

No. 21-1333

IN THE
Supreme Court of the United States

REYNALDO GONZALEZ, *et al.*,
Petitioners,

v.

GOOGLE LLC,
Respondent.

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

**BRIEF OF ANTI-DEFAMATION LEAGUE
AS *AMICUS CURIAE*
IN SUPPORT OF NEITHER PARTY**

STEVEN M. FREEMAN
Counsel of Record
EILEEN B. HERSHENOV
LAUREN T. KRAPF
ANTI-DEFAMATION LEAGUE
605 Third Avenue
New York, NY 10158
(212) 885-7700
sfreeman@adl.org

Counsel for Amicus Curiae

December 7, 2022

TABLE OF CONTENTS

	Page
TABLE OF AUTHORITIES.....	iii
STATEMENT OF INTEREST	1
INTRODUCTION AND SUMMARY OF ARGUMENT	3
ARGUMENT.....	8
I. SOCIAL MEDIA PLATFORMS FOMENT HATE AND EXTREMISM, RADICAL- IZE INDIVIDUALS, AND CONNECT THOSE RADICALIZED INDIVIDUALS TO NEW COMMUNITIES.....	8
A. Platforms actively support and grow new extremist communities by amplify- ing hateful and extremist content, recommending it to users, and recom- mending users join groups that engage in extreme and conspiratorial rhetoric.....	8
B. Platforms fomenting hate online play a role in facilitating offline violence, which disproportionately impacts vulnerable communities.....	11
C. Platforms are aware that their con- duct produces harmful consequences, yet they take little meaningful action	14
II. THE COURT SHOULD MAKE CLEAR THAT § 230(c)(1) DOES NOT IMMUN- IZE A PLATFORM FROM THE CONSE- QUENCES OF ITS OWN TORTIOUS CONDUCT	17

TABLE OF CONTENTS—Continued

	Page
III. THE PLAIN TEXT OF § 230(c)(2) PROTECTS A PLATFORM'S POWER TO EXERCISE EDITORIAL JUDGMENT ...	20
CONCLUSION	24

TABLE OF AUTHORITIES

CASES	Page(s)
<i>Force v. Facebook</i> , 934 F. 3d 53 (2d Cir. 2019)	<i>passim</i>
<i>FTC v. LeadClick Media, LLC</i> , 838 F. 3d 158 (2d Cir. 2016)	7, 23
<i>Manhattan Cmty. Access Corp. v. Halleck</i> , 204 L. Ed. 2d 405, 139 S. Ct. 1921 (2019).....	23
<i>Miami Herald Publishing Co. v. Tornillo</i> , 418 U.S. 241, 94 S. Ct. 2831 (1974),.....	23
CONSTITUTION	
U.S. Const. amend. I	23
STATUTES	
47 U.S.C. § 230	<i>passim</i>
47 U.S.C. § 230(c)(1).....	<i>passim</i>
47 U.S.C. § 230(c)(2).....	<i>passim</i>
47 U.S.C. § 230(c)(2)(A).....	22
OTHER AUTHORITIES	
<i>After Long Fight, ADL is Relieved at Facebook Announcement That It Will Remove Holo- caust Denial Content</i> , Anti-Defamation League (Oct. 12, 2020), https://www.adl.org/news/press-releases/after-long-fight-adl-is-relieved-at-facebook-announceme-nt-that-it-will-remove	21

TABLE OF AUTHORITIES—Continued

	Page(s)
Brandy Zadrozny, ‘Carol’s Journey’: What Facebook knew about how it radicalized users, NBC NEWS (Oct. 22, 2021), https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rcna3581.....	15
<i>Domestic Terrorism and Violent Extremism: Examining the Threat of Racially, Ethnically, Religiously, and Politically Motivated Attacks, Part II Before the S. Comm. on Homeland Security and Government Affairs</i> , 117 Cong. 6 (2021) (statement of Jonathan Greenblatt, CEO & National Director, Anti-Defamation League), https://www.adl.org/sites/default/files/adl-testimony-hsgac-domestic-terrorism-violent-extremism-2021-08-05.pdf	8, 11, 12
<i>Exposure to Alternative and Extremist Content on YouTube</i> , ANTI-DEFAMATION LEAGUE CENTER FOR TECHNOLOGY & SOCIETY (May 3, 2022), https://www.adl.org/sites/default/files/pdfs/2022-05/FINAL_FINAL_ADL-Report-Single-Final-Design.pdf	9, 10, 21
FBI, Anti-Government, Identity Based, and Fringe Political Conspiracy Theories Very Likely Motivate Some Domestic Extremists to Commit Criminal, Sometimes Violent Activity (May 30, 2019), https://www.justsecurity.org/wp-content/uploads/2019/08/420379775-fbi-conspiracy-theories-domestic-extremism.pdf	13

TABLE OF AUTHORITIES—Continued

	Page(s)
<i>FBI informant: Facebook led me to infiltrate plot to kidnap Gretchen Whitmer</i> , Bridge Michigan (Mar. 5, 2021), https://www.bridgemi.com/michigan-government/fbi-informant-facebook-led-me-infiltrate-plot-kidnap-gretchen-whitmer	12
<i>Holding Big Tech Accountable: Legislation to Build a Safer Internet Before the H. Comm. on Energy and Commerce Subcomm. on Consumer Protection and Commerce</i> , 117 Cong. (2021) (Statement of Jonathan Greenblatt, CEO & National Director, Anti-Defamation League), https://www.adl.org/sites/default/files/adl-testimony-house-cmte-energy-commerce-holding-big-tech-accountable-2021-12-09.pdf	14, 15
Jeff Horowitz & Deepa Seetharam, <i>Facebook Executives Shut Efforts to Make the Site Less Divisive</i> , WALL ST. J. (May 26, 2020), https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499	15
Jonathan E. Bromwich, <i>Before Massacre Began, Suspect Invited Others to Review his Plan</i> , N.Y. TIMES (May 17, 2022), https://www.nytimes.com/2022/05/17/nyregion/buffalo-shooting-discord-chat-plans.html	12

TABLE OF AUTHORITIES—Continued

	Page(s)
Letter dated Jan. 20, 2020 from the Chair of the Security Council Committee pursuant to resolutions 1267 (1999), 1989 (2011) and 2253 (2015) concerning Islamic State in Iraq and the Levant (Da'esh), Al-Qaida and associated individuals, groups, undertakings and entities addressed to the President of the Security Council, U.N. Doc. S/2020/53 (Jan. 20, 2020).....	12
Max Fisher, <i>THE CHAOS MACHINE: THE INSIDE STORY OF HOW SOCIAL MEDIA REWIRED OUR MINDS AND OUR WORLD</i> (2022).....	11
Newly Purnell & Jeff Horowitz, <i>Facebook Services Are Used to Spread Religious Hatred in India, Internal Documents Show</i> , WALL ST. J. (Oct. 23, 2021), https://www.wsj.com/articles/facebook-services-are-used-to-spread-religious-hatred-in-india-internal-documents-show-11635016354	16
<i>One Year After Ban, Holocaust Denial Remains on Facebook</i> , ADL (Dec. 1, 2021), https://www.adl.org/blog/one-year-after-ban-holocaust-denial-remains-on-facebook	22

TABLE OF AUTHORITIES—Continued

	Page(s)
<i>Online Hate and Harassment: The American Experience 2022</i> , ANTI-DEFAMATION LEAGUE (June 20, 2022), https://www.adl.org/sites/default/files/pdfs/2022-09/Online-Hate-and-Harassment-Survey-2022.pdf	13, 22
<i>Online Hate and Harassment: The American Experience 2021</i> , small caps: ANTI-DEFAMATION LEAGUE (Mar. 3, 2022), https://www.adl.org/sites/default/files/pdfs/2022-05/Online%20Hate%20and%20Harassment%202021_042821_V17%282%29.pdf	14
Paul Lewis, <i>'Fiction is outperforming reality': how YouTube's algorithm distorts truth</i> , GUARDIAN (Feb. 2, 2018), https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth...	9
Paul Mozur, <i>A Genocide Incited on Facebook, With Posts From Myanmar's Military</i> , N.Y. TIMES (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html	16
Rebecca Tushnet, <i>Power Without Responsibility: Intermediaries and the First Amendment</i> 76 GEO. WASH. L. REV. 986 (2008).....	17

TABLE OF AUTHORITIES—Continued

	Page(s)
Sergii Denysenko, <i>Key Considerations For Ensuring Brand Safety</i> , FORBES (Mar. 20, 2022), https://www.forbes.com/sites/forbesagencycouncil/2022/03/30/key-considerations-for-ensuring-brand-safety/?sh=3f735e2d2000	20
<i>State of Social Media Regulation: Misinformation, Exploitation, Harassment, and Radicalization Before the S. Comm. on Judiciary</i> , 2021 Leg. (Cal. 2021) (statement of Lauren T. Krapf, ADL Technology Policy & Advocacy Counsel), https://sjud.senate.ca.gov/sites/sjud.senate.ca.gov/files/lauren_krapf_testimony_-_adl.pdf	20
<i>The ISIS Impact on the Domestic Islamic Extremist Threat</i> , ANTI-DEFAMATION LEAGUE (MAR. 22, 2016), https://www.adl.org/sites/default/files/documents/assets/pdf/combating-hate/CR_4473_HomegrownExtremismReport-2009-2015_web2.pdf	12
<i>The Rate of Hate: Measuring Antisemitism Prevalence and Enforcement on Reddit and Twitter</i> , ANTI-DEFAMATION LEAGUE CENTER FOR TECHNOLOGY & SOCIETY (MAY 3, 2022), https://www.adl.org/sites/default/files/pdfs/2022-05/How%20Platforms%20Rate%20on%20Hate%202022_OH_I_V10.pdf	11, 22

TABLE OF AUTHORITIES—Continued

	Page(s)
<i>YouTube’s Filter Bubble Problem is Worse for Fox News Viewers</i> , Tech Transparency Project (Oct. 24, 2021), https://www.techtransparencyproject.org/articles/youtubes-filter-bubble-problem-worse-fox-news-viewers	18
<i>YouTube Regrets</i> , MOZILLA (July 2021), https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf	10

STATEMENT OF INTEREST¹

The Anti-Defamation League (“ADL”) submits this amicus curiae brief in support of neither party. Founded in 1913 in response to an escalating climate of antisemitism and bigotry, ADL’s timeless mission is to stop the defamation of the Jewish people and to secure justice and fair treatment to all. ADL continues to fight all forms of bigotry and hate with the same vigor and passion and is often the first call when acts of antisemitism occur. A recognized leader in exposing extremism, delivering anti-bias education, and fighting hate online, ADL’s ultimate goal is a world in which no group or individual suffers from bias, discrimination, or hate.

ADL believes that fighting hate online is central to robustly pursuing its mission of fighting antisemitism and securing justice and fair treatment for all. ADL has been devoting resources and attention to the issue of hate online since the early days of dial-up.

Today, ADL brings decades of experience and expertise to the fight against hate and extremism online. Its Center on Extremism (“COE”) examines the ways extremists across the ideological spectrum exploit the online ecosystem to spread their messages, recruit adherents, finance hate, and commit acts of terrorism. COE works directly with threatened communities on the ground as well as law enforcement agencies across the country to prevent attacks and identify perpetra-

¹ No counsel for a party authored any part of this brief and no counsel or party made a monetary contribution intended to fund the preparation or submission of the brief. Only the amicus have paid for the filing and submission of this brief. Pursuant to Rule 37.3(a), all parties have granted blanket consent to the filing of amicus curiae briefs.

tors. ADL’s Center for Technology and Society (“CTS”) works across four key areas—policy, research, advocacy, and incident response—to generate advocacy-focused solutions to make digital spaces safer and more equitable. For years, CTS has researched how platforms amplify hate and extremism through their user interfaces, recommendation engines, and algorithms. CTS also has years of experience researching the consequences of that amplification, including the radicalization of users and the spread of online hate to offline violence. CTS engages directly and regularly with major social media platforms, as well as with legislators and regulators, to push for policy and product changes, making a measurable difference in fighting online extremism.

ADL believes that several of the lower courts—particularly the Ninth and Second Circuits—have interpreted § 230(c)(1) far beyond its text and original legislative intent and too broadly as a matter of policy. *See* 47 U.S.C. § 230(c)(1) (providing that “[n]o provider or user of an interactive computer services shall be treated as the publisher or speaker of any information provided by another information content provider”). Congress never intended § 230(c)(1) to shield platforms from any possibility of accountability for amplifying hate and extremism, least of all where those platforms acted with recklessness or intentionality, resulting in injury. The lower courts’ interpretation of § 230(c)(2), however, is an important bulwark protecting platforms from liability for moderating hateful and extremist content and should be affirmed. *See* 47 U.S.C. § 230(c)(2) (granting immunity for “any action voluntarily taken in good faith to restrict access to or availability of material. . . whether or not such material is constitutionally protected.”).

ADL respectfully asks this Court to make clear that a social media platform is not immunized for its own conduct under § 230(c)(1) merely because that conduct also involves third-party speech. At the same time, however, this Court should affirm that immunity under § 230(c)(2) necessarily protects platforms when they exercise their editorial discretion by publishing and moderating third-party content.

INTRODUCTION AND SUMMARY OF ARGUMENT

The past several years have witnessed a tectonic shift in the proliferation of hate online and the destabilizing and violent consequences of that proliferation, both online and off. It is by now well-established that social media platforms pushing volume and virality in service of their bottom lines endanger vulnerable communities most at risk of online harassment and related offline violence.

These platforms are not merely holding up a mirror that reflects society's existing state of divisiveness and polarization. Rather, the companies that run these platforms deliberately construct their core product mechanics—including user interfaces, recommendation engines and algorithms, as well as other targeting tools—to keep users scrolling and clicking. Collecting and leveraging enormous amounts of personal user data, platforms target their users with customized feeds and recommendations to push them to spend more and more time on their platform. In doing so, these companies have actively pulled users toward extremism and hate, creating new trends, new communities, and new harms.

In one stunning example, Facebook researchers determined that the platform was responsible for pushing users into “rabbit holes” and feeding them

ever-more extreme content. Company researchers set up a test account, “Carol,” and assigned it interests in political conservatism, politics, parenting, and Christianity. Five days later, Facebook was recommending that “Carol” join conspiratorial groups and inundating “Carol’s” feed with “a barrage of extreme, conspiratorial, and graphic content” that violated Facebook’s own content rules, including those on hate speech.

These platforms know that their recommendation engines at times contribute to extremist, terrorist, and anti-democratic violence around the world. In spite of this they have taken little meaningful action. This is, in part, due to the sweeping immunity they enjoy under the lower courts’ misinterpretation of § 230.

As the late Judge Katzmann explained, Congress enacted § 230 at the dawn of the commercial internet—before social media existed—to protect families and children from inappropriate material online, while also protecting the nascent internet industry from potentially fatal liability for third-party postings on platform-supported static message and bulletin boards. *See Force v. Facebook*, 934 F. 3d 53, 78 (2d Cir. 2019) (Katzmann, J., concurring in part). But under some lower courts’ misguided and overbroad interpretations of § 230(c)(1), internet companies have come to enjoy near-total protection from judicial scrutiny. This means that even where internal documents show that platforms are aware that their conduct created significant social harms, and that those platforms have actively chosen to do nothing or double down on the tools that lead to such harm, the victims of those harms remain unable to seek redress through the courts.

When Congress passed § 230 in 1996, approximately 36 million people were online. Today, that number is 5.47 billion. The early internet of static messaging boards is a far cry from today's sophisticated platforms, which harvest reams of personal data about their users and deploy that data through carefully designed interfaces, recommendation engines and algorithms, and other targeting tools to keep users on their platforms in order to keep generating revenue. Moreover, in many countries, these platforms are not just a part of the internet: they are the internet.

In short, one of the world's most global and influential industries—which its own experts admit radicalizes individuals and spreads hateful and extremist content, often with egregious consequences—has developed without an attendant body of law. This lack of legal accountability has led to strong disincentives for platforms to address the serious harm created by their conduct.

The spread of hateful and extremist content has systemic effects, but it also impacts individuals on a daily basis. In its nationally representative annual survey, ADL found that approximately forty percent of people report having experienced online harassment. More than half of respondents who reported experiencing any harassment reported experiencing severe online harassment, a category that includes physical threats, sustained harassment, stalking, sexual harassment, doxing, and swatting. Vulnerable communities—including Jewish Americans, Asian Americans, and members of the LGBTQ+ communities—report higher than average rates of harassment. The rates of online harassment are also higher for youth. Among youth aged thirteen to seventeen, nearly half reported experiencing online harassment, and among vulnerable

youth the rate was almost two-thirds. Twice as many Asian Americans reported experiencing online harassment in 2022 as in 2021, and at the same time, there has been a spike in offline violent and fatal attacks against Asian Americans. The increase in reports of online harassment comes at a time when antisemitic and other hate incidents and hate crimes have also been on the rise. While most hate speech online is not unlawful, the fact that platforms' features may facilitate a transition into unlawful speech and conduct is cause for concern. This is just one example of the readily apparent correlation between increased online hate and on-the-ground violence.

Following decades of advocacy by ADL and other groups, and faced with the threat of legislative and regulatory action, some platforms are beginning to do more to moderate content. The percentage of instances where platforms deleted content that users reported for containing physical threats against them increased from fourteen percent in 2021 to nineteen percent in 2022. Still, this is far too little, and social media companies' self-regulation of hate and extremism is halting and uncertain. For instance, since Elon Musk instituted a new policy of unmoderated speech at Twitter, hate speech has thrived on a platform that used to be an industry leader in transparency and genuine partner in content moderation. The daily instance of racial slurs against Black Americans nearly tripled and Twitter went from taking action against sixty percent of tweets reported by ADL to thirty percent. This example illustrates the importance of content moderation and argues strongly against weakening § 230 in a way that would result in even stronger disincentives against moderating hateful content across all platforms.

Content moderation is the first line of defense against the continued mainstreaming of extremism and hate. Under § 230(c)(2)'s Good Samaritan protection, platforms can moderate content without fear of liability. *See Force v. Facebook*, 934 F. 3d 53, 81 (2d Cir. 2019) (Katzmann, J., concurring in part) (citing *FTC v. LeadClick Media, LLC*, 838 F. 3d 158, 174 (2d Cir. 2016)). It is evident to ADL that any removal or curtailment of this protection will result in even more hate and extremism online. Therefore, it is crucial that this court affirm § 230(c)(2)'s Good Samaritan protection.

In short, there is now a tremendous amount of research, including from the companies themselves, indicating that social media platforms knowingly design their recommendation engines, algorithms, and other technological tools to draw users ever-more efficiently into deeper engagement. Far too often, that can result in spreading hate, radicalizing users, and recruiting extremists, at times significantly influencing violent actors. Importantly, it is difficult to test this research and confirm the way recommendation engines and targeting tools pursue engagement because of platforms' lack of transparency and the overbroad immunity under § 230(c)(1). Given the plain language of § 230(c)(1), the intent of the drafters, the growing evidence of harm, and the acknowledgments from the platforms themselves, this Court should make clear that § 230(c)(1) does not immunize platforms for their own conduct merely because that conduct involves third-party speech. At the same time, however, this Court should affirm the core protection of the statute as embodied in the Good Samaritan provision of § 230(c)(2): that platforms are immune from liability for content created by third parties and for the traditional editorial functions of publishing, including the decision to keep or remove content.

ARGUMENT

I. SOCIAL MEDIA PLATFORMS FOMENT HATE AND EXTREMISM, RADICALIZE INDIVIDUALS, AND CONNECT THOSE RADICALIZED INDIVIDUALS TO NEW COMMUNITIES

A. Platforms actively support and grow new extremist communities by amplifying hateful and extremist content, recommending it to users, and recommending users join groups that engage in extreme and conspiratorial rhetoric.

There is little doubt that social media companies have helped radicalize users and normalize both online and offline extremism.² While most users are not radicalized when they come into contact with hate content, mainstreaming extremism so that even a relatively small number of extremists act on these ideas has real, on-the-ground effects. Given the scale of social media platforms, the results are anything but small. Section 230 should not preemptively immunize social media companies for virtually any role they play, no matter how key, in facilitating crimes committed by even one person.

It has long been clear to ADL and other researchers that social media companies take extremist views

² See *Domestic Terrorism and Violent Extremism: Examining the Threat of Racially, Ethnically, Religiously, and Politically Motivated Attacks, Part II Before the S. Comm. on Homeland Security and Government Affairs*, 117 Cong. 6 (2021) (statement of Jonathan Greenblatt, CEO & National Director, Anti-Defamation League), <https://www.adl.org/sites/default/files/adl-testimony-hsgac-domestic-terrorism-violent-extremism-2021-08-05.pdf> [hereinafter *Domestic Terrorism Greenblatt Testimony*].

from the fringe into the mainstream. They do this by pursuing two strategies through their product design. First, they leverage the troves of user data they collect to identify and target individuals susceptible to hateful and extremist content and then recommend that content to them. Second, they build online communities by effectively mobilizing and rewarding outrage, one of the best ways to ensure more and deeper engagement. ADL and other researchers have shown that these resulting communities and the conspiracies they espouse are frequently laced with antisemitism and other types of hate.

These strategies are motivated by a relentless pursuit of user engagement. They produce environments that foster hate and extremism at a scale that would not otherwise exist. This is because, as study after study has shown, extremist and outrage-provoking content is among the best performing types of content when engagement is the primary goal, as platforms pull users towards escalating moral outrage and tribalism.³

A landmark ADL survey of YouTube recommendation exposure found that YouTube, a platform with 2.5 billion users, targets and delivers extremist content to highly susceptible users.⁴ Overall, exposure was

³ See Paul Lewis, *'Fiction is outperforming reality': how YouTube's algorithm distorts truth*, GUARDIAN (Feb. 2, 2018), <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>.

⁴ See *Exposure to Alternative and Extremist Content on YouTube*, ANTI-DEFAMATION LEAGUE CENTER FOR TECHNOLOGY & SOCIETY, 6 (May 3, 2022), https://www.adl.org/sites/default/files/pdfs/2022-05/FINAL_FINAL_ADL-Report-Single-Final-Design.pdf [hereinafter *Exposure to Alternative and Extremist Content on YouTube*].

disturbingly high: one in ten survey participants viewed at least one video from an extremist channel, and two in ten viewed at least one video from an alternative channel that can serve as a gateway to more extremist content.⁵ Survey participants who watched such videos were more likely to receive recommendations to other extreme and alternative content, and consumption of such content was concentrated among people with negative racial views.⁶ In other words, despite recent laudable changes to YouTube's recommendation algorithm to better deal with "borderline content and misinformation," YouTube's recommendation engines continue to serve up hateful and extreme content to users particularly susceptible to such content.⁷

Unfortunately, YouTube also drives engagement by recommending hateful and extremist content to users watching unrelated content. Users participating in a Mozilla study of YouTube recommendations used a browser extension in a bid to reduce harmful content.⁸ The study revealed that seventy-one percent of the content flagged by the participating users was recommended to them by YouTube's algorithm and recommended videos were forty percent more likely to be unrelated to videos the users had searched for.⁹ In

⁵ *See id.*

⁶ *See id.* at 7.

⁷ *See id.* at 7, 12.

⁸ *See* Mozilla, YouTube Regrets 5 (July 2021), https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf.

⁹ *See id.* at 13 ("In 43.3% of cases where we have data about trails a volunteer watched before a Regret, the recommendation was completely unrelated to the previous videos that the volunteer watched.").

other words, YouTube is recommending to users hateful, extreme, or inappropriate content, often unrelated to what the user was previously watching.

In addition to amplifying and targeting hateful and extremist third-party content posted on their sites, mainstream platforms lend their unparalleled reach to content that first incubates on more fringe platforms. The reach provided by large platforms results in the exponential spread of fringe content into the mainstream.¹⁰ In one study, ADL's Online Hate Index, a machine-learning system that detects hate targeting vulnerable groups online, measured the potential reach of antisemitic tweets over one week to be 130 million people.¹¹

B. Platforms fomenting hate online play a role in facilitating offline violence, which disproportionately impacts vulnerable communities.

Social media platforms' amplification of radicalizing and hateful has translated online hate to offline violence at home and abroad.¹² Before the deadly attacks in Paris that took the life of Nohemi Gonzalez, ISIS flooded social media platforms with recruitment propaganda intended to inspire Muslims living in the

¹⁰ See *Domestic Terrorism Greenblatt Testimony*, at 9-11.

¹¹ See *The Rate of Hate: Measuring Antisemitism Prevalence and Enforcement on Reddit and Twitter*, ANTI-DEFAMATION LEAGUE CENTER FOR TECHNOLOGY & SOCIETY 4, 12 (MAY 3, 2022), https://www.adl.org/sites/default/files/pdfs/2022-05/How%20Platforms%20Rate%20on%20Hate%202022_OHI_V10.pdf [hereinafter *The Rate of Hate*].

¹² See MAX FISHER, *THE CHAOS MACHINE: THE INSIDE STORY OF HOW SOCIAL MEDIA REWIRED OUR MINDS AND OUR WORLD* 312-316 (2022).

West to commit violent acts of terror.¹³ Internal whistleblower reports show that Facebook amplified business pages for ISIS and other terrorist organizations, *see Force*, 934 F. 3d at 85 (Katzmann, J., concurring in part), and a UN Security Council report noted that Facebook enables ISIS to fundraise by selling antiquities.¹⁴ Closer to home, one of the men who participated in a plot to kidnap Michigan Governor Gretchen Whitmer joined the Wolverine Watchmen group that planned the attack after Facebook recommended he join.¹⁵ And domestic terrorist attacks and acts of violent extremism in Charleston, Charlottesville, Pittsburgh, Poway, El Paso, and Buffalo were all connected to gunmen radicalized on social media by communities of like-minded extremists brought together by social media platforms.¹⁶

¹³ *See The ISIS Impact on the Domestic Islamic Extremist Threat*, ANTI-DEFAMATION LEAGUE, 3, 22-24 (MAR. 22, 2016), https://www.adl.org/sites/default/files/documents/assets/pdf/combatting-hate/CR_4473_HomegrownExtremismReport-2009-2015_web2.pdf.

¹⁴ Letter dated Jan. 20, 2020 from the Chair of the Security Council Committee pursuant to resolutions 1267 (1999), 1989 (2011) and 2253 (2015) concerning Islamic State in Iraq and the Levant (Da'esh), Al-Qaida and associated individuals, groups, undertakings and entities addressed to the President of the Security Council, U.N. Doc. S/2020/53 (Jan. 20, 2020).

¹⁵ *See FBI informant: Facebook led me to infiltrate plot to kidnap Gretchen Whitmer*, Bridge Michigan (Mar. 5, 2021), <https://www.bridgemi.com/michigan-government/fbi-informant-facebook-led-me-infiltrate-plot-kidnap-gretchen-whitmer>.

¹⁶ *See Domestic Terrorism Greenblatt Testimony*, at 1, 7, 8; Jonathan E. Bromwich, *Before Massacre Began, Suspect Invited Others to Review his Plan*, N.Y. TIMES (May 17, 2022), <https://www.nytimes.com/2022/05/17/nyregion/buffalo-shooting-discord-chat-plans.html>.

Where platforms draw users into conspiracies by recommending conspiracy theory content, they do more than play an active role in spreading that content and attracting more recruits. Platforms are increasingly bringing disparate conspiracy adherents and conspiracies together, making it possible for adherents—and those susceptible to radicalization—to be absorbed into all-encompassing stories that have the broadest appeal and which the FBI deem to pose potential domestic terrorism threats.¹⁷ This is one of the dangerous consequences of platforms' pursuit of engagement and profit at the expense of user safety.

In addition to conspiracy theories and large-scale tragic attacks, online hate can also inspire daily, smaller scale violence that is no less pervasive and corrosive. For example, this year Asian Americans reported the largest increase in online harassment ever, from twenty-one percent in 2021 to thirty-nine percent in 2022, mirroring an increase in violent and fatal attacks offline, often against Asian American women.¹⁸ Similarly, sixty-six percent of the LGBTQ+ community reported experiencing online harassment in 2022, following the single deadliest year on record for transgender individuals.¹⁹ After the 2020 murder

¹⁷ See FBI, Anti-Government, Identity Based, and Fringe Political Conspiracy theories Very Likely Motivate Some Domestic Extremists to Commit Criminal, Sometimes Violent Activity (May 30, 2019), <https://www.justsecurity.org/wp-content/uploads/2019/08/420379775-fbi-conspiracy-theories-domestic-extremism.pdf>.

¹⁸ See *Online Hate and Harassment: The American Experience 2022*, Anti-Defamation League (June 20, 2022) at 5, 24-28, <https://www.adl.org/sites/default/files/pdfs/2022-09/Online-Hate-and-Harassment-Survey-2022.pdf>. [hereinafter *Online Hate and Harassment 2022*]

¹⁹ See *id.* at 5, 9.

of George Floyd, ADL reported that anti-Black posts on Facebook had quadrupled,²⁰ and the number of white supremacist propaganda incidents has nearly doubled to the highest number of incidents since ADL started tracking the phenomenon.²¹ And 2021 saw the largest number of antisemitic incidents since ADL began tracking in 1979, including a 167% increase year on year in incidents of assault and a 106% increase in incidents at non-Jewish K-12 schools.²² The connection between increased online hate and offline harassment and violence is inescapable.

C. Platforms are aware that their conduct produces harmful consequences, yet they take little meaningful action.

Social media companies know that their platforms foster hate online and inspire hate offline. The most recent window into a major platform operating with the same engagement-based business model as Respondent Google's YouTube comes from the internal Facebook documents leaked by Frances Haugen. Those documents stated that the company had "evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook and the family of apps are affecting societies around the world."²³ The documents linked the spread of such

²⁰ See *Online Hate and Harassment: The American Experience 2021*, ANTI-DEFAMATION LEAGUE (Mar. 3, 2022) at 9, https://www.adl.org/sites/default/files/pdfs/2022-05/Online%20Hate%20and%20Harassment%202021_042821_V17%282%29.pdf. [hereinafter *Online Hate and Harassment 2021*]

²¹ See *Domestic Greenblatt Testimony*, at 4.

²² See *Online Hate and Harassment 2021*, at 9.

²³ See *Holding Big Tech Accountable: Legislation to Build a Safer Internet Before the H. Comm. on Energy and Commerce Subcomm. on Consumer Protection and Commerce*, 117 Cong.

content directly to the platform’s mechanics and design. “Our core products’ mechanics, such as virality, recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish.”²⁴

The documents show that Facebook knows that engagement is driven by extreme, hateful, titillating, and divisive content.²⁵ “If left unchecked,” an internal Facebook presentation observed, the platform would offer users “more and more divisive content in an effort to gain user attention & increase time on the platform.”²⁶ When this internal Facebook review raised the alarm that the platform was exploiting “the human brain’s attraction to divisiveness” and recommended changes, senior executives quashed the review and refused to implement changes.²⁷

The documents also show that Facebook knows that the harm extends beyond speech to harmful and unlawful conduct. The documents prove that Facebook is aware of its role in amplifying “combustible election

(2021) (Statement of Jonathan Greenblatt, CEO & National Director, Anti-Defamation League), <https://www.adl.org/sites/default/files/adl-testimony-house-cmte-energy-commerce-holding-big-tech-accountable-2021-12-09.pdf>. [hereinafter *Holding Big Tech Accountable Testimony*]

²⁴ *See id.*

²⁵ *See* Jeff Horowitz & Deepa Seetharam, *Facebook Executives Shut Efforts to Make the Site Less Divisive*, WALL ST. J. (May 26, 2020), <https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>; Brandy Zadrozny, ‘Carol’s Journey’: What Facebook knew about how it radicalized users, NBC News (Oct. 22, 2021), <https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rna3581>.

²⁶ *Id.*

²⁷ *See id.*

misinformation;”²⁸ in spreading inflammatory, hateful content in India that was linked to deadly religious riots;²⁹ and in enabling the Myanmar military to conduct a genocide of the Rohingya.³⁰ As noted above, YouTube and other social media companies lack transparency around product affordances and optimization tactics. Despite detailed knowledge of the devastating social consequences of their conduct and lack of protection to consumers, platforms are not incentivized to make meaningful and lasting changes. Instead, they rely on a sweeping interpretation of § 230(c)(1) immunity to protect their profits and business models.

Recommendation engines may be pieces of computer code but they are coded by humans who direct specific outcomes. Platforms may pay lip service to going after hate online, but their own research—as well as that of countless external experts—shows that that they sacrifice the safety of their online spaces for user engagement at every turn. They do so, in no small part, because they are immunized by the overbroad interpretations of several lower courts, which interpreted § 230(c)(1) far beyond its text and purpose.

²⁸ See *Holding Big Tech Accountable Testimony*, at 6.

²⁹ See Newly Purnell & Jeff Horowitz, *Facebook Services Are Used to Spread Religious Hatred in India, Internal Documents Show*, WALL ST. J. (Oct. 23, 2021), <https://www.wsj.com/articles/facebook-services-are-used-to-spread-religious-hatred-in-india-internal-documents-show-11635016354>.

³⁰ See Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar’s Military*, N.Y. TIMES (Oct. 15, 2018), <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>.

II. THE COURT SHOULD MAKE CLEAR THAT § 230(c)(1) DOES NOT IMMUNIZE A PLATFORM FROM THE CONSEQUENCES OF ITS OWN TORTIOUS CONDUCT

The Ninth and Second Circuits' interpretation of § 230(c)(1) is inconsistent with the statute's text and history and overbroad as a matter of policy. The lower courts' misinterpretation of § 230(c)(1) is not the only reason for the proliferation of online hate and extremism and escalating offline consequences. But this misreading and overbroad application has deprived the victims and targets of unlawful content and conduct of the opportunity to pursue a viable remedy, prevented courts from developing a jurisprudence for the digital world, and enabled platforms to operate with legal impunity. In other words, the lower courts' incorrect reading of § 230 ensures that platforms can enjoy "power without responsibility."³¹

Congress never intended for § 230(c)(1) to completely immunize internet providers from the consequences of their actions. Rather, Congress intended a broad "Good Samaritan" protection for internet providers that engaged in content moderation, underscored by a narrow liability shield. To that end, "Congress emphasized the narrow liability shield that became § 230(c)(2), rather than the broad rule of construction laid out in § 230(c)(1)." *Force*, 934 F. 3d at 78 (Katzmann, J., concurring in part). This is not surprising. Both the Senate and House efforts that became § 230 had at their core the protection of families and children from inappropriate material housed in online spaces. *See id.*

³¹ Tushnet, Rebecca, *Power Without Responsibility: Intermediaries and the First Amendment* 76 GEO. WASH. L. REV. 986 (2008).

Instead of charging the FCC with regulating obscene material online, as the Senate proposed, Congress empowered platforms to self-regulate. *See id.* To keep platforms inside in the regulatory effort, the drafters made sure to overrule a then-recent New York State trial court decision that found an online service liable for defamation as a publisher of objectionable content. In enacting § 230, Congress intended to empower platforms to protect families and children from objectionable content while fostering a nascent and innovative industry.

It is now clear that extending § 230(c)(1) to immunize all platform conduct produces incongruous and unjust results. This is because platform behavior extends far beyond § 230's scope. In addition to hosting extremist content posted by users, platforms play an active role in auto-generating, recommending, and amplifying radicalizing hateful content as well as connecting users and recommending groups and communities to join. As noted above, Facebook auto-generates business pages for terrorist groups, including Al Qaeda and ISIS, giving such groups the imprimatur of legitimacy even as it purports to target those groups for content moderation.³² *See Force v. Facebook*, 934 F. 3d at 85 (Katzmann, J., concurring in part). In fact, Facebook's amplification of terrorist pages and content goes one step further. Facebook "integrates users' photos and other media to generate videos commemorating their previous year. Militants get a ready-made propaganda clip, complete with a thank-you message

³² *See also, YouTube's Filter Bubble Problem is Worse for Fox News Viewers*, TECH TRANSPARENCY PROJECT (Oct. 24, 2021), <https://www.techtransparencyproject.org/articles/youtubes-filter-bubble-problem-worse-fox-news-viewers>.

from Facebook.” *Id.* at 86 (Katzmann, J., concurring) (citations omitted).

Platforms may not be liable for all of the harms mentioned in this brief, but Congress never intended to confer near-total, preemptive immunity on internet providers for all conduct implicated in their design, deployment and ongoing decision-making merely because there is some nexus with third-party content.³³

There is no more “plausible deniability.” Platforms are aware that their conduct produces violent and destabilizing consequences around the world.³⁴ Yet because the lower courts misinterpreted § 230(c)(1) to confer immunity to platforms for virtually all tortious conduct, the massive social media industry has grown up largely shielded from the courts and the normal development of a body of law. It is highly irregular for a global industry that wields staggering influence to be protected from judicial inquiry. Congress never intended for internet providers to be immunized from the get-go from nearly all secondary liability.

Section 230 has already shielded platforms from numerous accusations related to the consequences of enabling hate and radicalism to spread online. Considering the vast reach and impact of platforms’ recommendation engines, and the seriousness of the

³³ See *Holding Big Tech Accountable: Legislation to Build a Safer Internet Before the H. Comm. on Energy and Commerce Subcomm. on Consumer Protection and Commerce*, 117 Cong. 1 (2021) (statement of Jonathan Greenblatt, CEO & National Director, Anti-Defamation League) <https://www.adl.org/sites/default/files/adl-testimony-house-cmte-energy-commerce-holding-big-tech-accountable-2021-12-09.pdf>.

³⁴ See *supra* pp. 14-16.

consequences they produce, “caution is warranted before courts extend the CDA’s reach any further.” *Id.* at 78 (Katzmann, J., concurring in part).

III. THE PLAIN TEXT OF § 230(c)(2) PROTECTS A PLATFORM’S POWER TO EXERCISE EDITORIAL JUDGMENT

Content moderation is one of the strongest tools in a platform’s arsenal against online hate.³⁵ Platforms have the most visibility about what is posted on their sites and the most power to keep or remove unlawful conduct or content that violates their own terms of service. When users encounter hateful, extremist, or unlawful content, they must rely on platforms to act.

This Court should affirm that § 230(c)(2) protects a platform’s power to exercise editorial judgment through publishing and removing third-party content to mitigate the virality and impact of hateful and extremist content that may or may not be unlawful. Eviscerating or curtailing § 230(c)(2) would likely put an effective stop to the very type of platform content moderation that ADL and many others have dedicated enormous time and energy to promoting, and that some market forces are also now pushing toward.³⁶

³⁵ See *State of Social Media Regulation: Misinformation, Exploitation, Harassment, and Radicalization Before the S. Comm. on Judiciary*, 2021 Leg. (Cal. 2021) (statement of Lauren T. Krapf, ADL Technology Policy & Advocacy Counsel), https://sjud.senate.ca.gov/sites/sjud.senate.ca.gov/files/lauren_krapf_testimony_-_adl.pdf

³⁶ See e.g. Sergii Denysenko, *Key Considerations For Ensuring Brand Safety*, FORBES (Mar. 20, 2022), <https://www.forbes.com/sites/forbesagencycouncil/2022/03/30/key-considerations-for-ensuring-brand-safety/?sh=3f735e2d2000> (explaining how advertisers can protect their message from being promoted next to inapposite content).

Content moderation alone will not solve the problem of hate online, but it is a necessary component of any robust solution.

Recently, some platforms have taken more significant steps towards engaging in content moderation to reduce the spread of hate and extremism. In 2020, after a decade of lobbying by ADL, Facebook announced that Holocaust denialism would finally be classified as hate speech and removed from the platform.³⁷ YouTube has made over thirty recent changes to its recommendation engine to reduce the spread of harmful content.³⁸ And yet, as this brief has illustrated, much more is needed to mount a successful fight against hateful and extremist content online.

These advances in using content moderation to combat egregious, but lawful, hate and extremist content make it all the more crucial to retain the Good Samaritan provision of § 230(c)(2) while limiting the current overbroad sweep of § 230(c)(1). The forward momentum is almost certain to come crashing to a halt if changes to the law meant that such moderation opened the door to liability.

ADL is dedicating significant resources to addressing the current state of online hate, including by continuing to measure and disclose platform enforcement of their own policies. Platforms are still far too unlikely to remove hateful content, mitigate its virality

³⁷ See *After Long Fight, ADL is Relieved at Facebook Announcement That It Will Remove Holocaust Denial Content*, ANTI-DEFAMATION LEAGUE (Oct. 12, 2020), <https://www.adl.org/news/press-releases/after-long-fight-adl-is-relieved-at-facebook-announcement-that-it-will-remove>.

³⁸ See *Exposure to Alternative and Extremist Content on YouTube*, at 7.

or impact, or otherwise sanction users who post such content.³⁹ When ADL's Online Hate Index tracked antisemitic posts on Twitter and Reddit to measure the platforms' responses to hateful content, it found that approximately seventy to seventy-five percent of the posts tracked remained on the platforms more than two months after their initial discovery.⁴⁰ Even after ADL notified the platforms about the antisemitic posts, more than fifty-five percent of them remained up one week after notification.⁴¹ This example illustrates the unfortunate fact that content moderation is not yet a priority for social media platforms.

Unless platforms can continue to engage in content moderation without fear of legal and financial liability, they will continue to facilitate and normalize on and offline hate and extremism. It is therefore crucial that this Court affirm immunity for a platform's exercise of traditional editorial functions under § 230(c)(2).

Courts agree that a plain text reading of the statute clearly protects a platform from liability for the third-party content it publishes and the editorial decisions it takes regarding that content. Section 230(c)(2), which immunizes platforms for "any action voluntarily taken in good faith to restrict access to or availability of material. . . whether or not such material is constitutionally protected," 47 U.S.C. § 230(c)(2)(A), "bars lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional edito-

³⁹ See *supra* at 6; *Online Hate and Harassment 2022*, at 33; see also *One Year After Ban, Holocaust Denial Remains on Facebook*, ADL (Dec. 1, 2021), <https://www.adl.org/blog/one-year-after-ban-holocaust-denial-remains-on-facebook>.

⁴⁰ See *The Rate of Hate*, at 5, 12

⁴¹ See *id.*

rial functions—such as deciding whether to publish, withdraw, postpone or alter content” provided by another for publication.” *Force*, 934 F. 3d at 81 (Katzmann, J., concurring in part) (citing *FTC v. LeadClick Media, LLC*, 838 F. 3d 158, 174 (2d Cir. 2016)). This plainly protects a platform from liability arising out of the content posted by a third party, as well as the decision to keep or remove that content, for the simple reason that the content is “provided by another information content provider.” 47 U.S.C. § 230(c)(1).

Affirming immunity for the exercise of traditional editorial functions, including the type of content moderation that ADL has long urged upon platforms, is consistent with this Court’s precedents on editorial discretion. Although these precedents relate to government action, they offer strong analogs to social media platforms’ exercise of editorial discretion. In *Miami Herald Publishing Co. v. Tornillo*, 418 U.S. 241, 94 S. Ct. 2831 (1974), the Court found unconstitutional a Florida statute that required newspapers to give equal reply space to those they criticized in editorials, noting that the choices of material “constitute the exercise of editorial control and discretion,” which was protected by the First Amendment. Similarly, in *Manhattan Cmty. Access Corp. v. Halleck*, 204 L. Ed. 2d 405, 139 S. Ct. 1921, 1926 (2019), this Court noted the First Amendment right to editorial discretion of a private entity operating public access channels, observing that “merely hosting speech by others is not a traditional, exclusive public function and does not alone transform private entities into state actors subject to First Amendment constraints.”

Since content moderation is one of social media platforms’ most powerful weapons against the continued spread of hateful and extremist content, and it is

expressly immunized by § 230(c)(2), this Court should not hesitate to affirm that immunity.

CONCLUSION

For the foregoing reasons, this Court should make clear that § 230(c)(1) does not immunize a platform for its conduct merely because that conduct involves third-party speech. The Court should also affirm that platforms retain, as part of their exercise of traditional editorial functions, § 230(c)(2) immunity for publishing and moderating third party content.

Respectfully submitted,

STEVEN M. FREEMAN

Counsel of Record

EILEEN B. HERSHENOV

LAUREN T. KRAPF

ANTI-DEFAMATION LEAGUE

605 Third Avenue

New York, NY 10158

(212) 885-7700

sfreeman@adl.org

Counsel for Amicus Curiae

December 7, 2022