

The Digital Services Act and the Brussels Effect on Platform Content Moderation

Dawn Carla Nunziato*

Abstract

The EU's latest regulation of social media platforms—the Digital Services Act (DSA)—will create tension and conflict with the U.S. speech regime applicable to social media platforms. The DSA, like prior EU regulations of social media platforms, will further instantiate the Brussels Effect, whereby EU regulators wield powerful influence on how social media platforms moderate content on the global scale. This is because the DSA's regulatory regime (with its huge penalties for noncompliance) will incentivize the platforms to skew their global content moderation policies toward the EU's instead of the U.S.'s balance of speech harms versus benefits. The Act's incentives for platforms to moderate harmful content, if implemented globally as is likely, will also create tension with recently enacted U.S. state laws like those adopted in Texas and Florida, and those proposed at the federal level, which prohibit platforms from moderating content in a viewpoint-discriminatory manner.

* The Pedas Family Endowed Professor of Intellectual Property and Technology Law, The George Washington University Law School. I am extremely grateful to David Markallo, Garrett Dowell, and Ken Rodriguez for their excellent and expert research and librarian assistance in connection with this Article. I am also very grateful to Todd Peterson for his insightful comments on this draft and to the editors at the *Chicago Journal of International Law*, especially Tori Keller, Miriam Kohn, and Michael Antosiewicz.

Table of Contents

| | |
|---|-----|
| I. Introduction | 117 |
| II. The DSA’s Substantive Content Moderation Requirements | 118 |
| A. Platform Liability Under EU and U.S. Speech Regimes | 118 |
| B. The Broad Sweep of EU Speech Restrictions | 119 |
| C. The DSA’s Expansion of the Brussels Effect..... | 120 |
| D. The DSA’s Tension with U.S. Laws | 123 |
| III. The DSA’s Procedural Content Moderation Requirements..... | 125 |
| IV. Conclusion..... | 127 |

I. INTRODUCTION

The European Union (EU) is at it again—engaging in extensive regulation of social media platforms in ways that will affect how the platforms are regulated, not just in the EU but the world over. The latest regulation—the broad-ranging Digital Services Act (DSA)—will create a host of tensions and outright conflicts with the United States (U.S.) speech regime applicable to social media platforms. The DSA, like other recent EU regulations of social media platforms, will also further enhance the Brussels Effect,¹ whereby European regulators strongly influence how social media platforms globally moderate content, which will incentivize the platforms to moderate much more (allegedly) harmful content than they have in the past. This extensive regulatory regime will incentivize platforms to skew their global content moderation policies toward the EU’s instead of the U.S.’s balance of speech harms and benefits. While the EU and its constituent nations generally prioritize protecting against dignitary, reputational, and societal harms more highly than absolute freedom of expression, and they generally hold platforms accountable for their role in facilitating harmful content, the U.S. takes the opposite approach.

The DSA will likely incentivize platforms to skew their content moderation policies toward the EU’s approach. This is because the DSA levies huge financial penalties for violating its provisions, including maximum fines of six percent of a platform’s annual worldwide turnover.² The DSA’s incentives for platforms to moderate harmful content, if implemented globally (as is likely), will also create tension with recently enacted U.S. state laws like those adopted in Texas and Florida, and proposed in federal legislation, which prohibit platforms from moderating content in a viewpoint-discriminatory manner. Finally, the DSA’s onerous procedural requirements imposed on platforms—including the Act’s extensive notice, right of appeal, and explanation requirements—will create tension, and in some cases conflict, with the procedural requirements imposed on platforms under U.S. state laws. These onerous procedural requirements would likely be considered unduly burdensome and violative of the First Amendment if implemented by the U.S. federal or state governments.

¹ See, e.g., ANU BRADFORD, *THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD* (2020) (arguing that the EU, by promulgating regulations that govern the international business environment and lead to the Europeanization of many important aspects of global commerce, shapes policy in areas such as online hate speech and data privacy, among others); see also Asha Allen & Ophélie Stockhem, *A Series on the EU Digital Services Act: Tackling Illegal Content Online*, CTR. FOR DEMOCRACY & TECH. (Aug. 22, 2022), <https://perma.cc/HG99-U28Q> (stating that the DSA is “set to be a legislative driving force, with the *Brussels Effect* in its full stride.”).

² Commission Regulation 2022/2065, *Shaping Europe’s Digital Future: The Digital Services Act Package*, 2022 J.O. (277), 1, 2, art. 52(2)–(3) [hereinafter DSA], <https://perma.cc/J6AF-BZHE>.

II. THE DSA'S SUBSTANTIVE CONTENT MODERATION REQUIREMENTS

A. Platform Liability Under EU and U.S. Speech Regimes

The DSA, which entered into force on November 16, 2022, and became operational on very large social media platforms in early 2023, maintains the conditional liability regime imposed over two decades ago under the EU E-Commerce Directive.³ Under this prior regulation, platforms were only liable for hosting harmful content if they had notice of such content and failed to remove it. The DSA generally provides that platforms are not liable for the third-party content they host, provided they act expeditiously upon notice of such allegedly illegal content.⁴ Once a platform receives a credible notification that it is hosting illegal content on its site, this triggers the duty for it to “expeditiously” remove such content or risk liability.⁵ Such notice can be issued by a number of different entities. The DSA contemplates a regime in which individuals, country-level authorities, and “trusted flaggers”—which are private, non-governmental entities or public entities with expertise of some type—can identify content⁶ that they believe to be illegal⁷ under EU country-specific laws.⁸ The statutory text states that platforms hosting third party content

shall not be liable for the information stored at the request of a recipient of the service, on condition that the provider: (a) does not have actual knowledge of illegal activity or illegal content and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or illegal content is apparent; or (b) *upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the illegal content.*⁹

Thus, upon receiving notice of allegedly illegal content hosted on their platform, providers must expeditiously remove such content—or risk liability.¹⁰

³ *Id.*; Council Directive 2000/31/EC, Directive on Electronic Commerce, 2000 O.J. (L 178) [hereinafter EU E-Commerce Directive].

⁴ DSA art. 22.

⁵ EU E-Commerce Directive, *supra* note 3, art. 13(1)(e).

⁶ DSA art. 22.

⁷ As defined in the DSA, “‘illegal content’ means any information that . . . is not in compliance with Union law or the law of any Member State . . . irrespective of the precise subject matter or nature of that law. . . .” *Id.* art. 3(h).

⁸ *See id.* art. 16(1). “Providers of hosting services shall put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content. Those mechanisms shall be easy to access, user-friendly, and allow for the submission of notices exclusively by electronic means.” *Id.*

⁹ *Id.* art. 6(1)(a)–(b) (emphasis added).

¹⁰ *Id.* *See generally* Joan Barata, *Europe’s Tech Regulations May Put Free Speech at Risk*, (May 18, 2022), <https://perma.cc/L34L-LU39> (arguing that the DSA will incentivize the platforms to overly moderate content in a manner that will put free speech at risk).

The “Notice and Action” provisions of the DSA stand in sharp contrast to the comparable U.S. regime under § 230(c) of the Communications Decency Act (CDA). Section 230(c), which is the main piece of legislation applicable to general platform liability in the U.S., immunizes platforms from many forms of liability for hosting third party content. In contrast to the DSA, § 230(c) imposes no conditions on platforms to receive immunity from liability.¹¹ Since the CDA’s passage in 1996, § 230(c) has been consistently interpreted by U.S. courts to provide broad immunity to platforms for hosting and facilitating a wide range of illegal content—from defamatory speech to hate speech to terrorist and extremist content.¹² Notice of illegal content is irrelevant to such immunity.¹³ Thus, even if a platform like YouTube is repeatedly and clearly notified that it is hosting harmful content (such as ISIS propaganda videos), the platform remains immune from liability for hosting such harmful content.¹⁴

B. The Broad Sweep of EU Speech Restrictions

Many EU countries have speech regimes under which speech is deemed illegal according to widely different (and compared to the U.S., vastly less protective) standards. The DSA will consequently push platforms to quickly remove a wide swath of allegedly illegal content or risk losing immunity from liability. Several categories of speech are illegal under European law but would be protected in the U.S. under the First Amendment—some for better, some for worse. For example, several EU countries restrict Holocaust denial and minimization as well as glorification of Nazi ideology. In Germany and other EU member states,¹⁵ Holocaust denial and glorification are illegal. German law, which is similar to that of several other EU nations, makes it a crime to deny or downplay an act committed under the rule of National Socialism¹⁶ or to glorify or

¹¹ 47 U.S.C. § 230(c) (2018) [hereinafter § 230], <https://perma.cc/8R2V-NU3P>.

¹² See generally David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373 (2010); Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45 (2020); *Gonzalez v. Google LLC*, 2 F.4th 871 (9th Cir. 2021).

¹³ See § 230(c).

¹⁴ Platforms’ immunity from liability under § 230 was recently before the U.S. Supreme Court, although the Court declined to apply § 230 to the complaint because it held that the complaint did not state a plausible claim for relief. *Gonzalez v. Google LLC*, 143 S. Ct. 1191 (2023).

¹⁵ Many EU countries have analogous statutes to those in Germany that criminalize Holocaust denial. See generally Piotr Bąkowski et al., *Holocaust Denial in Criminal Law: Legal Frameworks In Selected EU Member States*, EUR. PARLIAMENTARY RSCH. SERV. (Jan. 2022) (describing similar statutes in Belgium, France, Italy, Greece, and other member states).

¹⁶ STRAFGESETZBUCH [STGB] [Penal Code], § 130, para. 3, <https://perma.cc/RT22-B6EA> (Ger.); see also BRITTAN HELLER & JORIS VAN HOBOKEN, TRANSATLANTIC WORKING GRP., FREEDOM OF EXPRESSION: A COMPARATIVE SUMMARY OF UNITED STATES AND EUROPEAN LAW 8–9 (May 3, 2019), <https://perma.cc/D6UC-49V4>.

justify National Socialist tyranny and arbitrary rule¹⁷—speech that is protected under the First Amendment. Such criminal prohibitions carry over to the online realm and to a comparable notice and action regime in Germany through the Network Enforcement (NetzDG) Act.¹⁸ Yet, illegal content in European countries also includes categories of content that would be deemed valuable and are protected under the U.S. free speech regime.¹⁹ These include French laws prohibiting criticism and parody of the president, such as by depicting him as Hitler (which was recently held to violate French insult and public defamation laws²⁰); Austrian²¹ and Finnish²² laws that criminalize blasphemy; Hungarian laws that prohibit a range of pro-LGBTQ+ content accessible to minors;²³ and laws in various European countries (Turkey, France, and Russia among them) that ban certain types of offensive humor.²⁴ The DSA’s Notice and Action regime, which allows entities in the EU to flag content that is illegal under their country’s laws and requires the platforms to expeditiously remove such content, will likely incentivize platforms to remove a vast amount of content that would be deemed protected—and indeed valuable—under other countries’ speech laws, including the U.S.’s, such as political criticism, satire, parody, and pro-LGBTQ+ content.

C. The DSA’s Expansion of the Brussels Effect

The DSA, like other European codes before it, will likely further instantiate the Brussels Effect, whereby platforms shape their globally applicable content moderation policies and practices to conform to the dictates of EU regulations. Such an effect has already been observed in conjunction with platforms’ compliance with several recent EU regulatory schemes. For example, in 2016, the EU incentivized platforms to “voluntarily” adopt the EU Code of Conduct on

¹⁷ STGB, § 130, para. 4; *see also* HELLER & VAN HOBOKEN, *supra* note 16.

¹⁸ *See* Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [NetzDG] [Network Enforcement Act], Sept. 1, 2017, BGBl. I at 3352 (Ger.).

¹⁹ *See, e.g.*, Jacob Mchangama, *Op-Ed: Don’t Be Too Tempted by Europe’s Plan to Fix Social Media*, L.A. TIMES (Dec. 23, 2022), <https://perma.cc/Q3FD-Q6PM>.

²⁰ *French Billboard Owner Fined €10,000 for Depicting Macron as Hitler in Poster Protesting COVID Rules*, EURONEWS (Sept. 17, 2021), <https://perma.cc/8M7R-9HKD>.

²¹ The “disparagement of religious doctrines” is illegal under Article 188 of the Austrian Criminal Code. *See, e.g.*, Can Yeginsu & John Williams, *Criminalizing Speech to Protect Religious Peace? The ECtHR Ruling in E.S. v. Austria*, JUST SEC. (Nov. 28, 2018), <https://perma.cc/DJ8P-QLUV>.

²² For a discussion of Finnish law prohibiting blasphemy, *see* U.S. DEP’T OF STATE, OFF. OF INT’L RELIGIOUS FREEDOM, 2021 REPORT ON INTERNATIONAL RELIGIOUS FREEDOM: FINLAND, (2022), <https://perma.cc/ZD3P-YN47/>.

²³ Jennifer Rankin, *Hungary Passes Law Banning LGBT Content in Schools or Kids’ TV*, GUARDIAN (June 15, 2021), <https://perma.cc/VVG3-2HBJ>.

²⁴ *See, e.g.*, Alberto Godioli et al., *Laughing Matters: Humor, Free Speech and Hate Speech at the European Court of Human Rights*, 35 INT’L J. SEMIOTICS L. 2241, 2243–46 (Nov. 2, 2022), <https://perma.cc/7U5Q-SD9C>.

Countering Illegal Hate Speech Online.²⁵ Under the Code, platforms agreed to remove various types of illegal hate speech within 24 hours of receiving notice of such content.²⁶ The platforms agreed to adopt this Code of Conduct in part to forestall regulation by the EU. As Danielle Citron explains, after terrorist attacks in Paris and Brussels in 2015, European regulators threatened the platforms with extensive regulations unless the platforms undertook meaningful measures to effectively police and remove hate speech and extremist speech.²⁷ Faced with this prospect of regulation, four major platforms—Facebook, Microsoft, Twitter, and YouTube—entered into a voluntary agreement with the EU.²⁸ The Code that was adopted provided that, while hate speech would continue to be regulated by the states under applicable criminal law, “this work must be complemented with actions geared at ensuring that illegal hate speech online is expeditiously acted upon by online intermediaries and social media platforms, upon receipt of valid notification, in an appropriate time-frame.”²⁹ The Code also *indirectly* encouraged the platforms to conform their globally-applicable terms of service to the contours of European hate speech regulation. In particular, in adopting and agreeing to the Code, the platforms agreed to:

- [1] [H]ave in place clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content. . . .
- [2] [H]ave in place Rules or Community Guidelines clarifying that they prohibit the promotion of incitement to violence and hateful conduct. . . .
- [3] Upon receipt of a valid removal notification, . . . review such requests against their rules and community guidelines and where necessary national laws transposing the Framework Decision 2008/913/JHA, with dedicated teams reviewing requests. . . .

²⁵ Under this Code, the major platforms commit to put in place a system whereby they can receive notifications of alleged illegal hate speech on their platforms, to review the majority of valid notifications for removal of illegal hate speech in less than twenty-four hours, and remove or disable access to such content, if necessary. This Code has now been adopted by the major social media platforms, including Facebook, Microsoft, Twitter, YouTube, Instagram, Snapchat, TikTok, and LinkedIn. See *The EU Code of Conduct on Countering Illegal Hate Speech Online*, EUR. COMM’N (June 2022), <https://perma.cc/DFK6-VMSQ>.

²⁶ Not surprisingly, such regulatory mechanisms have been heavily criticized on the basis that they circumvent rule of law safeguards, delegate determinations of illegality to private entities, and lead to over-moderation of content. See *Letter to European Commissioner on Code of Conduct for “Illegal” Hate Speech Online*, CTR. FOR DEMOCRACY & TECH. (June 3, 2016), <https://perma.cc/7VNW-PLA8>.

²⁷ See, e.g., Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035 (2018).

²⁸ *Id.* at 1037–38.

²⁹ EUR. COMM’N, *supra* note 25.

[4] [R]eview the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.³⁰

Under Framework Decision 2008/913/JHA, in turn, it is a crime to “publicly incit[e] to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin” or to “publicly condon[e], deny[] or grossly trivialis[e] crimes of genocide, crimes against humanity and war crimes.”³¹

Although it was initially contemplated that the EU Code of Conduct would apply to content moderation only within the EU, the platforms unsurprisingly found it easier to comply with the Code’s requirements globally.³² As Citron argues:

Companies’ presumptive deletion of hate speech [under the EU Code of Conduct] is bound to have a global impact because [Terms of Service (TOS)] agreements are involved rather than court orders or other forms of legal process. TOS agreements are typically the same across the globe. Thus, decisions to delete or block content as TOS violations mean content will be deleted or blocked everywhere the platform is viewed. . . . Terms of service related to hateful and extremist speech apply to the Tech Companies’ global operations. Because the [EU Code of Conduct] is operationalized through terms of service, a presumption of removal would mean worldwide removal.³³

A similar Brussels Effect likely flowed from the recently adopted EU Code of Practice on Disinformation.³⁴

In short, the DSA’s substantive content moderation and notice and take down provisions will likely incentivize the platforms to remove large swaths of content—including political speech, criticism of political figures, parody, and pro-LGBTQ+ speech—that may be flagged by private entities as illegal under EU countries’ laws. These types of flagged content would no doubt be protected in the U.S. under the First Amendment. And the platforms will likely alter their globally applicable terms of service and content moderation guidelines in response to the DSA’s mandates in ways that will be speech-restrictive worldwide.

³⁰ *Id.*

³¹ Council Framework Decision 2008/913/JHA of Nov. 28, 2008 on Combating Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law, 2008 O.J. (L 328) art. 1(1)(a), (c).

³² *See, e.g.,* Citron, *supra* note 27 (arguing that the platforms modified their globally applicable terms of services in response to pressure from EU lawmakers and in an attempt to stave off regulation by the EU).

³³ *Id.* at 1055–56.

³⁴ As David Morar suggests, existing EU codes, like the EU Code of Practice on Disinformation—and perhaps the EU Code of Conduct on Illegal Hate Speech—will likely be considered applicable/enforceable Codes of Conduct under the DSA. *See* David Morar, *The Digital Services Act’s Lesson for U.S. Policymakers: Co-regulatory Mechanisms*, BROOKINGS INST. (Aug. 23, 2022), <https://perma.cc/QV8Y-TVM3>.

D. The DSA's Tension with U.S. Laws

Social media platforms' global modification of their content moderation policies, as incentivized by the DSA, may generate tension with recently enacted U.S. laws that prohibit platforms from moderating content in a manner that discriminates based on viewpoint. In 2021, Texas adopted House Bill 20 (HB 20), which heavily regulates large platforms' content moderation practices.³⁵ HB 20 prohibits large social media platforms (those with 100 million users in a calendar month) from censoring expression based on viewpoint,³⁶ where "censor" is defined as "to block, ban, remove, deplatform, demonetize, de-boost, restrict, deny equal access or visibility to, or otherwise discriminate against expression."³⁷ The law, which was recently upheld by the Fifth Circuit against the platforms' constitutional challenge (but is currently subject to a stay by the Supreme Court),³⁸ prohibits platforms from moderating content in a way that discriminates based on viewpoint. Under the Texas law, a platform's removal of content that, for example, denies or questions the extent of the Holocaust, or that is critical of immigration policies or immigrants or COVID-19 vaccines, would likely be considered illegal viewpoint discrimination in content moderation. Yet, a platform's *refusal* to remove such content upon notice would likely violate the terms of the DSA. As the Center for Democracy and Technology argues:

[Under HB 20,] providers will be forced to rescind or not enforce their viewpoint-based content policies. The effect will be the unchecked proliferation of content that, for example, . . . demeans Christians, Jews, Muslims and people of other religious faiths. . . . Faced with the prospect of

³⁵ Regulated platforms are those with "more than 50 million active users in the United States in a calendar month." TEX. BUS. & COM. CODE ANN. § 120.002 (West 2021), <https://perma.cc/J6BR-DBNP>.

³⁶ TEX. CIV. PRAC. & REM. § 143A.002 (West 2021), <https://perma.cc/EAD6-PHCD>.

³⁷ *Id.* § 143A.001, <https://perma.cc/EAD6-PHCD>.

³⁸ The Fifth Circuit upheld these provisions against the platforms' First Amendment challenge. *NetChoice LLC v. Paxton*, 49 F.4th 439 (5th Cir. 2022). The Fifth Circuit held that H.B. 20 correctly classified the platforms as common carriers and that the law's prohibition against viewpoint-discriminatory content moderation was consistent with historical regulation of common carriers. *Id.* at 469–73. The court held, in the alternative, that even if the platforms were not properly viewed as common carriers and were instead viewed as entities that engage in protected speech when they moderate content, the law was nonetheless constitutional because it would be subject only to an intermediate level of scrutiny and would be upheld applying such scrutiny because it was content neutral and because the state's interest in promoting the free exchange of ideas from a variety of sources was a sufficiently important government interest and the law was the least speech-restrictive means of advancing this interest. *Id.* at 480–88. This Fifth Circuit's decision stands in contrast to the Eleventh Circuit's decision in a similar case which struck down the provisions of Florida Senate Bill 7072 that limited the platforms' ability to engage in deplatforming, censorship, shadow-banning, or prioritization and prohibited them from deplatforming or restricting the content of political candidates or journalistic enterprises. *NetChoice v. Florida*, 34 F.4th 1196 (11th Cir. 2022).

many such suits under HB 20, platforms may decide that they face less legal risk if they decline to moderate any content³⁹

A platform's decision not to moderate any content, however, would render it in violation of the DSA.

Proposed federal legislation in the U.S. would also prohibit platforms from engaging in viewpoint discrimination in their content moderation practices. If enacted, such prohibitions would further conflict with the DSA's provisions. For example, the proposed Disincentivizing Internet Service Censorship of Online Users and Restrictions on Speech and Expression Act (DISCOURSE Act) would also prohibit platforms from engaging in viewpoint discrimination, whether human-moderated or algorithmic.⁴⁰ The Act would censure social media companies that “engage[] in a content moderation activity”⁴¹ or “a pattern or practice of content moderation activity that reasonably appears to express, promote, or suppress a discernible viewpoint for a reason that is not protected [under the Act].”⁴² Additionally, the Act would amend § 230 to specify that the CDA's limitation of liability does not apply unless the platforms comply with the Act's terms.⁴³ Section 230(c)(2) currently provides that no platform “shall be held liable on account of any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, *or otherwise objectionable*, whether or not such material is constitutionally protected.”⁴⁴ The DISCOURSE Act would condition this limitation of liability on an objective reasonableness standard and limit the permissible bases for which a platform could restrict access to content.⁴⁵ Under the Act, in order to continue enjoying § 230's limitation on liability, a platform would only be permitted to restrict access to content if it had “an objectively reasonable belief” that the content is “obscene, lewd, lascivious, filthy, excessively violent, harassing, *promot[es] terrorism,*” is determined to be “*unlawful,*” or promotes “*self-harm.*”⁴⁶

³⁹ Motion for Leave to File Brief Amici Curiae and Brief for Center for Democracy & Technology et al. as Amici Curiae Supporting Applicants at 15–17, *NetChoice v. Paxton* 49 F.4th 439 (5th Cir. 2022) (No. 21A720), 2022 WL 2376280.

⁴⁰ DISCOURSE Act, S. 2228, 117th Cong. (2021), <https://perma.cc/8LXU-T9ND>.

⁴¹ *See id.* § 2(a).

⁴² *Id.* § 2(a)(2)(B).

⁴³ *See* Press Release, Senator Marco Rubio, Rubio Introduces Sec 230 Legislation to Crack Down on Big Tech Algorithms and Protect Free Speech (June 24, 2021), <https://perma.cc/C5CE-U52E>.

⁴⁴ § 230(c)(2)(a) (emphasis added).

⁴⁵ *See* DISCOURSE Act, S. 2228 § 2(b).

⁴⁶ § 230(c)(2)(a) (emphasis added). This section includes a religious liberty clause, which states explicitly that (c)(2) does not extend liability protections to decisions that restrict content based on their religious nature. *See* DISCOURSE Act, S. 2228 § 2(c).

In addition, Congress's proposed 21st Century Foundation for the Right to Express and Engage in Speech Act would impose nondiscrimination, transparency, and due process requirements on—and would limit the immunity enjoyed by—social media platforms.⁴⁷ Specifically, the Act would prohibit platforms from making or giving “any undue or unreasonable preference or advantage to any particular person, class of persons, political or religious groups or affiliation, or locality.”⁴⁸ If a platform discriminated against content in violation of the Act's provisions, it would be liable for suit by affected private individuals and by states' attorneys general.⁴⁹ Further proposed federal legislation, such as the Curbing Abuse and Saving Expression In Technology (CASE-IT) Act, would regulate viewpoint discrimination by platforms by removing § 230 immunity for market-dominant platforms that “make[] content moderation decisions pursuant to policies or practices that are not reasonably consistent with the First Amendment.”⁵⁰ The CASE-IT Act would also create a private right of action for users negatively affected by such decisions.⁵¹

In sum, both state and proposed federal legislation in the U.S. would prohibit platforms from engaging in viewpoint discrimination in their content moderation practices, and such prohibitions conflict with the DSA's provisions mandating the removal upon notice of certain types of allegedly illegal speech in order to preserve a platform's immunity from liability.

III. THE DSA'S PROCEDURAL CONTENT MODERATION REQUIREMENTS

The DSA also imposes extensive procedural and due process requirements on platforms, some of which would likely be unconstitutional under the First Amendment. Procedural requirements regarding content moderation are generally scrutinized more deferentially under the First Amendment than substantive requirements. The DSA's procedural requirements, however, impose quite burdensome requirements on the platforms, which would create tension and conflict with the procedural requirements imposed on platforms by recent U.S. state laws. These procedural requirements would likely be viewed as unconstitutional if viewed through the lens of First Amendment law.

First, the DSA requires platforms to provide adversely affected users with notice, explanation, and an opportunity to appeal every adverse content moderation decision. Whenever a platform engages in an act of content

⁴⁷ 21st Century Foundation for the Right to Express and Engage in Speech Act, S. 1384, 117th Cong. § 2(a) (2021), <https://perma.cc/V33C-DLYP>.

⁴⁸ *Id.* § 232(c)(1)(C).

⁴⁹ *Id.*

⁵⁰ CASE-IT Act, H.R. 573, 118th Cong. § 2(a)(4)(A)(ii) (2022).

⁵¹ *Id.*

moderation—in response to receiving notice of illegal content, or on its own initiative—the platform will be required to provide the user with a detailed explanation and an opportunity to appeal the decision, both internally with the platform and externally.⁵² In particular, when the platform deems that a user has posted illegal content, the platform must provide the user with a “statement of reasons.”⁵³ This statement must indicate whether the content has been removed or deprioritized and whether the user has been suspended or terminated from the service; whether the content was removed because of a complaint; why the content was deemed illegal under the applicable law or as a violation of the platform’s terms of service; and the redress available to the recipient.⁵⁴ An adversely affected user must also be given the right to appeal the platform’s decision to the platform for at least six months following the decision,⁵⁵ and the platform must decide whether to reverse the restriction “without undue delay” and must offer other methods of redress, including out-of-court dispute resolution.⁵⁶ While such a burdensome requirement could ordinarily be expected to *disincentivize* the platforms’ acts of content moderation, in conjunction with the DSA’s conditional immunity regime and severe financial penalties for noncompliance discussed above, such a disincentivizing effect is unlikely.

Second, although entities like social media platforms may constitutionally be subject to limited due process, disclosure, and transparency requirements, if these requirements are unduly burdensome, they run afoul of the First Amendment. In *Zauderer v. Office of Disciplinary Counsel of the Supreme Court of Ohio*⁵⁷ and its progeny, for example,⁵⁸ the Supreme Court has held that, while entities can be constitutionally subject to certain disclosure and transparency requirements, such requirements are unconstitutional if they are overly burdensome (or require the disclosure of controversial, not purely factual, information). Consistent with this line of precedent, the Eleventh Circuit recently struck down the provisions of Florida Senate Bill 7072 that require platforms to provide notice and justification for each of their content moderation actions on the grounds that these notice and

⁵² See DSA art. 17.

⁵³ *Id.*

⁵⁴ *Id.* art. 17(1), (3).

⁵⁵ *Id.* art. 20(1).

⁵⁶ *Id.* art. 20(4)–(5).

⁵⁷ As the Supreme Court held in *Zauderer*, regulations that merely require an entity to disclose factual and uncontroversial information are constitutional and do not violate the entity’s First Amendment rights. 471 U.S. 626 (1985).

⁵⁸ See *National Institute of Family and Life Advocates v. Becerra*, 138 S. Ct. 2361, 2366 (2018) (holding that the disclosure requirement mandated by the state was unconstitutional because it went beyond “purely factual and uncontroversial information about the terms under which . . . services will be available,” among other reasons) (citations omitted).

justifications provisions were unduly burdensome.⁵⁹ Under this standard, it is likely that the DSA’s procedural content moderation requirements would be viewed as imposing unduly burdensome and therefore unconstitutional requirements on the platforms under the First Amendment.

In addition, the DSA’s procedural requirements will conflict with the procedural requirements imposed under Texas HB 20 and Florida SB 7072, and with similar requirements in laws proposed at the federal level.⁶⁰ The Texas law, which was upheld by the Fifth Circuit,⁶¹ requires platforms to notify users when they remove content, provide users with the opportunity to appeal such decisions, provide an “easily accessible” system for users to submit complaints about content removals, and act on these complaints within 48 hours. The DSA’s and the Texas law’s procedural requirements regarding content moderation will impose conflicting, burdensome, and perhaps insurmountable hurdles on the platforms’ acts of content moderation.

IV. CONCLUSION

The recent adoption of the DSA will bring with it tension and in some cases outright conflict with the U.S. free speech regime applicable to social media platforms. The DSA, like other recent EU regulations of social media platforms, will further instantiate the Brussels Effect, whereby European regulators will continue to strongly influence how social media platforms globally moderate content and will incentivize the platforms to moderate much more (allegedly) harmful content than they have in the past. This extensive regulatory regime will incentivize the platforms to skew their global content moderation policies toward the EU’s (instead of the U.S.’s) approach to balancing the costs and benefits of free speech—especially given the DSA’s huge financial penalties for violating its provisions.

The DSA’s incentives for platforms to moderate harmful content, if implemented globally as is likely, will create tensions with recently enacted U.S. state laws and proposed federal laws, which prohibit platforms from moderating content in a viewpoint-discriminatory manner. However, because these provisions of state laws are unlikely to withstand First Amendment scrutiny and because the proposed federal laws are unlikely to be enacted (or to be upheld by courts if

⁵⁹ *NetChoice v. Florida*, 34 F.4th at 1222–23, 1230–31. In contrast, the Fifth Circuit upheld the Texas law’s transparency reporting requirements and requirements. *NetChoice v. Paxton*, 49 F.4th at 439.

⁶⁰ See, e.g., Dawn Carla Nunziato, *Protecting Free Speech and Due Process Values on Dominant Social Media Platforms*, 73 HASTINGS L.J. 1255 (2022) (discussing similar provisions of proposed legislation like the DISCOURSE Act, the 21st Century FREE Speech Act, the PRO-SPEECH Act, and the PACT Act).

⁶¹ The Fifth Circuit concluded that all of these provisions were constitutional under the lower level of constitutional scrutiny applicable to disclosure requirements under *Zauderer*. See *Paxton*, 49 F.4th at 485–88.

enacted), the EU will likely continue to exercise unchecked power and influence over the platforms' global content moderation practices. In short, the DSA's substantive regulatory regime—in conjunction with its huge penalties for noncompliance—will incentivize the platforms to continue to skew their global content moderation policies toward the EU's approach, for better or for worse.

Finally, the DSA's onerous procedural requirements imposed on platforms—including the Act's extensive notice, right of appeal, and explanation requirements—will require platforms to greatly expand the resources devoted to content moderation and to expend far greater resources on complying with such requirements.

In short, the DSA will result in the continuation of the trend of EU regulations incentivizing platforms to skew their global content moderation policies toward Europe's balance of speech harms and benefits—instead of the U.S.'s balance—and will solidify the EU's position of being the global driver of internet content moderation policies.