

Nos. 22-277, 22-555

IN THE
Supreme Court of the United States

ASHLEY MOODY, ATTORNEY
GENERAL OF FLORIDA, *et al.*,

Petitioners,

v.

NETCHOICE, LLC, D/B/A NETCHOICE, *et al.*,

Respondents.

NETCHOICE, LLC, D/B/A NETCHOICE, *et al.*,

Petitioners,

v.

KEN PAXTON, ATTORNEY GENERAL OF TEXAS,

Respondent.

ON WRITS OF CERTIORARI TO THE UNITED STATES COURT OF
APPEALS FOR THE FIFTH AND ELEVENTH CIRCUITS

**BRIEF OF *AMICUS CURIAE* THE TRUST & SAFETY
FOUNDATION IN SUPPORT OF RESPONDENTS IN
NO. 22-277 AND PETITIONERS IN NO. 22-555**

BETHANY J. M. PANDHER
GREENBERG TRAUIG, P.A.
333 S.E. 2nd Avenue
Miami, FL 33131

IAN C. BALLON
Counsel of Record
LORI CHANG
GREENBERG TRAUIG, LLP
1840 Century Park East,
Suite 1900
Los Angeles, CA 90067
(310) 586-7700
ballon@gtlaw.com

Counsel for Amicus Curiae

TABLE OF CONTENTS

	Page
TABLE OF CONTENTS	i
TABLE OF CITED AUTHORITIES.....	iii
STATEMENT OF INTEREST.....	1
SUMMARY OF ARGUMENT.....	2
ARGUMENT.....	3
I. TRUST & SAFETY PROFESSIONALS PERFORM THE ESSENTIAL FUNCTION OF CONTENT MODERATION IN THE MODERN INTERNET	3
A. The Trust & Safety Profession Has Developed Alongside Platforms to Lead All Aspects of Content Moderation.....	3
B. Content Moderation in Today’s Digital Forums Is Complex and Benefits Users and Platforms Alike	8
II. THE REGULATORY CONSTRAINTS ON CONTENT MODERATION IMPOSED BY TEXAS AND FLORIDA LIKELY WILL YIELD UNINTENDED CONSEQUENCES THAT SUPPRESS SPEECH.....	13
A. Increased Costs Associated with Exercising Editorial Control Likely Will Result in Policy and Enforcement Changes That Suppress Speech	15

B. Mandatory Notice and Appeal Requirements Incentivize Platforms to Reduce Opportunities for Online Speech and Encourage Consolidation of Existing Platforms	21
C. Sweeping Carriage Obligations Undermine Platform Functionality and Contravene User Desirability.....	26
CONCLUSION	29

TABLE OF CITED AUTHORITIES

Page(s)

Cases:

Cox v. Reddit, Inc.,
 No. S22-87J1 (Just. Ct. Denton Cnty.,
 Tex. May 17, 2022)..... 19

Netchoice, LLC v. Paxton,
 573 F. Supp. 3d 1092 (W.D. Tex. 1099)..... 26

Statutes

Fla. Stat. § 2041(h)..... 26

Fla. Stat. § 501.204(2)(d)..... 15

Fla. Stat. § 501.2041..... 1

Fla. Stat. § 501.2041(3) 10

Fla. Stat. § 501.2041(3)(b)..... 15

Fla. Stat. § 501.2041(6)(a)..... 13, 15

Fla. Stat. § 501.2051(3) 15

Tex. H.B. § 143.007(b)(1)..... 13

Tex. H.B. § 143.007(d) 13

Tex. H.B. § 143.007(e) 13

Tex. H.B. 20 1, 19. 26

Tex. H.B. 20 § 101 15

Tex. H.B. 20 § 102 15

Tex. H.B. 20 § 103 15

Tex. H.B. 20 § 104	15
Tex. H.B. 20 § 120.101	23
Tex. H.B. 20 § 143A.002.....	26

Other Authorities:

<i>Americans Support Free Speech Online but Want More Action to Curb Harmful Content</i> , KNIGHT FDN. (June 16, 2020), https://knightfoundation.org/press/releases/ americans-support-free-speech-online-but- want-more-action-to-curb-harmful-content .	27, 28
Bernard Marr, <i>How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read</i> , FORBES (May 21, 2018), https://www.forbes.com/sites/bernardmarr/20 18/05/21/how-much-data-do-we-create-every- day-the-mind-blowing-stats-everyone-should- read/	8
<i>Community Guidelines: Overview</i> , YOUTUBE (2023), https://www.youtube.com/howyoutubeworks/ policies/community-guidelines/	5
<i>Community Standards Enforcement Report: Q3 2023 report</i> , META (2023), https://transparency.fb.com/reports/communi ty-standards-enforcement/	12

- Corrections and Adjustments: Updated Methodology for Appealed Content on Facebook and Instagram*, META (Aug. 2022), <https://transparency.fb.com/policies/improving/corrections-adjustments>..... 24
- Daphne Keller, *Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List*, STANFORD L. SCH.: CTR. FOR INTERNET & SOC'Y (Feb. 8, 2021), <https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws/>..... 14
- Daphne Keller, *Lawful but Awful? Control of Legal Speech by Platforms, Governments, and Internet Users*, U. CHI. L. REV. ONLINE ARCHIVE (June 28, 2022), <https://lawreviewblog.uchicago.edu/2022/06/28/keller-control-over-speech>..... 5
- Daphne Keller, *Platform Transparency and the First Amendment*, 4 J. Free Speech L. 1 (2023), <https://journaloffreespeechlaw.org/keller2.pdf>..... 10, 17, 26
- Daphne Keller, *Reported Monthly Active Usage Data for Content Hosting Platforms*, at 15 (Apr. 29, 2022), <https://www.judiciary.senate.gov/imo/media/doc/Keller%20Testimony1.pdf> ... 26
- DSA Transparency Database: Analytics, TikTok*, EUROPEAN COMMISSION, <https://transparency.dsa.ec.europa.eu/analytics> 9

- DSA Transparency Database: Submission of clear and specific statements*, EUROPEAN COMMISSION, <https://transparency.dsa.ec.europa.eu/page/documentation>..... 19
- Emmet Lyons, *Elon Musk’s X Platform Fueled Far-Right Riots in Ireland, Experts Say*, CBS NEWS (Dec. 5, 2023), <https://www.cbsnews.com/news/elon-musk-ireland-x-twitter-far-right-dublin-immigration/>..... 27
- Facebook Submission in Response to Subcommittee Questions for the Record, *Hearing on Disinformation Nation: Social Media’s Role in Promoting Extremism and Misinformation, Subcomm. On Commc’n & Tech. and the Subcomm on Consumer Prot. & Com. of the H. Comm. On Energy & Com.*, 117th Cong. (2021), <https://bit.ly/3QU0R4w>..... 10
- Facebook Unveils Appeal Process For When It Removes Posts*, YAHOO NEWS (Apr. 24, 2018), <https://sg.news.yahoo.com/facebook-unveils-appeal-process-removes-posts-092814445.html> 7
- Glyn Moody, *Copyright as Censorship: Abuse of the DMCA To Try to Delete Online News Is Rampant*, TECHDIRT (May 24, 2022), <https://www.techdirt.com/2022/05/24/copyright-as-censorship-abuse-of-the-dmca-to-try-to-delete-online-news-is-rampant/>..... 18

- Google Submission in Response to
 Subcommittee Questions for the Record,
*Hearing on Disinformation Nation: Social
 Media’s Role in Promoting Extremism and
 Misinformation, Subcomm. On Commc’n &
 Tech. and the Subcomm on Consumer Prot.
 & Com. of the H. Comm. On Energy & Com.*,
 117th Cong. (2021), <https://bit.ly/3He9sMf>..... 10
- Hanlin Li, *et al.*, *Measuring the Monetary Value
 of Online Volunteer Work*, 16 PROCEEDINGS
 OF THE SIXTEENTH INT’L AAAI CONF. ON WEB
 & SOCIAL MEDIA (2022), [https://aaai.org/
 papers/00596-19318-measuring-the-
 monetary-value-of-online-volunteer-work/](https://aaai.org/papers/00596-19318-measuring-the-monetary-value-of-online-volunteer-work/) 11
- Harsha Bhatlapenumarthy & James Gresham,
*Content Moderation and Operations: What Is
 Content Moderation?*, TR. & SAFETY PROF.
 ASS’N, [https://www.tspa.org/curriculum/ts-
 fundamentals/content-moderation-and-
 operations/what-is-content-moderation/](https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/what-is-content-moderation/) 6
- Harsha Bhatlapenumarthy & James Gresham,
User Appeals, TR. & SAFETY PROF. ASS’N,
[https://www.tspa.org/curriculum/ts-
 fundamentals/content-moderation-and-
 operations/user-appeals/](https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/user-appeals/) 21
- Help Center: I Don’t Think Facebook Should
 Have Taken Down My Post*, FACEBOOK,
[https://www.facebook.com/help/20908563312
 03011?helpref=faq_content](https://www.facebook.com/help/2090856331203011?helpref=faq_content)..... 22
- Help Center: Submit an Appeal*, PINTEREST,
[https://help.pinterest.com/en/article/review-
 your-reports-and-violations](https://help.pinterest.com/en/article/review-your-reports-and-violations) 22

<i>Integrity Reports, Third Quarter 2023</i> , META, https://transparency.fb.com/integrity-reports-q3-2023	24
Jan Eissfeldt, Jeff Lazarus & Pia Shah, <i>Industry Overview</i> , TR. & SAFETY PROF. ASS'N, https://www.tspa.org/curriculum/ts-fundamentals/industry-overview/intro-to-ts	4
Jennifer M. Urban <i>et al.</i> , <i>Notice and Takedown in Everyday Practice</i> , UC BERKLEY L. RES. PAPER NO. 2755628 (Mar. 22, 2017), http://dx.doi.org/10.2139/ssrn.2755628	14
<i>Macrotrends, Meta Platforms: Number of Employees 2010–2022</i> , META, https://www.macrotrends.net/stocks/charts/META/meta-platforms/number-of-employees	7
Maggie Engler, Jeff Lazarus & James Gresham, <i>Automated Systems and Artificial Intelligence</i> , TR. & SAFETY PROF. ASS'N, https://www.tspa.org/curriculum/ts-fundamentals/automated-systems-and-ai/uses	11, 12
Mike Masnick, <i>Masnick's Impossibility Theorem: Content Moderation At Scale Is Impossible To Do Well</i> (Nov. 29, 2019), https://techdirt.com/2019/11/20/masnicks-impossibility-theorem-content-moderation-scale-is-impossible-to-do-well/	8
North America Man/Boy Love Association, WIKIPEDIA, https://en.wikipedia.org/wiki/North_America_n_Man/Boy_Love_Association	27

Number of Worldwide Social Network Users from 2017 to 2027, STATISTA (June 2022), <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>..... 8-9

Shubham Atreja *et al.*, *Remove, Reduce, Inform: What Actions Do People Want Social Media Platforms to Take on Potentially Misleading Content*, at 1 (October 2023), <https://bit.ly/3jTauUo>..... 27

Sup. Ct. R. 37.6..... 1

User Appeals: Processing Appeals, TR. & SAFETY PROF. ASS'N, <https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/user-appeals/>..... 23

STATEMENT OF INTEREST¹

The Trust & Safety Foundation (TSF) is a 501(c)(3) nonprofit charitable organization that brings together key stakeholders to engage in interdisciplinary dialogue, education, and research programs to improve society's understanding of the field of trust and safety (T&S). TSF's related organization, the Trust & Safety Professional Association (TSPA), is a 501(c)(6) nonpartisan membership association that supports the global community of professionals who develop and enforce principles, policies, and practices that define acceptable behavior and content online. TSPA works to create and foster a global community of T&S professionals, collaborating with them to build a community of practice, and providing support as they do the critical work of keeping online platforms safe. Neither TSF nor TSPA make recommendations about the editorial rules that platforms should apply to particular speech.

T&S operations and practices are complex and develop in tandem with the fast-paced landscape of the modern Internet to benefit both users and platforms. The goal of this brief is to explain the operational realities of content moderation by T&S teams and the practical impact of Texas's H.B. 20 and Florida Statute § 501.2041, through which certain

¹ Pursuant to Rule 37.6, counsel for amici state that no party's counsel authored any part of this brief and no party's counsel or party made a monetary contribution intended to fund the preparation or submission of the brief. Only the amici and its attorneys have funded the preparation and submission of this brief.

state governments seek to regulate platform operators' content moderation efforts, which in turn are likely to chill speech.

SUMMARY OF ARGUMENT

As the Internet has grown and social media platforms have flourished, content moderation has emerged as a critical tool deployed by platforms to ensure safety and privacy, support free expression, and generate trust, revenue, and growth. To that end, the T&S profession leads the way in crafting policy, designing processes, and executing enforcement strategies that allow users to connect, share, and interact.

Content moderation is challenging, nuanced, and ever-changing. And there is no "one size fits all." With the advent of user-generated content and global trends of increased Internet use, T&S provides critical support to maintain functioning and viable digital communities. T&S teams have matured from an initial reactive posture of responding to online abuse to developing forward looking policies and overseeing operations to curate and organize content in diverse ways that promote positive user experiences. This includes the development and use of predictive machine learning and artificial intelligence (AI) tools, supported by human oversight, to manage the sheer volume of content available online.

Texas and Florida endeavor to broaden the channels for digital user speech by legislating against platforms' exercise of editorial control. They claim that by imposing automatic notice and appeal requirements, along with mandates for viewpoint-neutral policies, users will benefit from fuller

opportunities for expression and discourse. These burdens, however, are likely to hamper the ability and willingness of platforms to host user-generated content and are likely to yield unintended consequences that will suppress free speech. To preserve the utility and diversity of online forums, the laws should not be permitted to stand.

ARGUMENT

I. TRUST & SAFETY PROFESSIONALS PERFORM THE ESSENTIAL FUNCTION OF CONTENT MODERATION IN THE MODERN INTERNET.

A. The Trust & Safety Profession Has Developed Alongside Platforms to Lead All Aspects of Content Moderation.

Online services and digital technologies have ushered in an unprecedented era of connection and access. In this innovation wave, technology companies have increasingly recognized that while platforms and services drive growth and facilitate engagement, the threat of abuse by bad actors engaging in harmful and unwanted behaviors is real and pervasive. Online companies and service providers therefore must understand the types of content and behaviors that are potentially harmful or could lead to violence and intimidation. The T&S profession has emerged to fulfill this role, among other critical functions that ensure the health and safety of society and the Internet by protecting users from potentially dangerous experiences online.

T&S involves review of user-generated content to vet for compliance with (1) all relevant laws and (2) the platform's specific policies and guidelines. The

process of moderating content and enforcing policy is done in different ways, depending on the maturity, scale, and needs of each platform. Companies and organizations may rely on user reporting to identify potential violations or proactively enforce their community guidelines using manual review by people, through automation, or some combination of both. Along with content moderation, many T&S teams contribute data to train machine learning algorithms and artificial intelligence (AI) to enforce platforms' chosen rules preemptively and advise product teams on design recommendations to minimize abuse.

Early trust and safety teams responded to common forms of online abuse like fraud, scams, phishing, and spam, which surfaced during the commercialization of the Internet and its rapid growth in the 1990s. Jan Eissfeldt, Jeff Lazarus & Pia Shah, *Industry Overview*, TR. & SAFETY PROF. ASS'N, <https://www.tspa.org/curriculum/ts-fundamentals/industry-overview/intro-to-ts>. Since then—particularly with the advent of social media and the resulting proliferation of user-generated content—T&S has evolved from merely identifying, minimizing, and removing content based on narrow legal mandates or hastily adopted rules, to the more significant role of creating the nuanced policies that govern a company's products and services and define its brand, as well as developing the tools, systems, and techniques for enforcing them. Often called “community standards” or “community guidelines,” these policies work as editorial rules within the platform to set uniform standards for behavior. These often vary significantly from company to company.

An unsurprising hallmark of many policies is the prohibition against illegal or unsafe speech. *E.g.*, *Community Guidelines: Overview*, YOUTUBE (2023), <https://www.youtube.com/howyoutubeworks/policies/community-guidelines/> (“Community Guidelines” covering various categories, including deceptive practices, violent or dangerous content, sensitive content, and regulated goods). Although some categories of speech do not rise to the level of illegality, companies and organizations may still decide to exclude or limit them on grounds ranging from founders’ personal beliefs to advertisers’ and consumers’ preferences. These categories often consist of “material that cannot be prohibited by law but that profoundly violates many people’s sense of decency, morality, or justice.” Daphne Keller, *Lawful but Awful? Control of Legal Speech by Platforms, Governments, and Internet Users*, U. CHI. L. REV. ONLINE ARCHIVE (June 28, 2022), <https://lawreviewblog.uchicago.edu/2022/06/28/keller-control-over-speech>. Once developed, guidelines are public and typically shared within a platform’s Terms of Service or Transparency page.

These guidelines must be broad enough to cover individual content, actors, and behavior on a platform. Individual content is the user-generated content that we traditionally associate with online platforms, including videos, posts, or comments shared by users. Actors are those users or entities who share the content and engage in online behavior. Such behavior includes initially policy-compliant behavior that becomes abusive in the aggregate over time and over the course of multiple actions. For example, a user may violate platform rules if he repeatedly sends the

same message to someone, sends objectionable messages to multiple users, uses bots or other means to artificially boost the popularity of content, or engages in acts designed to enable other violations.

Simplified, a healthy content moderation system has three key components: the policy team, the product and engineering team, and the operations team. Harsha Bhatlapenumarthy & James Gresham, *Content Moderation and Operations: What Is Content Moderation?*, TR. & SAFETY PROF. ASS'N, <https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/what-is-content-moderation/>. *Policy* teams create the rules that govern the content moderation process. *Product and engineering* teams support that process by building the required tools and infrastructure for enforcement, as well as the overall experience for people using the website or product. This includes building public-facing tools for users to report objectionable content as well as creating the internal systems and tools for both manual enforcement (*i.e.*, review by content moderators) and automated systems or AI. *Operations* teams establish the appropriate workforce model and build sustainable processes for platform moderators to monitor and iterate on the content moderation system. Together, these teams collaborate to develop a platform's strategy for T&S. The way T&S teams are structured to execute these functions varies widely; however, collaboration is essential to building and maintaining a robust content moderation system.

For any particular content moderation decision, the journey typically begins when content or an account is flagged for review. This may occur by a user report or

a platform's proactive identification of policy violations, often using machine learning or AI. For larger platforms, once a report is created, automated tools may assess whether the content is policy-violating, and then either enforce (for violative content) or take no action (for permissible content). If automation does not resolve the issue, the content is queued for manual review by a content moderator. Some platforms may use AI to rank and prioritize content for review based on factors including potential harm to the user. Whether the review is done manually or through automation, the policy assessment process is the same and the same enforcement actions result.²

A myriad of results flow from content moderation. Depending on the company and the nature of the content, as well as other editorial and business considerations, outcomes range from content deletion, banning, temporary suspension, feature blocking, reduced visibility, labeling, demonetization, withholding payments, and referral to law enforcement. T&S teams execute with consistency as a goal, and aim to craft policies and design processes

² It is noteworthy that Texas's and Florida's laws impose notice and appeal burdens that giants in the social media industry shouldered only very late in their growth. Indeed, "Facebook first allowed users to appeal removals of photos, videos, and posts beginning in 2018, when the company was worth \$374 billion and had some 35,000 employees." *Facebook Unveils Appeal Process For When It Removes Posts*, *YAHOO NEWS* (Apr. 24, 2018), <https://sg.news.yahoo.com/facebook-unveils-appeal-process-removes-posts-092814445.html>; *Macrotrends, Meta Platforms: Number of Employees 2010–2022*, META, <https://www.macrotrends.net/stocks/charts/META/meta-platforms/number-of-employees>).

to get as close to that result as they can. But perfect consistency across millions of individual decisions is fundamentally impossible. If one million content moderation decisions on a platform are made every day, even if the platform realizes 99.9% accuracy, 1,000 determinations will be incorrect. Mike Masnick, *Masnick's Impossibility Theorem: Content Moderation At Scale Is Impossible To Do Well* (Nov. 29, 2019), <https://techdirt.com/2019/11/20/masnick-impossibility-theorem-content-moderation-scale-is-impossible-to-do-well/>. Of course, the largest platforms far surpass a million content moderation decisions daily—indeed, *every minute* of the day 456,000 tweets are sent on Twitter, 46,740 photos are posted on Instagram, and 510,000 comments are posted on Facebook, such that even a near perfect 0.1% error rate would yield well over 1,000 mistaken determinations each day. Bernard Marr, *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*, FORBES (May 21, 2018), <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>.

B. Content Moderation in Today's Digital Forums Is Complex and Benefits Users and Platforms Alike.

With the proliferation of user-generated content on digital platforms, the task of content moderation is more daunting than ever. In 2022, over 4.59 billion people used social media globally. This number is projected to increase to nearly 6 billion in 2027. *Number of Worldwide Social Network Users from 2017 to 2027*, STATISTA (June 2022),

<https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>.

Last year, users spent an average of 151 minutes daily on social media and messaging apps, creating an ever-increasing volume of user-generated content for real time review and action. *Id.* In 2023, user engagement—and therefore platform review and moderation—has increased significantly. During a 20-day period in the third quarter of 2023, TikTok reported processing 34,038,045 notices (over 1.7 million per day) related to removal or restriction of user content in EU jurisdictions. *DSA Transparency Database: Analytics, TikTok*, EUROPEAN COMMISSION <https://transparency.dsa.ec.europa.eu/analytics> (as of December 3, 2023) (“*DSA Transparency*”). Analyzing the grounds for these content moderation decisions over 90 days in 2023, less than 1% were due to illegal content; 99% were coded as “Content incompatible with terms and conditions.” *Id.* (as of December 3, 2023, reporting 1,399,314 of 655,363,603 total notices as concerning “illegal content;” the remainder “incompatible with terms and conditions”). Although filters used to remove duplicate content aid platforms, like TikTok, in enforcing rules and issuing notices, the demand on platforms is unrelenting.

Moreover, these numbers reported in compliance with the EU’s DSA do not capture the scale required under Texas’s and Florida’s laws. *See DSA Transparency, supra.* Unlike Texas’s and Florida’s laws, the DSA does not require, for example, platforms like YouTube to report on content moderation decisions made related to user comments. With this added burden, the number of decisions processed per day would certainly be much higher. It

is also unclear whether notices of the sort issued using automated systems under EU legal standards would meet requirements like Florida's for a "precise and thorough" explanation of how the platform became aware of content, what role algorithms played, and on what basis the platform took action. Fla. Stat. § 501.2041(3); *see also* Daphne Keller, *Platform Transparency and the First Amendment*, 4 J. Free Speech L. 1 (2023), <https://journaloffreespeechlaw.org/keller2.pdf> ("Platform Transparency").

As a result of this growth, large social media websites with millions or even billions of users have built large T&S teams to review the huge volumes of potentially violative content. Google reported in 2021 that its T&S team comprised nearly 22,000 employees. Google Submission in Response to Subcommittee Questions for the Record, *Hearing on Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation, Subcomm. On Commc'n & Tech. and the Subcomm on Consumer Prot. & Com. of the H. Comm. On Energy & Com.*, 117th Cong. (2021), at 5, <https://bit.ly/3He9sMf>. For its part, Facebook reported in 2021 that it employed 15,000 content moderators. Facebook Submission in Response to Subcommittee Questions for the Record, *Hearing on Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation, Subcomm. On Commc'n & Tech. and the Subcomm on Consumer Prot. & Com. of the H. Comm. On Energy & Com.*, 117th Cong. (2021), at 39, <https://bit.ly/3QU0R4w>.

By contrast, smaller platforms may rely on a limited number of T&S professionals. And topical

communities that exist within platforms, such as subreddits on Reddit and Groups on Facebook, rely on volunteer moderators. Hanlin Li, et al., *Measuring the Monetary Value of Online Volunteer Work*, 16 PROCEEDINGS OF THE SIXTEENTH INT’L AAAI CONF. ON WEB & SOCIAL MEDIA (2022), at 596, <https://aaai.org/papers/00596-19318-measuring-the-monetary-value-of-online-volunteer-work/> (“[S]ocial media platforms such as Facebook Groups, Reddit, and Discord prominently depend on fleets of volunteer moderators to build and manage communities with millions of users and, thereby, keep these platforms viable.”). These volunteers perform the content moderation tasks essential for the success of these communities, including setting community guidelines, admitting users to participate, and managing what content stays online. *Id.* at 597.

To manage the ever-increasing volume of content review for safety and legal compliance, many platforms employ automated tools and machine learning, which—although imperfect—can serve as an effective first line of defense to proactively detect policy-violating content and more efficiently route content for human review. Maggie Engler, Jeff Lazarus & James Gresham, *Automated Systems and Artificial Intelligence*, TR. & SAFETY PROF. ASS’N, <https://www.tspa.org/curriculum/ts-fundamentals/automated-systems-and-ai/uses>. Automation technologies improve efficiency in content moderation in three primary ways: (1) by reducing the number of human reviewers needed to make a decision; (2) by routing content or accounts for review to the most appropriate human reviewer (*e.g.*, reviewers with appropriate language skills or subject

matter expertise); and (3) by prioritizing reviews in the most impactful way possible. *Id.* Indeed, Meta reports that it relies on AI tools, not user reports, to identify 87.8% of bullying and harassing content that removes or restricts. *Community Standards Enforcement Report: Q3 2023 report*, META (2023), <https://transparency.fb.com/reports/community-standards-enforcement/> (“*Meta Enforcement Report 2023*”).

Because there is always a degree of subjectivity in any type of content moderation decision, including human ones that train AI, the process is not perfect, much in the same way that editorial decisions by newspapers as print publishers are imperfect. Unlike traditional publishers, however, online platforms must make split-second decisions about individual posts or respond to developments ranging from evolving local slang to natural disasters or acts of terrorism. Beyond serving a public good (as with promoting safety information during natural disasters) or advancing platform operations’ editorial policies, content moderation establishes goodwill with users. Policy differences—such as promoting child-friendly environments—identify and distinguish one platform from another and serve different user needs. Further, many platforms’ business models rely on subscriptions and advertising to attract and retain both user base and revenue. Content moderation ensures desired traffic to a platform by directly influencing user trust and brand reputation, which in turn drives growth and revenue. Overall, concrete benefits in efficiency, consistency, and accuracy are realized by the collaborative use of automated processes alongside human review, which contribute

to successful platforms that facilitate dynamic user speech.

II. THE REGULATORY CONSTRAINTS ON CONTENT MODERATION IMPOSED BY TEXAS AND FLORIDA LIKELY WILL YIELD UNINTENDED CONSEQUENCES THAT SUPPRESS SPEECH.

The premise supporting Texas's and Florida's new laws is that content moderation by platforms is bad and that editorial decisions by platforms to publish, edit, and delete material should be circumscribed. More accurately understood, Point I., *supra*, these processes facilitate speech and assist users by connecting them with content that aligns with their subjective interests. By precluding or disincentivizing platforms from performing these critical moderation functions, online communities will lose their distinctive identities and the usefulness of these platforms will decline.

By imposing mandatory user notice and appeal requirements, along with consistency or viewpoint-neutrality mandates, Texas and Florida burden platforms with skyrocketing compliance and litigation costs. Litigation is particularly likely given Florida's \$100,000 bounty for users who bring claims, Fla. Stat. § 501.2041(6)(a), and Texas's law does little to deter frivolous claims where an award of fees could be given even to plaintiffs bringing claims already rejected, or enforcing legal provisions already deemed unconstitutional, by other Texas courts, Tex. H.B. § 143.007(b)(1), (d)–(e). While T&S professionals are quite skilled, they typically aren't trained lawyers. When faced with the choice of tasking their T&S team members to individually assess the local legal merits

of each claim, or simply instructing their T&S teams to remove such content as a whole, platforms will not hesitate to choose the simpler option. History teaches the most likely path these platforms will choose is the wholesale removal of any content that even remotely expands risk of liability. *E.g.*, Jennifer M. Urban et al., *Notice and Takedown in Everyday Practice*, UC BERKLEY L. RES. PAPER NO. 2755628 (Mar. 22, 2017), <http://dx.doi.org/10.2139/ssrn.2755628> (noting online service providers' concern that "transparency reporting and public archiving could trigger negative attention from rights enforcement groups, exposing them to high-volume sending or even litigation."). This is particularly true for smaller companies who may simply not have the budget to adopt a policy other than "if in doubt, take it down." Daphne Keller, *Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List*, STANFORD L. SCH.: CTR. FOR INTERNET & SOC'Y (Feb. 8, 2021), <https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws/> ("The easiest, cheapest, and most risk-avoidant path for any [platform] is simply to process a removal request and not question its validity.").

Pairing the economic realities of reduced T&S teams and resources with the uptick in user-generated content, most platforms likely will lack the resources to take on the impossible task of ensuring consistency or viewpoint-neutrality across the tremendous range of ideas and topics addressed in users' posts, or to thoroughly explain each decision to users. Even the largest platforms will need to weigh the hard costs of

continuing to enforce nuanced speech rules or even host some speech at all, given the increased costs of explaining each decision and fighting continued challenges from those who claim enforcement is inconsistent or biased.

A. Increased Costs Associated with Exercising Editorial Control Likely Will Result in Policy and Enforcement Changes That Suppress Speech.

Both Texas’s and Florida’s laws mandate that platforms notify users affected by all content moderation decisions and provide a detailed, individualized explanation. Tex. H.B. 20 §§101–104; Fla. Stat. § 501.204(2)(d) (requiring notice), § 501.2051(3) (outlining details of notice). Texas’s law requires platforms to develop a complaint system that is “easily accessible” for users to submit and track complaints; new procedures for processing of complaints; detailed explanations related to a platform’s editorial decisions to permit or remove content; and appeal procedures to be carried out within fourteen days (*see* Point II.B, *infra*). H.B. 20 §§101–104. Florida’s law, while also mandating highly specific notice within seven days, also grants users a private right of action through which platforms that issue any notice deemed insufficiently “thorough” or “precise” may incur liability for up to \$100,000 in statutory damages. Fla. Stat. § 501.2041(3)(b) (seven-day rule), § 501.2041(6)(a) (cause of action). Here again, many platforms will find it more practical to simply provide customers fewer opportunities to express themselves online, rather than incur this business risk thousands of times a day.

Given the design and functioning of typical content moderation systems, these rules will generate billions of individualized notifications every year. Many of these notifications will provide information that users don't really care about or which they may find confusing or intrusive. The droves of people who posted restaurant reviews on Yelp many years ago, for example, likely do not want to receive notice if the platform decides to demote older reviews or put them in a less viable place. Constant and unwelcome emails announcing such changes would contribute to user fatigue, likely causing some users to tune out notifications entirely. As a whole, this will reduce users' ability to notice when something meaningful does happen. It also will harm platforms as users reflexively appeal each notified decision—generating wasteful additional work—or simply delete their accounts in response to excessive state-mandated messaging.

Faced with the documentation burdens resulting from mandatory notice obligations, platforms must weigh the greatly increased cost of exercising editorial control. Companies and organizations that curate speech on their platforms will need to vastly expand internal processes in order to thoroughly memorialize each action in notices to users in Texas and Florida and be prepared to defend each during Texas's mandatory appeals processes or in court. At scale, the costs as platforms seek to protect themselves from liability will be exorbitant. T&S teams would quickly be overwhelmed by the daunting task of spelling out the basis for every action. While this may conceivably be an option for the largest social media platforms—the stated targets of the laws—it changes the

incentives to offer speech-supportive features and enforce editorial policies, even for them.

Platforms unable to bear the steep cost associated with blanket review of every piece of user-generated content will face essentially three options to reduce the financial burden. First, platforms may choose to simply moderate less—the result lawmakers in Texas and Florida are hoping for. *Platform Transparency* at 31. Even that may backfire, of course, if platforms decide to simply terminate or “deplatform” individual users after a single violation, instead of bearing the notice and appeal cost for multiple violations before doing so. Second, they may adopt simpler, blunter policies that will more easily generate the required individualized notices and be less prone to dispute on appeal. For example, platforms might prohibit all racial slurs instead of having special rules for slurs appearing in historical documents, popular culture, or news reporting. Third, they may stop offering features that enable speech, including user comments and topical communities. All paths lead to the same destination: forfeiture of editorial control.

Current practice suggests that T&S teams are likely to employ stricter policies and enforcement to ensure consistency, removing content to avoid litigation and potential liability. This is the approach platforms already take with respect to intellectual property, where many err on the side of removal regardless of the lawfulness of the materials to avoid costs associated with litigation. T&S teams employ a “notice and takedown” practice in which they review legal claims of ownership—which may be false or fraudulent—and attempt to decide which user speech should be silenced. This scheme is burdensome for

reviewers and ripe for abuse by competitors and others posing as legitimate rights owners. Glyn Moody, *Copyright as Censorship: Abuse of the DMCA To Try to Delete Online News Is Rampant*, TECHDIRT (May 24, 2022), <https://www.techdirt.com/2022/05/24/copyright-as-censorship-abuse-of-the-dmca-to-try-to-delete-online-news-is-rampant/> (“This [Digital Millennium Copyright Act] ‘notice and takedown’ system allows the copyright industry to . . . demand[] that [infringing content] should be taken down. . . . This unbalanced nature of the system makes it ripe for fraud, whereby people falsely claim to be the owner of copyright material in order to get it removed from a Web site. . . . An entire business sector, called ‘reputation management,’ has sprung up to offer this kind of service.”).

If T&S teams are required to evaluate the individual legal risks of moderating each piece of content and articulate to the user the basis for any content moderation decision, there is a tremendous incentive to default to removing entire categories of content representing all possible “viewpoints” on particular issues. T&S teams (or AI tools) likely cannot make sufficiently consistent decisions quickly enough to enforce more nuanced speech policies. In practice, platforms will be less willing to host even *potentially* controversial user speech and may wholesale preclude discussion of important, yet sensitive topics like race, gender, sexual identity, politics, and healthcare, without regard to context.

The difficulty of defining and enforcing viewpoint neutral rules across all possible areas of disagreement between online speakers has already been illustrated.

In May 2022, a plaintiff sued Reddit under Texas’s H.B. 20 for a volunteer moderator’s decision to ban a user for calling the fictional Star Trek character Wesley Crusher a “soy boy” in the r/StarTrek subreddit. Petition: Small Claims Case, *Cox v. Reddit, Inc.*, No. S22-87J1 (Just. Ct. Denton Cnty., Tex. May 17, 2022) (“plaintiff claims to have been banned and/or deplatformed from r/StarTrek for posting a lawful opinion about a fictional character”). Faced with increased liability, platforms may lose their appetite for hosting not only politically controversial speech, but also user posts that while offensive may otherwise be lawful. In effect, Florida’s and Texas’s mandates encourage platforms to curtail opportunities for expression.

Development and implementation of new practices required for transparency reporting is a costly endeavor. It is unclear whether platforms will be willing or able to absorb the cost associated with the even broader notice requirements contemplated by Texas and Florida. For example, Texas’s requirements for mandatory user explanations are not the same as those required for DSA compliance.³ Thus, even a platform that has invested in the processes to fully comply with the DSA will be

³ The DSA requires a statement of reasons for content that is removed or restricted only. These statements must be “clear and specific,” including information on the type of restriction(s) imposed, territorial scope, and duration of the restriction; facts and circumstances relied on in making the decision; information on the use made of automated means; and legal or contractual grounds relied on in taking the decision. *DSA Transparency Database: Submission of clear and specific statements*, EUROPEAN COMMISSION
<https://transparency.dsa.ec.europa.eu/page/documentation>.

burdened with designing and implementing a new set of processes to meet Texas's and Florida's requirements, which also must operate in harmony with a platform's current scheme. Such expenses are entirely avoided if a platform chooses to limit the speech it hosts.

The United States should not emulate the EU in imposing detailed notice requirements, even if those rules only apply to larger platforms here. European law and policy may permit imposing burdens that will reshape platforms' editorial practices, but the First Amendment holds states to a higher standard. By the same token, European policymakers may be more willing to quash the next generation of startups, since the EU economy has seen few successful ones compared to the United States. Newer small and midsize social media platforms here, however, are already straining to compete with titans in the space. Startups that hope to one day become larger—which is to say, almost all of them—will need to plan for requirements like those in Texas and Florida now and acquire additional investment to meet the states' legal requirements. This may disincentivize these independent—and often diverse—platforms from entering the market in the first place, continuing to host user-generated content, or push them to cease competing with larger platforms flush with resources and waiting in the wings with an acquisition offer. Such absorption by larger platforms will naturally result in fewer opportunities for user speech.

B. Mandatory Notice and Appeal Requirements Incentivize Platforms to Reduce Opportunities for Online Speech and Encourage Consolidation of Existing Platforms.

No decision process for content moderation yields perfect results. Today, many platforms provide some avenue for users to challenge content moderation decisions if they believe the decision to be incorrect, unfair, or not aligned with their expectations. Appeals serve multiple purposes in T&S, including correction of erroneous decisions, serving as an avenue for users affected by decisions to have their voices heard, and giving useful insights into users' knowledge and understanding about a platform's policies, which may drive revisions and improvement.

The types of appeals that platforms generally offer vary in process and formality. The most common type is the "responsible actor appeal." Harsha Bhatlapenumarthy & James Gresham, *User Appeals*, TR. & SAFETY PROF. ASS'N, <https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/user-appeals/> ("*User Appeals*"). This is when a piece of content, account, or other entity is restricted or removed by a platform and the person responsible for the content seeks reconsideration of the decision. Some platforms also permit "reporter appeals," where the person who reported potentially violative content objects to the decision made—most commonly when no action or alternative action to removal is taken (e.g., when content is labeled as disturbing but retained). Platforms may also offer more informal forums for user objection like

disagreement reports, which provide valuable information on a user's disagreement with a content moderation decision but do not automatically lead to an appeal.

Appeals are initiated through a variety of different communication methods. For some platforms, the first source of appeals is a simple messaging system, like a Contact Us or Help page containing a form or email address. Large platforms with a significant number of users and the commensurate volume of user-generated content inevitably construct more sophisticated appeals processes. Simply put, the most cost-effective appeals strategies, like a direct messaging system managed by manual review, quickly become impractical.

Most midsize and large platforms thus introduce dedicated appeals channels for registering appeals and may employ automation tools. *E.g., Help Center: I Don't Think Facebook Should Have Taken Down My Post*, FACEBOOK, https://www.facebook.com/help/2090856331203011?helpref=faq_content; *Help Center: Submit an Appeal*, PINTEREST, <https://help.pinterest.com/en/article/review-your-reports-and-violations>. Use of automation in appeals, however, is often limited to prioritization, decisions with the lowest user impact, and rejections of users who have abused the system to preserve the safety net function that appeals serve in reviewing initial moderation decisions for error.

Consistency is of paramount importance in appeals because if discrepancies exist between appeals and moderation, penalties can be repeatedly restored and

revoked as the two processes disagree. *User Appeals: Processing Appeals*, TR. & SAFETY PROF. ASS'N, <https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/user-appeals/>. To foster consistency, appeal reviews often mimic the original review as closely as possible, with little to no additional information about the violation. Although this certainly increases consistency between the processes, it also increases the likelihood the same error that resulted in the incorrect moderation decision will plague the appeal.

More robust appeals systems that introduce additional information and processes like context from users, history of the appellant, and policy changes or problems that may justify reevaluation alleviate this concern. Maintaining such an appeals system, however, can require significant additional resources for development, implementation, and maintenance.

Texas's law creates an automatic right for users to appeal platform content moderation decisions and mandates that platforms provide individualized justification for moderation decisions. The law also requires platforms to build an "easily accessible complaint system to enable a user to submit a complaint in good faith and track the status" of appeals concerning illegal content or activity and any decision that results in removal of content posted by the user. Tex. H.B. 20 §120.101.

At first blush, it may appear that certain platforms' existing appeal processes already satisfy the requirements of the law. But even for the largest

platforms, deeper analysis suggests otherwise. For example, in the second quarter of 2022, Meta updated its appeals methodology for content on Facebook and Instagram “to account for all instances where content was submitted for additional review,” including when users disagreed with content moderation decisions. *Corrections and Adjustments: Updated Methodology for Appealed Content on Facebook and Instagram*, META (August 2022), <https://transparency.fb.com/policies/improving/corrections-adjustments>. Meta’s appeals process, however, still excludes instances where content is not submitted for additional review—as in many cases of spam—even if users report their disagreement with a decision.

It is also true that large platforms, including Meta, recently have updated their policies to fulfill additional DSA transparency and appeal requirements for EU jurisdictions. The result was an immediate meaningful increase in the number of appeals initiated. Instagram, for example, saw an increase in appealed content on Bullying and Harassment from 874,000 appeals in quarter two 2023 to a staggering 1.6 million appeals in quarter three. *Meta Enforcement Report 2023*, , <https://transparency.fb.com/reports/community-standards-enforcement/>; *Integrity Reports, Third Quarter 2023*, META, <https://transparency.fb.com/integrity-reports-q3-2023>. Significant increases in appealed content on Hate Speech were also observed, with appealed content increasing from 625,000 in Q2 2023 to 977,000 in Q3 2023. *Id.*

This surge of new appeals undoubtedly consumes

additional resources and, depending on the scale and sophistication of the platform, compliance will naturally require companies to either expand their appeals teams, invest in further automation and processes, or risk liability for inability to keep up. With the introduction of appeal mandates for new jurisdictions like Texas with detailed explanation requirements, the number of appeals platforms must process will only continue to balloon.

Because social media companies are businesses, funds allocated to other areas may be diverted to cover new costs associated with the influx of new appeals. This may implicitly encourage platforms to revisit policies to limit exposure and conserve resources, or to redistribute resources away from important safety functions addressing hate speech, terrorism, or cyber security assessment and management. It is not difficult to predict the types of policy changes likely to result. For example, to alleviate the burden of Texas's mandatory appeal requirement, platforms may introduce more punitive strike policies to avoid endless expensive appeals. Where previously a user might have been given 20 chances to correct an abusive behavior before their account is permanently terminated, they might now only receive two.

This is particularly problematic for midsize and smaller platforms that have less robust policies and procedures and less access to resources. Although Texas has advocated that its law will impact only the industry giants YouTube, Facebook, and Twitter, the law, as written, covers any platform with 50 million monthly active users in the United States. Under that metric, the law is likely to reach sites like Reddit, Quora, Skype, Rumble, LinkedIn, Picsart, Pinterest,

Discord, Twitch, Stack Exchange, Wikipedia, Glassdoor, Vimeo, Steam, and Minecraft. See *Platform Transparency* at 35; accord *Netchoice, LLC v. Paxton*, 573 F. Supp. 3d 1092, 1099 (W.D. Tex. 1099) (listing platforms and applications covered by Tex. H.B. 20); Daphne Keller, *Reported Monthly Active Usage Data for Content Hosting Platforms*, at 15 (Apr. 29, 2022), <https://www.judiciary.senate.gov/imo/media/doc/Keller%20Testimony1.pdf> (listing reported user data across platforms). Smaller platforms have a fraction of the resources of the triad social media giants and will feel a disparate impact if forced to shoulder new mandatory appeal obligations. For many, these obligations may simply be too prohibitive to continue operations.

C. Sweeping Carriage Obligations Undermine Platform Functionality and Contravene User Desirability.

Proponents of so-called “must carry” laws, like those in Texas and Florida, purport to promote platform neutrality as the solution to what they claim are threats to free speech online through moderation. Florida’s law requires platforms forego automated moderation of all speech posted by a political candidate—regardless of its content or character. Fla. Stat. § 2041(h). Texas, for its part, mandates that platforms’ policies on hate speech, disinformation, and other topics be neutral as to the speakers’ viewpoint. Tex. H.B. 20 §143A.002.

Hypothetical examples illustrate the absurdity of these rules. A child-focused social media platform could, for example, be required to carry a candidate

affiliated with the “North American Man/Boy Love Association,” a pedophilia and pederasty advocacy organization which works to abolish age-of-consent laws. North America Man/Boy Love Association, WIKIPEDIA, https://en.wikipedia.org/wiki/North_American_Man/Boy_Love_Association. Likewise, a conservative platform like Donald Trump’s Truth Social could be required to host content from Alexandria Ocasio-Cortez. The laws on their face would also require platforms to carry medical disinformation and conspiracy theories that lead to violence, *see* Emmet Lyons, *Elon Musk’s X Platform Fueled Far-Right Riots in Ireland, Experts Say*, CBS NEWS (Dec. 5, 2023), <https://www.cbsnews.com/news/elon-musk-ireland-x-twitter-far-right-dublin-immigration/>, despite their brand objectives and initiatives. This is also true for sexually suggestive material and other content that in a free market of ideas some services might want to carry but others would not.

A neutral version of the Internet disregards most users’ preferences for at least some level of content moderation that treats content differently based on its viewpoint. *E.g.*, Shubham Atreja *et al.*, *Remove, Reduce, Inform: What Actions Do People Want Social Media Platforms to Take on Potentially Misleading Content*, at 1 (Oct. 2023), <https://bit.ly/3jTauUo> (“We find a clear hierarchy of perceived severity of actions with a majority of raters wanting informational labels on the most articles and removal on the fewest. There was no partisan difference in terms of how many articles deserve platform actions but conservatives did prefer somewhat more action on content from liberal sources, and vice versa.”); *Americans Support*

Free Speech Online but Want More Action to Curb Harmful Content, KNIGHT FDN. (June 16, 2020), <https://knightfoundation.org/press/releases/americans-support-free-speech-online-but-want-more-action-to-curb-harmful-content> (“Most people support the removal of false or misleading health information from social media. Amid the pandemic, 85% of Americans are in favor of this, and 81% support removing intentionally misleading information on elections or other political issues.”).

So-called viewpoint-neutral platform requirements thus would be anything but neutral in practice. For example, it is routine for online community leaders to create digital communities around particular points of view, which encourages free and unfettered expression by community members to say what they really think. By forcing mandatory viewpoint neutrality on platforms, these topical online communities will effectively be eliminated.

Users are consumers who come to Internet websites to find a particular experience and interact with a particular set of people. By analogy, the government would not compel a home improvement store to carry a full inventory of makeup and tropical fish supplies. Customers should not have to slog through aisles of eyeshadow and fin rot medications to be able to locate the drywall anchors they came in for. No customer would patronize such an inconvenient home improvement store.

Applied to the digital landscape, imagine having to carry pro-San Francisco 49ers content in a Miami Dolphins football fan group. Internet platforms won't wait for a disgruntled 49ers fan to sue them before

removing a pro-49ers post in the Dolphins fan forum; they will simply stop offering these forums, which will have lost their appeal and utility anyway.

CONCLUSION

For these reasons, the Court should reverse the Fifth Circuit's judgment and affirm the Eleventh Circuit's judgment as to the issues before the Court.

Respectfully Submitted,

Ian C. Ballon
Counsel of Record
Lori Chang
GREENBERG TRAUIG, LLP
1840 Century Park East,
Suite 1900
Los Angeles, CA 90067
(310) 586-7700
ballon@gtlaw.com

Bethany J. M. Pandher
GREENBERG TRAUIG, P.A.
333 S.E. 2nd Avenue
Miami, FL 33131

Counsel for Amicus Curiae

December 7, 2023