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 AMICI CURIAE ARTICLE 19: GLOBAL CAMPAIGN FOR FREE EXPRESSION AND THE INTERNATIONAL JUSTICE CLINIC AT THE UNIVERSITY OF CALIFORNIA, IRVINE SCHOOL OF LAW IN SUPPORT OF RESPONDENT

DAVID KAYE
EMILE G. AYOUB
INTERNATIONAL JUSTICE CLINIC
UNIV. OF CALIFORNIA, IRVINE
SCHOOL OF LAW
401 E. PELTASON DRIVE
IRVINE, CA 92697
(949) 824-0066
DKAYE@LAW.UCI.EDU

ROBERT P. LATHAM
COUNSEL OF RECORD
MARC FULLER
HANNAH WALSH
JACKSON WALKER LLP
2323 ROSS AVENUE, STE. 600
DALLAS, TX 75201
(214) 952-6000
BLATHAM@JW.COM

JANUARY 19, 2023 *COUNSEL FOR AMICI CURIAE*

SUPREME COURT PRESS ♦ (888) 958-5705 ♦ BOSTON, MASSACHUSETTS

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page

FEDERAL CASES

INTERNATIONAL CASES

TABLE OF AUTHORITIES – Continued

Page

CONSTITUTIONAL PROVISIONS

FEDERAL STATUTES

TABLE OF AUTHORITIES – Continued

Page

INTERNATIONAL STATUTES

CONGRESSIONAL RECORD

TABLE OF AUTHORITIES – Continued

Page

OTHER AUTHORITIES

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

TABLE OF AUTHORITIES – Continued

Page

INTEREST OF AMICI CURIAE

Amici are organizations that share a strong interest in ensuring that individuals around the world are able to participate freely in online expression and debate on matters of public concern, including human rights issues.[1] As such, they share the view that Section 230 of the Communications Decency Act, 47 U.S.C. § 230, promotes the same values of access to information and freedom of expression that are guaranteed by international human rights law, in particular Article 19 of the International Covenant on Civil and Political Rights ("ICCPR"), which the United States ratified in 1992.[2]

---

[1] Pursuant to Rule 37.6, amici curiae state that no counsel for any party authored this brief in whole or in part and no person or entity, other than amici curiae, their members, or their counsel has made a monetary contribution to its preparation or submission.

[2] *See* International Covenant on Civil and Political Rights, Oct. 5, 1977, T.I.A.S. No. 92–908, 999 U.N.T.S 171 ("ICCPR"). Article 19 of the ICCPR commits state parties to respect and ensure, *inter alia*, everyone's right "to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." It further provides that any restriction on the right to freedom of expression must: (i) be "provided by law," (ii) imposed only for legitimate objectives as set out in article 19(3) of the ICCPR, and (iii) be necessary and proportionate to achieve a legitimate objective. ICCPR art. 19(3); *see also* U.N. Human Rights Council, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on the Regulation of User-Generated Online Content*, U.N. Doc. A/HRC/38/35 (2018).

ARTICLE 19: Global Campaign for Free Expression ("ARTICLE 19") is a nonpartisan non-governmental organization founded in 1987, with an international office in London, UK, and regional offices in the United States, The Netherlands, Brazil, Mexico, Senegal, Kenya, and Bangladesh, and several country offices in Asia and the Pacific region. The organization, named for the corresponding article of the Universal Declaration of Human Rights, advocates for freedom of expression as a fundamental human right, including in the digital environment. It has participated as *amicus curiae* in free expression cases around the world, including in the United States, and has intervened at regional bodies dealing with intermediary liability and freedom of expression. ARTICLE 19 also actively participates in discussions at the United Nations Human Rights Council and the United Nations General Assembly on issues related to counter-terrorism and human rights.

The International Justice Clinic at the University of California, Irvine School of Law ("IJC") promotes international human rights law at national, regional, international, and corporate levels, in the United States and globally. IJC is directed by Professor David Kaye, the former United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, who has written extensively on the enjoyment and protection of human rights in digital environments. IJC has extensive experience especially addressing threats to human rights in the digital realm, working alongside civil society organizations and other stake-holders from across the globe.

INTRODUCTION AND SUMMARY OF THE ARGUMENT

This Court is asked to decide the fate of one of the foundations of online speech: the liability shield for platforms that filter, organize, and display third-party content. The wrong outcome in this case could transform the internet for everyone. If petitioners' constrained interpretation of Section 230 prevails, free expression and its vital role in the protection of human rights will be among the first casualties.

Petitioners have narrowed their claims in an effort to reassure the Court that the reversal of the Ninth Circuit's judgment would not dramatically curtail free speech online. But regardless of whether their attacks on Section 230 are aimed directly at the display of third-party speech or indirectly at the algorithmic tools that recommend it, the result is the same, as petitioners' claims effectively turn on the illegality of the ISIS-related content posted by users. To protect themselves from potentially crushing liability, websites with the means to do so would err on the side of caution by removing or blocking any content that might even remotely touch on illegal behaviour.

Recognizing that human review is logistically and financially impossible to scale to the levels necessary to address the overwhelming volume of online content, some (including the Ninth Circuit) have suggested that automated tools have progressed to the point where they now can accurately and consistently identify all types of unlawful speech. Some *amici* have even suggested that the algorithmic tools used for ranking and recommendation can themselves serve this function. Not so. Automated tools cannot make complex

assessments of illegality of expression, which means that reliance on them will result in lawful speech being arbitrarily denied a platform. Marginalized groups and those who criticize violent extremism or who report on human rights violations will be denied a voice.

Section 230's primary purpose has always been to protect online expression and debate from such threats, both public and private. This protection is essential to achieving the aims of international human rights law. It is also consistent with the Court's First Amendment jurisprudence, justly celebrated worldwide as a gold standard in speech protection. That jurisprudence rests on fundamental values and assumptions about the role of free speech in democracies and the government's obligations to protect it. The same values and assumptions underlie international human rights law and have been recognized by other countries and international regulatory bodies. At the global level, international human rights standards preclude governments from imposing unnecessary or disproportionate limits on individuals' freedom of expression, whether directly or by requiring that private actors take such steps. Even without the First Amendment or statutory provisions akin to Section 230, these other countries and regulatory bodies have repeatedly rejected theories of intermediary liability like those effectively advanced by petitioners and their *amici* out of a concern for protecting individual free expression.

The Court should do the same in this case. *Amici* do not oppose rights-respecting regulation of internet platforms. But regulatory complexity requires careful legislative fact-finding and drafting, especially when it directly implicates fundamental constitutional rights. Whatever the merits of various reform proposals,

Section 230's strong protections for websites that organize, display, and recommend third-party content provide the best foundation for evaluating potential reforms. *Amici* thus respectfully urge the Court to affirm the Ninth Circuit's judgment.

ARGUMENT

I. THE COURT'S APPLICATION OF SECTION 230 SHOULD BE GUIDED BY FIRST AMENDMENT VALUES AND INTERNATIONALLY RECOGNIZED FREE SPEECH NORMS.

Regardless of whether the Court focuses on algorithmic recommendation systems (as petitioners now do), or considers more broadly websites' liability for organizing and displaying third-party content (as some *amici* urge), the claims in this case effectively turn on the illegality of content posted by users. The stakes for individual freedom of expression online cannot be overstated. If the Court were to reverse the Ninth Circuit's judgment or limit the broad immunity that the court of appeals applied, the inevitable result would be extensive and arbitrary removals of content and the suppression of user speech. Congress, this Court, and international judicial bodies have consistently acted to prevent such an outcome, recognizing that standards of intermediary liability directly affect the freedom of individuals' expression.

A. Congress Understood That Internet Users Are the Ultimate Beneficiaries of Broad Section 230 Immunity.

As its supporters in Congress envisioned at the time of its passage, Section 230 has been an engine

for the promotion of freedom of expression online.[3] This freedom has played an essential role in the protection of human rights over the past 25 years, allowing dissidents, investigators, and activists to share and access a wide range of content and information. Imposing liability on platforms for displaying or algorithmically recommending third-party content would reverse the judgment that Congress made in enacting Section 230 and the rationale that courts have long embraced in applying it.

Many websites might respond to the prospect of increased intermediary liability by severely restricting user speech.[4] Their actions would not be limited to unlawful speech. In an abundance of caution, they might also remove content that they believe complies with the law out of fear that a judge or jury might

---

[3] *See* 141 Cong. Rec. H8470 ("The Internet is a fascinating place and many of us have recently become acquainted with all that it holds for us in terms of education and political discourse. We want to make sure that everyone in America has an open invitation and feels welcome to participate in the Internet") (statement of Rep. Cox), ("We have the opportunity for every household in America, every family in America, soon to be able to have access to places like the Library of Congress, to have access to other major libraries of the world, universities, major publishers of information, news sources. There is no way that any of those entities . . . can take the responsibility to edit out information that is going to be coming in to them from all manner of sources onto their bulletin board.") (statement of Rep. Goodlatte).

[4] *See Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021*, BSR (Sept. 2022), https://www.bsr.org/reports/BSR_Meta_Human_Rights_Israel_Palestine_English.pdf (finding that, while some violent content was removed, Google had "higher error rates for Palestinian Arabic" which resulted in a disproportionate amount of removals based on language and location of the users).

disagree with their determinations.[5] Such a regime might also leave users vulnerable to bad-faith takedown requests targeting speech that is merely displeasing, not unlawful.[6] Additionally, over-removals could implicate other human rights, such as freedoms of religion, association, and assembly.[7] Political dissidents and religious minorities would be particularly vulner-

---

[5] Daphne Keller, *Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List*, Ctr. for Internet and Soc'y (Feb. 8, 2021), https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws (last accessed Dec. 4, 2022); Article 19, *Watching the Watchmen Content Moderation, Governance, and Freedom of Expression*, 29–30 (2021); *see, e.g.*, Paige Leskin, *A Year After Tumblr's Porn Ban, Some Users Are Still Struggling to Rebuild Their Communities and Sense of Belonging*, Business Insider (Dec. 20, 2019), https://www.businessinsider.com/tumblr-porn-ban-nsfw-flagged-reactions-fandom-art-erotica-communities-2019-8 (last accessed on Dec. 4, 2022).

[6] Keller, *supra* note 5; Lenka Fiala & Martin Husovec, *Using Experimental Evidence to Improve Delegated Enforcement*, 71 INT'L REV. L. & ECON. (2022), at 1; Daphne Keller & Paddy Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation*, in Social Media and Democracy: The State of the Field, Prospects for Reform, 220 (Nathaniel Persily & Joshua A. Tucker, Eds., Cambridge University Press, 2020).

[7] *See, e.g.*, Jillian C. York, *Silicon Valley's Sex Censorship Harms Everyone*, WIRED (March 18, 2022), https://www.wired.com/story/silicon-values-internet-sex-censorship/; Paige Leskin, *A Year After Tumblr's Porn Ban, Some Users Are Still Struggling to Rebuild Their Communities and Sense of Belonging*, Business Insider (Dec. 20, 2019), https://www.businessinsider.com/tumblr-porn-ban-nsfw-flagged-reactions-fandom-art-erotica-communities-2019-8.

able because their speech is often considered the most controversial.[8]

Courts recognized—and sought to avoid—this potential outcome in the earliest Section 230 cases. Shortly after Congress enacted Section 230, the Fourth Circuit observed in *Zeran v. America Online* that, because platforms have "millions of users" communicating a "staggering" amount of information,

> [t]he specter of tort liability in an area of such prolific speech would have an obvious chilling effect. It would be impossible for service providers to screen each of their millions of postings for possible problems. Faced with potential liability for each message republished by their services, interactive computer service providers might choose to severely restrict the number and type of messages posted. Congress considered the weight of the speech interests implicated and chose to immunize service providers to avoid any such restrictive effect.

Zeran v. America Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997).

While the ways people communicate online have changed somewhat since *Zeran*, that court's insight holds true today: websites will restrict freedom of expression if necessary to protect themselves from liability.

---

[8] Jillian C. York & Karen Gullo, *Offline/Online Project Highlights How the Oppression Marginalized Communities Face in the Real World Follows Them Online*, Electronic Frontier Foundation (March 6, 2018), https://www.eff.org/deeplinks/2018/03/offlineonline-project-highlights-how-oppression-marginalized-communities-face-real

Indeed, long before the enactment of Section 230, this Court recognized the connection between distributor liability standards and individuals' free speech rights. In *Smith v. California*, the Court held that unduly onerous liability standards for bookstores violate the First Amendment. *Smith v. California,* 361 U.S. 147, 153–54 (1959). This was not based merely on the bookstore's own rights, but on the rights of readers and authors who depended on bookstores' intermediary services—much as ordinary readers and speakers depend on internet platforms today. The Court wrote:

> [T]he bookseller's burden would become the public's burden, for by restricting him the public's access to reading matter would be restricted. If the contents of bookshops and periodical stands were restricted to material of which their proprietors had made an inspection, they might be depleted indeed. The bookseller's limitation in the amount of reading material with which he could familiarize himself, and his timidity in the face of his absolute criminal liability, thus would tend to restrict the public's access to forms of the printed word which the State could not constitutionally suppress directly. The bookseller's self-censorship, compelled by the State, would be a censorship affecting the whole public, hardly less virulent for being privately administered.[9]

---

[9] *Id.*; *see also Bantam Books, Inc. v. Sullivan*, 372 U.S. 58, 71-72 (1963).

The same principle applies even more forcefully in the internet context. The massive scale at which third-party information is uploaded makes it impossible for websites to identify and remove unlawful content with precision. Thus, Section 230 appropriately prescribes even broader immunity than that which protects physical booksellers. In doing so, it is consistent with First Amendment values and the Court's longstanding recognition of the connection between intermediary liability and individual free expression.

B. Foreign Courts Have Protected Users' Free Speech by Reducing the Specter of Liability for Platforms That Host That Speech.

Even where the First Amendment and Section 230 do not apply, foreign courts and regulatory bodies seeking to protect free expression online have adopted scienter requirements that protect websites from liability for third-party content that has not been adjudicated as unlawful. For example, the Supreme Courts of India and Argentina have held that platforms should face legal liability for user speech only if they know that a court or appropriate authority has adjudicated that speech to be unlawful, after full and fair legal process.[10] This "court order requirement" has also been enacted legislatively in some countries.[11]

---

[10] *See* Corte Suprema de Justicia de la Nación [CSJN] [National Supreme Court of Justice], 28/10/2014, "Rodríguez, María Belén c. Google Inc./daños y perjuicios," http://www.saij.gob.ar/corte-supremajusticia-nacion-federal-ciudad-autonoma-buenos-aires-rodriguez-maria-belen-google-incotro-danos-perjuicios-fa14000161-2014-10-28/123456789-161-0004-1ots-eupmocsollaf [https://perma.cc/6876-2G3P] (Arg.); Corte Suprema de Justicia de la Nación [CSJN] [National Supreme Court of Justice], 28/6/2022, Denegri,

Courts in Europe have similarly imposed knowledge requirements before platforms can be required by law to remove content. For example, the Court of Justice of the European Union ("CJEU") considered the issue of intermediary liability for copyright infringement in 2021.[12] The CJEU held that a platform's recommend-

---

Nathalia Ruth c. Google Inc.,/derechos personalísmos: Aciones relacionadas, https://sjconsulta.csjn.gov.ar/sjconsulta/documentos/verDocumentoByIdLinksJSP.html?idDocumento=7765751&cache=1656433432111 (Arg.); *Singhal v. Union of India*, (2015) 12 SCC 73, ¶¶ 100, 117 (India) holding that based on free expression considerations, a notice and takedown statute must be construed to mandate removal only based on court or other government order). *Royo v. Google* (Barcelona appellate court judgment 76/2013 of 13 February 2013 requiring court order prior to platform removal obligation) at Section 7; Asociación de *Internautas v. SGAE* (Spanish supreme court judgment 773/2009 of 9 December 2009), https://bit.ly/2HANz7t [https://perma.cc/62S3-NU2X] (holding that the EU legislation precludes requiring court orders for every removal); *see also Davison v. Habeeb* [2011] EWHC 3031 (QB) [68] (holding that notice of an allegedly defamatory blog post did not create actual or constructive knowledge where OSP was "faced with conflicting claims . . . between which it was in no position to adjudicate"); *Bunt v. Tilley* [2006] EWHC 407 (QB) [72] ("[I]n order to be able to characterise something as 'unlawful' a person would need to know something of the strength or weakness of available defences") (Eady, J.); *Kaschke v. Gray* [2010] EWHC 690 (QB) [100] (quoting Bunt, [2006] EWHC 407 (QB) [72]). *But see Tamiz v. Google Inc.* [2013] EWCA Civ 68 (holding that a blogging platform can be liable as a publisher of user content under defamation law, without consideration of eCommerce hosting defenses or standard for knowledge thereunder).

[11] *Marco Civil da Internet*, federal law 12.965, arts. 18–19; Law No. 20,430-modifying Law 17,336 on Intellectual Property, art. 85, Mayo 4, 2010, Diario Oficial [D.O.] (Chile) (English translation available at https://www.cdt.org/files/file/ChileanLaw20430-ModifyingLaw17336.pdf).

[12] Judgement of 22 June 2021, *Peterson v. Google LLC*, C-682/18

ations of content were not enough to establish "actual knowledge" and therefore remove the EU's statutory immunity for Internet platforms:

> [T]he fact that the operator of an online content-sharing platform automatically indexes content uploaded to that platform, that that platform has a search function and that it recommends videos on the basis of users' profiles or preferences is not a sufficient ground for the conclusion that that operator has "specific" knowledge of illegal activities carried out on that platform or of illegal information stored on it.[13]

The CJEU reasoned that, to protect freedom of expression, a platform could only be said to have "actual knowledge" of the infringing nature of speech in relation to a specific piece of content under certain circumstances.[14]

International human rights law also conditions mandatory content removal on the existence of an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity, and legitimacy.[15] Requiring

---

& C-683-18, ECLI:EU:2021:503.

[13] *Id.* at ¶ 114.

[14] *Id.* at ¶ 118.

[15] *See* Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on the regulation of user-generated online content, ¶ 7, U.N. Doc. A/HRC/38/35 (2018); *Google Spain v. Commission Nationale De L'informatique Et Des Libertés (CNIL)*, Case C-507/17, Submission by Article 19, ¶ 5 (Nov. 29, 2017) https://www.article19.org/wp-content/uploads/2017/12/Google-v-CNIL-A19-intervention-EN-11-

companies to remove content under the threat of liability in the absence of judicial process constitutes an undue delegation of regulatory functions to private actors that lack basic tools of independence and accountability.[16] In that sense, the then-U.N. Special Rapporteur for Freedom of Expression, David Kaye,[17] noted that "complex questions of fact and law should generally be adjudicated by public institutions, not private actors whose current processes may be inconsistent with due process standards and whose motives are principally economic."[18] Relying on private platforms to adjudicate the law, under circumstances in which their safest and most self-interested course is simply to censor user expression, inevitably would lead to the suppression of online speech.

Thus, international authorities are in accord with Section 230 policy and First Amendment values. To protect the freedom of expression of individuals, liability standards must protect websites that host such speech—at least until appropriate judicial process has determined the specific content to be unlawful. Legal regimes that attempt to delegate this duty to private

---

12-17-FINAL-v2.pdf.

16 *See* Eliska Pirkova & Estelle Massé, *EU Court decides on Two Major "Right to Be Forgotten" Cases: There Are No Winners Here*, Access Now (Oct. 23, 2019, 4:02 AM) https://www.accessnow.org/eu-court-decides-on-two-major-right-to-be-forgotten-cases-there-are-no-winners-here/.

17 David Kaye is currently the director of IJC, one of the *amici* submitting this brief.

18 Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on the regulation of user-generated online content, ¶ 17, U.N. Doc. A/HRC/38/35 (2018).

websites—as petitioners and their *amici* advocate—violate these well-established and widely recognized norms.[19]

II. AUTOMATED CONTENT MODERATION TOOLS CANNOT ADEQUATELY PROTECT FREE EXPRESSION.

These risks to freedom of expression will not be ameliorated by automated content moderation tools. Even the best, most accurate tools rely on calibration by humans. In the absence of immunity, website operators will design their automated systems to err even more conservatively on the side of removing controversial speech to avoid steep financial penalties.[20] The more severe the threatened liability, the more motivated they will be to tweak the tools so that such content is removed in its entirety, as quickly as possible.[21] Thus, automated content moderation is susceptible to the same incentives as human review when it comes to protecting websites against liability based on third-party content.

A. Automated Systems Cannot Accurately and Objectively Identify "Terrorist" Content and Other Harmful Speech.

Any discussion about the effectiveness of automated content moderation tools must also grapple

---

19 *Id.*

20 *See* Brief of Center for Democracy & Technology, et al. as Amici Curiae, p. 19, *Twitter, Inc. v. Taamnah*, No. 21-1496 (U.S. Dec. 6, 2022), https://www.supremecourt.gov/DocketPDF/21/21-1496/249112/20221205165314486_21-1496tsacCenterForDemocracyTechnology.pdf.

21 *See* Keller, *supra* note 5, at 4.

with their current and inherent shortcomings. Too often these shortcomings are ignored, as some courts and commentators have suggested that technology can thread the needle between the removal of unlawful speech and the protection of lawful speech. The Ninth Circuit made precisely this suggestion, opining that "Section 230's sweeping immunity is likely premised on an antiquated understanding of the extent to which it is possible to screen content posted by third parties." *Gonzalez v. Google LLC*, 2 F.4th 871, 913 (9th Cir. 2021). The court of appeals based its belief on reports that some large websites were "making laudable strides to develop tools to identify, flag, and remove inherently illegal content such as child pornography." *Id.* at 912–13. But the court of appeals conducted no further analysis of this reported progress and did not consider whether the same was true outside the narrow category of child sexual abuse material ("CSAM"). *Id.*

In fact, the success of automated content moderation tools in identifying CSAM cannot be replicated in the anti-terrorism context or any other context that requires sensitivity to context, tone, and nuance. CSAM is uniquely amenable to identification by automated tools. There is an international consensus on the illegality of CSAM, and there are clear parameters for which content should be flagged and removed. The duplicate detection technologies relied on by large websites can perform these tasks with a high degree of accuracy.[22] This success has been possible in large

---

[22] *See* Spandana Singh, *Everything in Moderation – An Analysis of How Internet Platforms are Using Artificial Intelligence to Moderate User-Generated Content* at 7, (July 22, 2019), https://www.newamerica.org/oti/reports/everything-moderation-analysis-

part because CSAM is illegal regardless of context and tone. Its legality does not generally depend on how the material is presented.[23]

More complex speech categories, such as "terrorist" speech, are different. Definitions of "terrorist" speech are notoriously vague and viewpoint-based.[24] Objective, consistent, and accurate identification of "terrorist" content evades the capabilities of automated tools, which are often unable to comprehend the tonal and contextual elements of speech or to identify when speech is satire or published for reporting purposes.[25]

To proactively monitor content at scale, websites now rely primarily on digital hash technology,[26] image

---

how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation.

[23] Even in the context of CSAM, automated content moderation tools can wrongfully flag and remove content. *See*, *e.g.*, Kashmir Hill, *A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal*, N.Y. TIMES (August 21, 2022), https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html.

[24] Stuart Macdonald, et al., *Regulating Terrorist Content on Social Media: Automation and the Rule of Law*, 15(2) INT'L J. L. IN CONTEXT 183, 188–89 (2019).

[25] *See* Singh, *supra* note 22.

[26] Hash matching assigns a unique digital "fingerprint" to previously detected harmful images and videos. Newly-identified harmful user-generated content can then be automatically removed if the computed hash matches a hash stored in the database of known harmful content. *See* Cambridge Consultants, *Use of AI in Online Content Moderation*, Ofcom, 9 (July 2019), https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf.

recognition,[27] or natural language processing ("NLP").[28] Websites also use these tools to flag and prioritize specific content for human review.[29] Major platforms already use these tools to filter and remove more content than the law currently requires, including content of a potentially "terrorist" nature.

Human rights and international legal observers are intimately familiar with the problems caused by overreliance on such tools.[30] They know that false positives—where websites flag and remove extremely valuable speech—are particularly prevalent in two categories of online speech: (1) the dissemination of news about terrorism, and (2) speech in languages other than English.

First, screening and removing terrorist content without contextual appreciation risks suppressing journalistic coverage and documentation of human rights violations. This has been studied by the Syrian Archive, a project by the human rights organization, Mnemonic, which relies on user-generated content from

---

[27] While digital hash technologies utilize image recognition, the technique can also be used more broadly for instance to identify specific objects within an image, such as a weapon. *See* Singh, *supra* note 22, at 14.

[28] NLP is a technique by which text is parsed in order to make predictions about the meaning of the text, for example whether it expresses a positive or negative opinion. *See* Natasha Duarte, et al., *Mixed Messages? The Limits of Automated Social Media Content Analysis*, Ctr. for Democracy & Tech., 9 (Nov. 28, 2017), https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/.

[29] *See* Singh, *supra* note 22, at 7.

[30] *See, e.g.*, Macdonald, *supra* note 23, 188–89.

social media to build criminal cases and conduct human rights research related to the ongoing conflict in Syria.[31] The Syrian Archive has tracked the removal of hundreds of thousands of videos and social media posts documenting crimes and human rights violations from various platforms, which has resulted in the loss of essential evidence that could have been used in international accountability efforts.[32] Documentation of crimes and human rights violations that focus on conflicts in Yemen, Sudan, and Ukraine are vulnerable to the same over-removal of content.[33]

Second, content moderation tools developed primarily in English are significantly less accurate when applied to speech in other languages and cultures. Automated tools are limited in their ability to parse variances in language that may result from different demographic and regional factors.[34] The effects of algorithmic biases are amplified through their continued use.[35] This may lead to the disproportionate removal of lawful content involving languages other than English. For example, during a May 2021 crisis in Israel, content from Palestinian Arabic speakers was disproportionately and inaccurately flagged and

---

[31] *See Lost and Found: Syrian Archive's Work on Content Taken Down From Social Media Platforms*, Syrian Archive (last updated Nov. 2022), https://syrianarchive.org/en/lost-found.

[32] *Id.*

[33] *See Our Work*, Mnemonic, https://mnemonic.org/en/our-work (last accessed Jan. 15, 2023).

[34] *See* Singh, *supra* note 22, at 18.

[35] Duarte, *supra* note 28, at 6; *see also* Cambridge Consultants, *supra* note 26, at 41.

removed from Facebook as terrorist content.[36] As later reported, this was at least in part the result of NLP classifiers which had "higher error rates for Palestinian Arabic."[37]

Germany's recent experiences with the pitfalls of identifying "terrorist" speech illustrate the difficulties inherent in the task—even when modern, automated tools are available. In 2016, Germany introduced the Network Enforcement Act (*Netzwerkdurchsetzungsgesetz*), better known as NetzDG.[38] The law requires platforms to remove "manifestly unlawful" content within 24 hours of notification under the threat of heavy fines.[39] After NetzDG came into force, two members of the far-right Alternative for Germany ("AfD") party posted incendiary anti-Muslim tweets that were flagged and deleted from Twitter as hate speech.[40] Also removed, however, were tweets from the satirical magazine, *Titanic*, which were obvious parodies of the AfD speech that had been removed by Twitter.[41] Even

---

[36] *See Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021*, BSR (Sept. 2022), https://www.bsr.org/reports/BSR_Meta_Human_Rights_Israel_Palestine_English.pdf.

[37] *Id.*

[38] Netzdurchsetzunggesetz [NetzDG] [Network Enforcement Act] Sept. 1 2017, BGBl I at 3352, (Ger.) (for an English translation, *See* https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2).

[39] *Id.* at art. 1, §§ 3-4.

[40] *See* Linda Kinstler, *Germany's Attempt to Fix Facebook is Backfiring*, ATLANTIC, May 18, 2018, https://www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/.

[41] *See id.*

several years later, researchers have been unable to show that NetzDG has resulted in the reduction of terrorist messaging, hate speech, and other harmful content online.[42]

B. A Lack of Transparency Compounds These Problems.

Any regime that relies on algorithmic systems to identify and remove "harmful" speech raises questions of transparency. A lack of transparency and accountability often accompanies the application of automated tools, raising further concerns for the freedom of expression rights of individuals whose content or accounts may have been subject to restrictive actions (such as demoting, demonetizing or removing) because of automated decision-making. This lack of transparency has been observed around how datasets are compiled, how accurate automated content moderation tools are, and how much content is removed both correctly and incorrectly.[43] Without sufficient transparency, necessity and proportionality of measures cannot be determined.

---

[42] William Echikson & Olivia Knodt, *Germany's NetzDG: A Key Test for Combatting Online Hate* (CEPS Research Report No. No. 2018/09, Nov. 2018), https://www.ceps.eu/ceps-publications/germanys-netzdg-key-test-combatting-online-hate/ .

[43] *See* Singh, *supra* note 22, at 16.

Notably, ARTICLE 19 has raised these transparency concerns in criticizing the Global Internet Forum to Counter Terrorism ("GIFCT"), which is probably the most advanced and certainly the most commonly used algorithmic tool available for combatting terrorist content. GIFCT began as a joint initiative by major websites to combine their efforts to counter extremism on their platforms. GIFCT created a shared hash database containing image and video hashes to moderate extremist content across platforms. In July 2020, ARTICLE 19, along with the American Civil Liberties Union, Human Rights Watch, and other organizations, expressed a concern that "GIFCT is maintaining a shared content removal resource that cannot be objectively evaluated to determine whether, for example, protected speech is being censored, or evidence of war crimes or other valuable evidence is being destroyed."[44]

A further problem arises where artificial intelligence and machine learning techniques are applied. Decisions made based on these systems involve hidden layers that may be unexplainable even to their creators.[45] Such concerns are particularly serious if upload filters are used, through which content may be blocked before it is even posted. This

---

44 *Joint Letter to New Executive Director, Global Internet Forum to Counter Terrorism*, Human Rights Watch (July 30, 2020), https://www.hrw.org/news/2020/07/30/joint-letter-new-executive-director-global-internet-forum-counter-terrorism.

45 *See* Cambridge Consultants, *supra* note 26, at 26.

makes it challenging to even know if the content has been removed in error.[46]

Users need to understand the reasoning behind the restriction of their speech to enable accountability and redress mechanisms, such as appeals of improper content removals. While platforms have started providing users with more transparency on their enforcement practices and greater access to appeal, users sometimes still lack the information required to exercise their rights (*e.g.*, information on which rules exactly were violated or whether content was removed following an automated decision-making process). This limits the effectiveness of the appeals processes.[47]

In sum, automated content moderation systems are vulnerable to the same incentives that plague human-review systems. When intermediary liability exposure is increased, the platforms' incentives to remove speech will also increase. Moreover, the accuracy, consistency, and transparency concerns that accompany even the most advanced automated systems cannot be ignored. Thus, technology should not be viewed as a substitute for the immunity that Section 230 provides when it comes to protecting lawful speech while combatting unlawful speech.

---

[46] Brittan Heller, *Combatting Terrorist-Related Content through AI and Information Sharing*, The Carr Center for Human Rights Policy, at 3 (April 26, 2019), https://www.ivir.nl/publicaties/download/Hash_sharing_Heller_April_2019.pdf.

[47] Angel Diaz & Laura Hecht Felella, *Double Standards in Social Media Content Moderation*, Brennan Center for Justice, (Aug. 4, 2021) at 18, https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation.

III. THE COURT SHOULD DEFER TO CONGRESS TO EVALUATE AND ENACT ANY POTENTIAL REFORMS TO SECTION 230.

The Court should also resist the siren songs of "reform" proponents who advocate the evisceration of current Section 230 protections as a means to speed the adoption of alternative regulatory regimes. To be sure, there is a lively debate on the merits of a wide range of regulatory proposals, including those that would apply to websites' use of algorithmic recommendation tools. *Amici* do not dispute that, despite the obvious benefits of many types of recommendation systems,[48] some have at times facilitated the spread

---

[48] Recommendation systems allow users to navigate the internet, providing users tools to sort through the vast realm of information that is available online. *See* Chris Meserole, *How Do Recommender Systems Work on Digital Platforms?*, Brookings (Sept. 21, 2022), https://www.brookings.edu/techstream/how-do-recommender-systems-work-on-digital-platforms-social-media-recommendation-algorithms/; *Section 230 of the Communications Decency Act*, Electronic Frontier Foundation, https://www.eff.org/issues/cda230. They can amplify individuals' stories and support small businesses in sharing their products. *See, e.g.*, Isabella Mercado, *The Black Lives Matter Movement: An Origin Story*, Underground Railroad Education Center, https://undergroundrailroadhistory.org/the-black-lives-matter-movement-an-origin-story/ (last accessed Dec. 22, 2022) (explaining the ability of social media algorithms and hashtags to amplify stories of racism and police brutality); *Small Businesses Blowing up on TikTok*, Zoella (June 19, 2021), https://zoella.co.uk/2021/06/19/small-businesses-blowing-up-on-tiktok/ (sharing stories of small business success on TikTok). They extend past traditional social media platforms, including search engines like Google and shopping sites such as Amazon, and allow users to discover new music, art, friends, and ideas. Francesco Ricci, Lior Rokach & Bracha Shapira, *Recommender Systems: Techniques, Applications, and Challenges*, IN RECOMMENDER SYSTEMS HANDBOOK, 3RD ED. (2021). Users have come to appreciate individualized recommend-

of misinformation and widely-reviled content.[49] As discussed above, transparency is also a legitimate topic of debate. Because websites rarely disclose sufficient information about how their recommendation systems work or what user data is being collected, users do not have the necessary information to make an informed choice over the content they get to see online. They are thus unable to protect themselves from potentially harmful content that is recommended by algorithms calibrated to prioritize user engagement.[50]

It may be necessary and wise to address these challenges through regulatory reform, but that cannot be done in this proceeding. Congress is already considering comprehensive Section 230 reform proposals,[51] including some that focus specifically on

---

ations and view them as a "feature" of a platform. Aurora Harley, *Individualized Recommendations: Users' Expectations & Assumptions,* Nielsen Norman Group (September 30, 2018), https://www.nngroup.com/articles/recommendation-expectations/.

[49] *See, e.g.,* Zeynep Tufecki, *YouTube, the Great Radicalizer*, N.Y. TIMES (Mar. 10, 2018) https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

[50] *Id.*

[51] Valerie C. Brannon & Eric N. Holmes, Cong. Rsch. Serv., *Section 230: An Overview,* R46751, 29-50 (2021); *see also* Quinta Jurecic, *The Politics of Section 230 Reform: Learning from FOSTA's Mistakes*, Brookings, March 1, 2022, https://www.brookings.edu/research/the-politics-of-section-230-reform-learning-from-fostas-mistakes; Search Results for 117th Congressional Session Hearings on Section 230, GOVINFO.GOV, https://www.govinfo.gov/ (search for "Section 230" and check the 117th congressional session filter on the congressional hearing database).

recommendation systems.[52] Issues that should be the subject of such reform proposals—including transparency, competition, equity, and procedural fairness—require exhaustive fact-finding and careful consideration of competing interests. This is quintessentially a legislative task, not a judicial one. Congress, not the judicial branch, is in the best position to design comprehensive reform that addresses all of these concerns while continuing to respect the free speech rights of website operators and users.

That legislative debate began long before this case, and it will continue if the court of appeals' judgment is affirmed. Imposing liability on platforms under the Anti-Terrorism Act or state tort law regimes for their recommendation systems will not guarantee the adoption of better regulation. The only guaranteed outcome will be the curtailment of free expression. The best foundation for future regulatory reform, including refinements to algorithmic recommendation systems, is the prevailing Section 230 jurisprudence that the court of appeals correctly applied in this case.

---

[52] *See* Citrus Policy Labs, *Section 230 Legislation Database*, https://citrispolicylab.org/section230/ (providing an up-to-date list of Congressional proposals related to Section 230).

CONCLUSION

For these reasons, the Court should affirm the Ninth's Circuit's judgment.

Respectfully submitted,

DAVID KAYE
EMILE G. AYOUB
INTL. JUSTICE CLINIC
U. OF CALIFORNIA, IRVINE
SCHOOL OF LAW
401 E. PELTASON DRIVE
IRVINE, CA 92697
(949) 824-0066
DKAYE@LAW.UCI.EDU

ROBERT P. LATHAM
COUNSEL OF RECORD
MARC FULLER
HANNAH WALSH
JACKSON WALKER LLP
2323 ROSS AVENUE, STE. 600
DALLAS, TX 75201
(214) 952-6000
BLATHAM@JW.COM

COUNSEL FOR AMICI CURIAE

JANUARY 19, 2023