in it disabling a punk rock band’s page because its name, Adrenochrome, is a chemical that was a central part of the QAnon conspiracy theory.²⁵ YouTube has removed videos documenting atrocities in Syria and elsewhere under its graphic violence policy.²⁶ YouTube has also been accused of restricting and demonetizing LGBTQ+ content.²⁷ Twitter has been repeatedly criticized for moderating pro-Palestinian tweets, including removing those reporting on the

²⁴ Ryan Mac, *Instagram Censored Posts About One of Islam’s Holiest Mosques, Drawing Employee Ire*, BuzzFeed News (May 12, 2021), <https://www.buzzfeednews.com/article/ryanmac/instagram-facebook-censored-al-aqsa-mosque>.

²⁵ *Facebook Treats Punk Rockers Like Crazy Conspiracy Theorists, Kicks Them Offline*, EFF, <https://www.eff.org/takedowns/facebook-treats-punk-rockers-crazy-conspiracy-theorists-kicks-them-offline> (last visited Sept. 21, 2023).

²⁶ Malachy Browne, *YouTube Removes Videos Showing Atrocities in Syria*, N.Y. Times (Aug. 22, 2017), <https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html>; Anderson, *supra* n.18.

²⁷ Megan Farokhmanesh, *YouTube Is Still Restricting and Demonetizing LGBT Videos—and Adding Anti-LGBT Ads to Some*, Verge (June 4, 2018), <https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-domentization-ads-algorithm>.

events in Sheikh Jarrah in 2021²⁸ and blocking accounts associated with a major Palestinian news publication.²⁹ In 2017, users protested that Twitter had marked tweets containing the word “queer” as offensive.³⁰

Some platforms structure their services so that some content moderation decisions are made by a broader community of users. Reddit and Discord rely on certain users to moderate content through the practice of “community moderation.”³¹ Reddit users manage and create thousands of communities, called “subreddits.” Although Reddit has an overriding content policy, a moderator makes the decisions within each community as guided by Reddit’s “Moderator Code of Conduct.”³² Discord employs a similar model.³³ Each site thereby

²⁸ Article 19, *Sheikh Jarrah: Facebook and Twitter Silencing Protestors, Deleting Evidence* (May 10, 2021), <https://www.article19.org/resources/sheikh-jarrah-facebook-and-twitter-silencing-protests-deleting-evidence>.

²⁹ *Twitter Suspends Accounts of Palestinian Quds News Network*, Al Jazeera (Nov. 2, 2019), <https://www.aljazeera.com/news/2019/11/2/twitter-suspends-accounts-of-palestinian-quds-news-network>.

³⁰ Taylor Wofford, *Twitter Was Flagging Tweets Including the Word “Queer” as Potentially “Offensive Content”*, Mic (June 22, 2017), <https://www.mic.com/articles/180601/twitter-was-flagging-tweets-including-the-word-queer-as-potentially-offensive-content>.

³¹ See *Moderator Code of Conduct*, Reddit, <https://www.redditinc.com/policies/moderator-code-of-conduct> (effective Sept. 8, 2022); *Role of Administrators and Moderators on Discord*, Discord, <https://discord.com/safety/360044103531-role-of-administrators-and-moderators-on-discord> (last visited Sept. 21, 2023); see also *Policy: Terms of Use*, Wikimedia, https://foundation.wikimedia.org/wiki/Policy:Terms_of_Use (last visited Sept. 21, 2023).

³² Reddit, *supra* n.31.

³³ Discord, *supra* n.31.

empowers some users to remove and down-rank other users' speech if that speech is against the community's rules.³⁴ Platforms also commonly rely on users to report content that violates the law or the platform's policies.³⁵

Most platforms, even those which do not employ community moderation, allow users to report or "flag" content they believe violates the platforms' rules or standards.³⁶ Indeed, such flags may account for a large amount of content moderation decisions. In the first quarter of 2023, YouTube removed 6.5 million videos, 360,000 of which were flagged by users.³⁷ During that same period, 34.8% of the 6.9 million posts Facebook actioned for bullying and harassment was reported by users, while 18% of the 10.7 million posts they actioned for hate speech was reported by users.³⁸ In the second half of 2021, users reported over 11

³⁴ See, e.g., *Reddiquette*, Reddit, <https://reddit.zendesk.com/hc/en-us/articles/205926439-Reddiquette> (last visited Sept. 21, 2023).

³⁵ See, e.g., *Report Content on Facebook*, Facebook, <https://www.facebook.com/help/181495968648557?rdrhc> (last visited Sept. 21, 2023).

³⁶ See generally Kate Crawford & Tarleton Gillespie, *What Is a Flag For? Social Media Reporting Tools and the Vocabulary of Complaint*, 18 *New Media & Soc'y* 410 (2014).

³⁷ *YouTube Community Guidelines Enforcement*, Google, <https://transparencyreport.google.com/youtube-policy/removals> (last visited Sept. 21, 2023).

³⁸ *Community Standards Enforcement Report: Bullying and Harassment*, Meta, <https://transparency.fb.com/reports/community-standards-enforcement/bullying-and-harassment/facebook/> (last visited Sept. 21, 2023); *Community Standards Enforcement Report: Hate Speech*, Meta, <https://transparency.fb.com/reports/community-standards-enforcement/hate-speech/facebook/> (last visited Sept. 21, 2023).

million accounts to Twitter as having violated at least one of its rules.³⁹

B. Content Moderation is Protected by the First Amendment.

1. The First Amendment protects the right to speak by curating the speech of others.

The Supreme Court has long held that private publishers have a First Amendment right to control the content of their publications, and specifically whether and how to publish things written by others. *See Miami Herald Publishing Co. v. Tornillo*, 418 U.S. 241, 254-55 (1974). *Cf. Manhattan Community Access Corp. v. Halleck*, 139 S. Ct. 1921, 1930 (2019) (reaffirming that “when a private entity provides a forum for speech,” “[t]he private entity may . . . exercise editorial discretion over the speech and speakers in the forum”). *See also Los Angeles v. Preferred Communications, Inc.*, 476 U.S. 488, 494 (1986) (recognizing cable television providers’ First Amendment right to “exercis[e] editorial discretion over which stations or programs to include in [their] repertoire[s]”).

Editorial freedom includes not only the right to devise and implement an editorial policy, but also protections for privacy in doing so. *See Bursey v. United States*, 466 F.2d 1059 (9th Cir. 1972) (overturning on First Amendment grounds grand jury contempt order that sought to compel disclosure of editorial decision-makers and their motives). Thus, in *Application of Consumers Union of U. S., Inc.*,

³⁹ *Rules Enforcement*, Twitter (July 28, 2022), <https://transparency.twitter.com/en/reports/rules-enforcement.html#2021-jul-dec>.

the district court quashed third-party subpoenas that sought the publisher's research methodology and "procedure by which he formed his conclusions." 495 F. Supp. 582, 585 (S.D.N.Y. 1980). The court explained that discovery seeking "to examine the reportorial and editorial processes . . . would represent a substantial intrusion on fact gathering and editorial privacy which are significant aspects of a free press." *Id.* at 586.⁴⁰ And although the Supreme Court declined to create a First Amendment evidentiary privilege when the publisher is the defendant and "there is a specific claim of injury arising from a publication that is alleged to have been knowingly or recklessly false," it stated that a "law that subjects the editorial process to private or official examination merely to satisfy curiosity or to serve some general end such as the public interest . . . would not survive constitutional scrutiny" under the First Amendment. *Herbert v. Lando*, 441 U.S. 153, 174 (1974).

It is well established that the First Amendment right of editorial freedom extends beyond the publication of one's own speech to the curation of others' speech. Even the typical newspaper is a mix of original writing and content created by others, including syndicated and wire services, advertisements, wedding, engagement, and birth announcements, and comics. Opinion pages, the specific forum targeted by the regulation that *Tornillo* struck down, typically publish a lot

⁴⁰ See also *In re Consumers Union of the U.S., Inc.*, 32 Fed. R. Serv. 2d 1373 (S.D.N.Y. 1981); *Kenneally v. Suzuki Motor Co., Ltd.*, No. M-8-85, 1994 WL 48840 (S.D.N.Y. Feb. 10, 1994).

of content created by others: opinion pieces, letters to the editor, syndicated editorial cartoons and columns.⁴¹ The Supreme Court’s most powerful pronouncement of freedom of the press, *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964), centered on *The Times* publishing someone else’s unsolicited content, a paid advertisement. The Court found that *The Times*’ role as a host for the speech of others was critical to its decision: newspapers are “an important outlet for the promulgation of information and ideas by persons who do not themselves have access to publishing facilities—who wish to exercise their freedom of speech even though they are not members of the press.” *Id.* at 266.

Tornillo, and the First Amendment more broadly, applies regardless of how selectively a platform publishes speech submitted to it. Print news media operates along the same continuum of selectivity as the social media content moderation processes described above. Pennysavers, for example, local newspapers either entirely or primarily composed of classified advertisements, coupons, life

⁴¹ See Jack Shafer, *The Op-Ed Page’s Back Pages: A Press Scholar Explains How the New York Times Op-Ed Page Got Started*, Slate (Sept. 27, 2010), <https://slate.com/news-and-politics/2010/09/a-press-scholar-explains-how-the-new-york-times-op-ed-page-got-started.html> (describing how the pages opposite newspapers’ editorial pages became a forum for outside contributors to express views different from those expressed by the paper’s editorial board); Michael J. Socolow, *A Profitable Public Sphere: The Creation of the New York Times Op-Ed Page*, Commc’n & Journalism Fac. Scholarship (2010), https://digitalcommons.library.umaine.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1001&context=cmj_facpub; *Op-Ed*, Wikipedia, <https://en.wikipedia.org/wiki/Op-ed> (last visited Sept. 21, 2023).

milestone announcements, congratulatory messages, recipes, public notices, and the like, have a long and storied history as a relatively non-selective print publication.⁴² Curators, theater directors and booksellers select plays to produce and books to publish or sell along a similar continuum; their First Amendment rights do not depend upon falling on the proper side of a constitutionally arbitrary selectivity line.

Tornillo is thus not limited to only newspapers or publishers that actively select the content they publish, or to media entities at all. The Supreme Court has applied it to a variety of entities that speak by curating the speech of others, including thrice in the 2018 term. *See Janus v. American Federation of State, County, & Municipal Employees, Council 31*, 138 S. Ct. 2448, 2463 (2018); *Nat'l Inst. of Family & Life Advocates v. Becerra*, 138 S. Ct. 2361, 2371 (2018); *Masterpiece Cakeshop, Ltd. v. Colorado Civil Rights Comm'n*, 138 S. Ct. 1719, 1745 (2018) (Thomas, J., concurring).

2. Social media and other interactive websites have a constitutional right to adopt, define, and implement their own editorial policies.

Given this, it is not surprising that numerous courts have applied *Tornillo* to social media platforms that primarily, if not exclusively, publish user-generated

⁴² *Pennysaver*, Wikipedia, <https://en.wikipedia.org/wiki/Pennysaver> (last visited Sept. 21, 2023).

content.⁴³

The Eleventh Circuit recently applied *Tornillo* to social media platforms and their user content curation. *NetChoice LLC v. Attorney General, Florida*, 34 F.4th 1196, 1212 (11th Cir. 2022) (“*NetChoice (Florida)*”). It explained:

Social-media platforms exercise editorial judgment that is inherently expressive. When platforms choose to remove users or posts, deprioritize content in viewers’ feeds or search results, or sanction breaches of their community standards, they engage in First-Amendment-protected activity. Social-media platforms’ content-moderation decisions are, we think, closely analogous to the editorial judgments that the Supreme Court recognized in *Miami Herald*, *Pacific Gas*,⁴⁴ *Turner*,⁴⁵ and *Hurley*⁴⁶. . . . [¶] All such decisions about what speech to permit, disseminate, prohibit, and deprioritize—decisions based on platforms’ own particular values and views—fit comfortably within the Supreme Court’s editorial-judgment precedents.

⁴³ See, e.g., *Children's Health Def. v. Facebook Inc.*, 546 F. Supp. 3d 909, 922-23 (N.D. Cal. 2021), *appeal docketed*, No. 21-1620 (9th Cir. July 21, 2021); *La'Tiejira v. Facebook, Inc.*, 272 F. Supp. 3d 981, 991 (S.D. Tex. 2017); *Zhang v. Baidu.com Inc.*, 10 F. Supp. 3d 433, 436-38 (S.D.N.Y. 2014); *Langdon v. Google, Inc.*, 474 F. Supp. 2d 622, 630 (D. Del. 2007); *Huber v. Biden*, No. 21-CV-06580-EMC, 2022 WL 827248, at *6 (N.D. Cal. Mar. 18, 2022), *aff'd*, No. 22-15443, 2022 WL 17818543 (9th Cir. Dec. 20, 2022); *O'Handley v. Padilla*, 579 F. Supp. 3d 1163, 1186 (N.D. Cal. 2022), *aff'd sub nom. O'Handley v. Weber*, 62 F.4th 1145 (9th Cir. 2023); *Amer. Freedom Defense Initiative v. Lynch*, 217 F. Supp. 3d 100, 106 (D.D.C. 2016); *Stossel v. Meta Platforms, Inc.*, 634 F. Supp. 3d 743, 760 (N.D. Cal. 2022), *appeal dismissed*, No. 22-16765, 2023 WL 3434064 (9th Cir. Mar. 1, 2023).

⁴⁴ *Pacific Gas & Elec. Co. v. Pub. Utilities Comm'n of California*, 475 U.S. 1 (1986).

⁴⁵ *Turner Broadcasting Sys., Inc. v. F.C.C.*, 512 U.S. 622 (1994).

⁴⁶ *Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Boston*, 515 U.S. 557 (1995).

Id. at 1213-14.

In so doing, the Eleventh Circuit rejected the argument, based on the Supreme Court’s decisions in *PruneYard Shopping Center v. Robins*, 447 U.S. 74 (1980), and *Rumsfeld v. Forum for Academic & Institutional Rights, Inc.*, 547 U.S. 47 (2006), that the government had a greater ability to meddle in social media content moderation because content moderation wasn’t itself expressive. *NetChoice (Florida)*, 34 F.4th at 1215-19. That those cases do not apply when a speaker’s ability to express their own message is directly burdened by a law was recently affirmed by the Supreme Court. *303 Creative LLC v. Elenis*, 600 U.S. ___, 143 S. Ct. 2298, 2317-18 (2023).

The Ninth Circuit has similarly rejected the argument that social media platforms are state actors that are limited by the First Amendment in their ability to select the speech of others. *Prager Univ. v. Google LLC*, 951 F.3d 991, 995 (9th Cir. 2020).

This Court should not follow the recent contrary decision of the Fifth Circuit in *NetChoice, LLC v Paxton* (“*NetChoice (Texas)*”), which quizzically saw the Supreme Court’s First Amendment jurisprudence as secondary to what the court divined as “the original public meaning of the First Amendment.” 49 F.4th 439, 454-55 (5th Cir. 2022).

In that case, the Fifth Circuit, analyzing whether Texas could compel large

social media platforms to publish users' posts in a viewpoint-neutral manner, distinguished *Tornillo*. The Fifth Circuit wrongly saw social media platforms as mere "conduits" for user speech that, "[u]nlike newspapers, . . . exercise virtually no editorial control or judgment"—despite the fact that the Texas law in question was motivated by perceived editorial decision-making. *Id.* at 459-60 (quoting *Tornillo*, 418 U.S. at 258).

Nevertheless, *NetChoice (Texas)* is readily distinguishable from this case in at least two material ways. First, section 394-ccc applies well beyond large social media companies and includes all websites with interactive features like reader comments. The Fifth Circuit's blindered view of content moderation as a mere conduit thus does not apply. Second, the Fifth Circuit further distinguished *Tornillo* and similar cases on the basis that the Texas law did not prohibit the platforms themselves from speaking. But section 394-ccc applies to any and all speech on the site that a user might deem "hateful," including the publisher's own statements. *See* N.Y. Gen. Bus. Law § 394-ccc(2).

This protection for editorial and curatorial freedom does not disable governments from regulating online services. Regulatory measures that neither target the editorial process nor are enacted in retaliation against disfavored editorial policies and decisions may be acceptable. For example, governments can promote

user choice and control by encouraging competition⁴⁷ and platform interoperability.⁴⁸

II. Government Involvement in Content Moderation Raises Serious First Amendment Concerns.

Government involvement in private companies' content moderation processes, regardless of form or degree, raises broader serious human rights concerns. The recently revised Santa Clara Principles, of which *amicus curiae* Electronic Frontier Foundation is a co-author, specifically scrutinize "State Involvement in Content Moderation," and affirm that "state actors must not exploit or manipulate companies' content moderation systems to censor dissenters, political opponents, social movements, or any person."⁴⁹ "Special concerns are raised by demands and requests from state actors (including government bodies, regulatory authorities, law enforcement agencies and courts) for the removal of content or the suspension of accounts."⁵⁰

⁴⁷ *Comment on Evaluating the Competitive Effects of Corporate Acquisitions and Mergers*, EFF (Aug. 20, 2018), <https://www.eff.org/document/eff-comments-ftc-competition-0>.

⁴⁸ Bennett Cyphers & Cory Doctorow, *The New ACCESS Act Is a Good Start. Here's How to Make Sure It Delivers*, EFF (June 21, 2021), <https://www.eff.org/deeplinks/2021/06/new-access-act-good-start-heres-how-make-sure-it-delivers>.

⁴⁹ The Santa Clara Principles, <https://santaclaraprinciples.org> (last visited Sept. 21, 2023).

⁵⁰ *Id.* And reflecting this concern for government involvement, the Principles themselves are expressly not a template for regulation, but instead self-regulatory guidelines. *Id.*

As discussed above, content moderation systems are fraught, and governments have outsized influence to manipulate content moderation systems for their own political goals.

Courts must thus be sensitive to all government interference in content moderation, beyond direct mandates to publish or unpublish certain user posts. A government that indirectly or subtly coerces the adoption of an editorial decision or policy also violates the First Amendment.

III. The New York Law Unconstitutionally Coerces Interactive Websites to Adopt the State’s Hate Speech Philosophy at the Expense of Their Own Editorial Policies.

As Appellees ably argue, section 394-ccc seeks to unlawfully coerce them to replace their own editorial policies with the state’s.

The proper test to assess coercion, direct or indirect, originates from *Bantam Books, Inc. v Sullivan*, in which the Supreme Court found that the First Amendment prohibited not only direct censorship demands but also “system[s] of informal censorship” aimed at speech intermediaries. 372 U.S. 58, 71 (1963). The Supreme Court found that “the threat of invoking legal sanctions and other means of coercion, persuasion, and intimidation” against book distributors were enough to violate the book publishers’ First Amendment rights. *Id.* at 67.

In *Bantam Books*, the Court found that a Rhode Island commission unconstitutionally coerced a book distributor, even though the commission itself

did not have the ability to penalize the distributor. The commission issued notices to book distributors that “certain designated books,” published by plaintiffs, were “objectionable for sale,” and that it was the commission’s “duty to recommend to the Attorney General prosecution of purveyors of obscenity.” *Id.* at 61-62. The commission also circulated the notices to local police, who visited the distributor “to learn what action he had taken,” *id.* at 62-63, similar to the provision of section 394-ccc that requires websites to disclose how they will respond to complaints about hate speech, N.Y. Gen. Bus. Law § 394-ccc(3). Predictably, the distributors stopped selling the books. 372 U.S. at 64. The Court found that the publishers had a First Amendment remedy against the state commission, even though it was the distributor’s action that directly harmed the publishers’ sales, and the government did not actually seize any books or prosecute anyone. *Id.* at 64 n.6.

Cases following *Bantam Books* have applied a totality of the circumstances analysis and identified numerous factors relevant to determining when government improperly pressures a speech intermediary to censor its users. In *Backpage.com v. Dart*, for example, the Seventh Circuit followed *Bantam Books* to find unconstitutional government coercion. 807 F.3d 229, 230 (7th Cir. 2015). The case involved a sheriff’s campaign to shutdown Backpage.com’s adult section “by demanding that firms such as Visa and MasterCard prohibit the use of their credit cards to purchase any ads on Backpage.” *Id.* Among the more pertinent factors the

court considered was that there was clear government *intent* to coerce Visa and Mastercard to cooperate.⁵¹ Looking at the totality of the circumstances, the letter was not simply an effort to educate the recipients “about the nature and possible consequences of advertising for sex.” *Id.* at 237.

This Court has identified four factors to distinguish permissible “attempts to convince” from unconstitutional “attempts to coerce”: (1) the speaker’s “word choice and tone”; (2) “whether the speech was perceived as a threat”; (3) “the existence of regulatory authority”; and, “perhaps most importantly, (4) whether the speech refers to adverse consequences.” *National Rifle Ass’n v. Vullo*, 49 F.4th 700, 715 (2d Cir. 2022); *see also Okwedy v. Molinari*, 333 F.3d 339, 344 (2d Cir. 2003). Both the Fifth and Ninth Circuits have recently employed similar analyses. *See State of Missouri v. Biden*, No. 23-30445, 2023 WL 5821788, at *18-19 (5th Cir., Sept. 8, 2023) (finding that several federal Executive Branch agencies and officials unlawfully coerced social media platforms to remove the plaintiffs’ specific posts); *Kennedy v. Warren*, 66 F.4th 1199, 1207-10 (9th Cir. 2023) (finding no coercion by Senator Warren who merely expressed her disapproval of

⁵¹ A strategic memo recommended appealing to the intermediaries’ interest in avoiding liability; the sheriff took credit in a press release for “compelling” the companies’ actions with his “demand”; and his office sent urgent communications to the companies following up on his letter which “imposed another layer of coercion due to [their] strong suggestion that the companies could not simply ignore [the sheriff].” *Id.* at 232, 237.

Kennedy's social media posts).

Applying those factors, as Appellees explain, the statute being challenged here is just one part of an overall effort by the governor, the attorney general, and the legislature to coerce interactive websites to replace their editorial policies regarding hateful speech, with that of the state. Though the statute itself does not directly require such adoption, it is clearly the state's goal. At a minimum, the statute is clearly intended to burden websites with complaints from users about its hate speech editorial policies.

This scheme, targeting legal speech the government itself cannot directly outlaw, is unconstitutional coercion.

CONCLUSION

Interactive websites have a First Amendment right to curate and edit their sites pursuant to their own editorial policies and without interference from the State. The New York law is part of a broader campaign to coerce interactive websites to replace their own editorial policies and practices with those mandated by the State and is thus unconstitutional. The Court should affirm the district court's order enjoining it.

Dated: September 26, 2023

Respectfully submitted,

/s/ David Greene

David Greene

Electronic Frontier Foundation

815 Eddy Street

San Francisco, California 94109
Tel.: (415) 436-9333
davidg@eff.org

Brian Hauss
American Civil Liberties Union Foundation
125 Broad St., 18th Fl.
New York, NY 10004
Tel.: (212) 549-2500
bhauss@aclu.org

*Attorneys for Amici Curiae
Electronic Frontier Foundation and
American Civil Liberties Union*

CERTIFICATE OF COMPLIANCE

Pursuant to Fed. R. App. P. 32(a)(7)(C), I certify as follows:

1. This Brief of *Amici Curiae* Electronic Frontier Foundation and American Civil Liberties Union in Support of Plaintiffs-Appellees and Affirmance complies with the type-volume limitation of Fed. R. App. P. 32(a)(7)(B) because this brief contains 5,446 words, excluding the parts of the brief exempted by Fed. R. App. P. 32(a)(7)(B)(iii); and

2. This brief complies with the typeface requirements of Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. P. 32(a)(6) because this brief has been prepared in a proportionally spaced typeface using Microsoft Word 2016, the word processing system used to prepare the brief, in 14-point font in Times New Roman font.

Dated: September 26, 2023

/s/David Greene
David Greene

CERTIFICATE OF SERVICE

I certify that on this 26th day of September 2023, I electronically filed the foregoing using the Court's CM/ECF system which will send notification of such filing to all parties of record.

Dated: September 26, 2023

/s/ David Greene
David Greene