

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

X CORP.,

Plaintiff,

-against-

LETITIA JAMES, in her official capacity, as
Attorney General of New York,

Defendant.

Case No. 1:25-cv-05068-JPC

**X CORP.'S MEMORANDUM OF LAW IN OPPOSITION TO DEFENDANT'S
MOTION TO DISMISS**

TABLE OF CONTENTS

INTRODUCTION	1
FACTUAL BACKGROUND.....	5
I. S895B’s Statutory Scheme and its Application to Plaintiff.....	5
A. Required Disclosure of Terms of Service (Terms of Service Requirement)	6
B. Terms of Service or Content Category Report	7
C. Penalty Provision	9
II. Section 1102(1)(c)’s Content Categories.....	9
III. Plaintiff’s Current Content-Moderation Efforts.....	10
IV. S895B, AB 587, and the New York Legislature’s Refusal to Reconsider S895B Following the Ninth Circuit’s Decision in <i>X Corp.</i>	11
ARGUMENT	12
I. The Content Category Report Provisions Violate the First Amendment as Applied to Plaintiff and On Their Face.....	12
A. <i>Volokh</i> and <i>X Corp.</i> Require Denial of Defendant’s Motion.....	12
B. Neither Intermediate Scrutiny nor <i>Zauderer</i> Applies	17
1. The Content Category Report Provisions Compel Noncommercial Speech, Precluding Intermediate Scrutiny and <i>Zauderer</i>	17
2. The Content Category Report Provisions Compel Disclosures that Are Neither “Purely Factual” Nor “Uncontroversial,” Precluding Application of <i>Zauderer</i>	19
C. Strict Scrutiny Applies Because the Content Category Report Provisions Are a Content- and Viewpoint-Based Regulation of Speech	22
D. The Content Category Report Provisions Fail Any Level of Scrutiny	24
1. Strict and Intermediate (<i>Central Hudson</i>) Scrutiny	24
2. <i>Zauderer</i>	26
E. Plaintiff Meets the Requirements for a Facial Challenge	28
II. The Content Category Report Provisions Violate and Are Preempted by 47 U.S.C. §230(c)(2)(A)	29
CONCLUSION.....	30

TABLE OF AUTHORITIES

Cases	Page(s)
<i>Americans for Prosperity Found. v. Bonta</i> , 594 U.S. 595 (2021).....	28
<i>Bad Frog Brewery, Inc. v. N.Y. State Liquor Auth.</i> , 134 F.3d 87 (2d Cir. 1998).....	17, 18
<i>Bolger v. Youngs Drug Prods. Corp.</i> , 463 U.S. 60 (1983).....	18
<i>CompassCare v. Hochul</i> , 125 F.4th 49 (2d Cir. 2025)	21
<i>Evergreen Ass’n, Inc. v. City of N.Y.</i> , 740 F.3d 233 (2d Cir. 2014).....	24
<i>Iancu v. Brunetti</i> , 588 U.S. 388 (2019).....	23
<i>Ibanez v. Fla. Dep’t of Bus. and Pro. Regul., Bd. of Acct.</i> , 512 U.S. 136 (1994).....	24, 25
<i>Int’l Dairy Foods Ass’n v. Amestoy</i> , 92 F.3d 67 (2d Cir. 1996).....	27
<i>Milavetz, Gallop & Milavetz, P.A. v. United States</i> , 559 U.S. 229 (2010).....	27
<i>Moody v. NetChoice, LLC</i> , 603 U.S. 707 (2024).....	17, 23, 28
<i>Nat’l Inst. of Fam. & Life Advocs. v. Becerra</i> , 585 U.S. 755 (2018).....	26
<i>Nat’l Retail Fed’n v. James (“NRF”)</i> , 2025 WL 2848212 (S.D.N.Y. Oct. 8, 2025).....	20
<i>Nat’l Rifle Ass’n of Am. v. Vullo</i> , 602 U.S. 175 (2024).....	21, 24, 30
<i>NetChoice, LLC v. Att’y Gen., Fla. (“NetChoice (Fla.)”)</i> , 34 F.4th 1196 (11th Cir. 2022)	15–16, 27
<i>NetChoice, LLC v. Paxton (“NetChoice (Tex.)”)</i> , 49 F.4th 439 (5th Cir. 2022)	15

<i>Pleasant Grove City, Utah v. Summum</i> , 555 U.S. 460 (2009).....	22
<i>Riley v. Nat’l Fed’n of the Blind of N.C., Inc.</i> , 487 U.S. 781 (1988).....	19
<i>Rosenberger v. Rector & Visitors of Univ. of Va.</i> , 515 U.S. 819 (1995).....	23
<i>Safelite Grp., Inc. v. Jepsen</i> , 764 F.3d 258 (2d Cir. 2014).....	24
<i>Smith v. California</i> , 361 U.S. 147 (1959).....	23, 24
<i>Volokh v. James</i> , 656 F. Supp. 3d 431 (S.D.N.Y. 2023).....	28
<i>Volokh v. James</i> , 148 F.4th 71 (2d Cir. 2025)	<i>passim</i>
<i>X Corp. v. Bonta</i> , 116 F.4th 888 (9th Cir. 2024)	<i>passim</i>
<i>Zauderer v. Off. of Disciplinary Couns.</i> , 471 U.S. 626 (1985).....	<i>passim</i>
Statutes	
47 U.S.C. §230.....	5, 29, 30
N.Y. Gen. Bus. Law §394-ccc (“Hateful Conduct Law”)	<i>passim</i>
N.Y. Gen. Bus. Law §§1100–1104 (“S895B”).....	<i>passim</i>
Other Authorities	
X Transparency, <i>Global Transparency Report</i> (H2 2024), available at https://tinyurl.com/5cfvw7wu	11
Meta, <i>Community Standards Enforcement Report</i> (Q2 2025), available at https://tinyurl.com/2vxuhvv4	11
<i>Rules</i> , X, https://help.x.com/en/rules-and-policies/x-rules	25

INTRODUCTION

This case concerns a constitutional and legal challenge to certain provisions of a New York statute, “S895B,” N.Y. Gen. Bus. Law §§1100–1104, that are *identical* in all material respects to those of a California law (Assembly Bill “AB” 587) that the Ninth Circuit struck down on First Amendment grounds after holding that the law triggered and could not survive strict scrutiny. *X Corp. v. Bonta*, 116 F.4th 888 (9th Cir. 2024). Significantly, in *Volokh v. James*, 148 F.4th 71 (2d Cir. 2025), *certified questions accepted*, 267 N.E.3d 1245 (2025), the Second Circuit recently endorsed the Ninth Circuit’s reasoning in expressly holding that laws that are analogous to AB 587 trigger and fail heightened scrutiny. Defendant’s Motion to Dismiss (Dkt. 29) (“Mot.”) is based entirely on arguments that were considered and rejected by the courts in both *Volokh* and *X Corp.* and must be denied.

S895B requires social media platforms, like X, to make disclosures specifically about certain highly controversial subject matters—“hate speech,” “racism,” “extremism,” “radicalization,” “disinformation,” “misinformation,” “harassment,” and “foreign political interference.” The challenged provisions (the “Content Category Report Provisions”)¹ force covered platforms to convey their opinions and policy views about whether and how to define and moderate those controversial categories of speech, and thus are different in character and kind from laws that require platforms to disclose their existing terms of service or content-moderation policies without reference to categories of controversial speech.

Specifically, the Content Category Report Provisions force platforms to disclose—under threat of financial penalties and lawsuits for injunctive relief, §1103—(i) “whether” they define the highly controversial content categories and, “if so,” how, §1102(1)(c); (ii) a “detailed

¹ Sections 1102(1)(c), 1102(1)(d)(i), and 1102(1)(e).

description” of how the platform’s “content moderation practices” are “intended to address” those categories, §1102(1)(d)(i); and (iii) statistics showing how often the platform “flagged,” “actioned,” “removed,” “demonetized,” and “deprioritized” content “belonging to any of the categories,” §1102(1)(e). Plaintiff does not challenge the remaining provisions of S895B—those not specifically tied to the controversial content categories selected by the State.

This focus on controversial content categories is legally significant; it makes the challenged provisions content- and viewpoint-based, thus requiring application of strict scrutiny. But Defendant never comes to grips with this aspect of the Content Category Report Provisions, as demonstrated by Defendant’s repeated arguments that the law must be upheld because it merely requires social media companies to disclose their “terms of service” and “content moderation policies,” which “constitute[] commercial speech” subject to a less-rigorous standard of review under *Zauderer v. Off. of Disciplinary Couns.*, 471 U.S. 626 (1985). *E.g.*, Mot. at 4, 9–10. The Motion barely even acknowledges the specific content categories—perhaps because Defendant has no valid argument that the Content Category Report Provisions are constitutional in light of them.

That the content categories in the challenged provisions are central to the applicable standard of review is not a novel concept. Both the Second Circuit and the Ninth Circuit focused on the content categories in concluding that the law here is both subject to heightened scrutiny and unconstitutional. Defendant never confronts the impact of those decisions on this case either.

As noted above, S895B is *identical* in all materials respects to AB 587,² which the Ninth Circuit struck down under strict scrutiny. *X Corp.*, 116 F.4th 888. In *X Corp.*, the Ninth Circuit

² See Dkt. 21-1 (redline between AB 587 and S895B, with highlighting to indicate the provisions challenged here by Plaintiff).

thoroughly analyzed the provisions of AB 587 identical to the provisions being challenged here, explaining that AB 587 compelled noncommercial speech by forcing “every covered social media company to reveal its policy opinion about contentious issues, such as what constitutes hate speech or misinformation and whether to moderate such expression.” *Id.* at 899; *id.* at 901 (“[W]hile a social media platform’s existing TOS and content moderation policies may be commercial speech, its opinions about and reasons for those policies are different in character and kind.”).

AB 587, although framed as “*only* a transparency measure,” compelled platforms to disclose “constitutionally protected speech that the State could not otherwise” compel “without satisfying strict scrutiny.” *Id.* at 902. It forced platforms to provide “insight” on whether, for instance, they consider “(1) a post citing rhetoric from on-campus protests to constitute hate speech; (2) reports about a seized laptop to constitute foreign political interference; or (3) posts about election fraud to constitute misinformation[.]” *Id.* AB 587’s “Content Category Report provisions” thus compelled noncommercial speech and triggered “strict scrutiny,” under which they did “not survive.” *Id.* at 898.

In *Volokh*, the Second Circuit fully endorsed the Ninth Circuit’s holding that AB 587’s Content Category Report Provisions violate the First Amendment. *Volokh* forecloses Defendant’s arguments and requires that the Motion be denied.

There, the Second Circuit held that whether New York’s Hateful Conduct Law, N.Y. Gen. Bus. Law §394-ccc (requiring social media networks to have a policy readily available that explains how the platform responds to and addresses reports of “hateful conduct”), violates the First Amendment, depends on “what the law actually requires” as a matter of statutory interpretation. *Volokh*, 148 F.4th at 89. The Court offered two possible interpretations: one where disclosure of policies that do not “specifically address hateful conduct” could satisfy the law and

another where compliance requires disclosure of policies specifically addressing hateful conduct. *See id.* at 90–94. The Court certified questions to the New York Court of Appeals to determine which interpretation is correct. *See id.* at 100–01.

The Court held that, under the first interpretation, the law would be subject to and withstand scrutiny under *Zauderer*. *Id.* at 90–91. And it held that, under the second interpretation, the law “*would be analogous to*” AB 587 (the “*California law considered by the Ninth Circuit in X Corp.*”), “*Zauderer would not apply,*” and “the statute would not satisfy the intermediate or strict scrutiny that would follow.” *Id.* at 93.³ And that makes perfect sense. If the Hateful Conduct Law required disclosures about a specific category of speech—there, “hateful conduct”—it would, like the Content Category Report Provisions of either AB 587 or S895B, be content-based and trigger heightened scrutiny, which it could not survive.

Because the Second Circuit held in *Volokh* that laws that are analogous to AB 587’s Content Category Report Provisions violate the First Amendment under either strict or intermediate scrutiny, and because the regulation of speech challenged in this case is *identical* to that law, Defendant’s Motion must be denied.

Moreover, as explained herein, the Content Category Report Provisions (i) compel noncommercial speech, which precludes the application of both intermediate (*Central Hudson*) scrutiny and *Zauderer*; (ii) compel information that is neither “purely factual” nor “uncontroversial,” which separately precludes *Zauderer*; (iii) are a content- and viewpoint-based regulation of speech, which triggers, at least, strict scrutiny; (iv) fail under any level of scrutiny; and (v) are properly subject to a facial challenge under the First Amendment.

³ Unless otherwise indicated, emphases in quotes are added and internal citations and quotations are omitted.

Finally, the Content Category Report Provisions also violate and are preempted by 47 U.S.C. §230(c)(2)(A) (“Section 230”), because they penalize interactive computer service (“ICS”) providers, such as Plaintiff, if they do not moderate content in the manner dictated by the State as part of a strategy to pressure the ICS providers to change their content-moderation practices. Section 230 imbues ICS providers with the autonomy to self-regulate content on their platforms, and the Content Category Report Provisions strike at the core of that autonomy.

FACTUAL BACKGROUND

I. S895B’s Statutory Scheme and its Application to Plaintiff

S895B requires “social media companies” to publish certain information about each of the “social media platforms” that the company “owns or operates.” *See* §§1101, 1102. X Corp. is a “social media company” and X is a “social media platform” under the statute. Compl. (Dkt. 1) ¶¶15–16; §§1100(4)–(5), 1104. X is also an “interactive computer service” under 47 U.S.C. §230(f)(2). Compl. ¶60.

S895B has three main components:

First, a “**Terms of Service Requirement**,” whereby social media companies must publicly post, for each of their social media platforms, “terms of service,” including processes for flagging content and potential actions that may be taken regarding flagged content. §1101.

Second, a “**Terms of Service Report**” (a/k/a “**Content Category Report**”), whereby social media companies must, for each of their platforms, submit a report to the New York Attorney General, who will disseminate the report to the public. §1102. The report must include, among other things, “statement[s]” and “detailed description[s]” about whether and if so, how, the company “defines” and “moderate[es]” “hate speech,” “racism,” “extremism,” “radicalization,” “disinformation,” “misinformation,” “harassment,” and “foreign political interference.”

§§1102(1)(c)–(d). The company must also provide statistics showing how often it censors content “belonging to” any of the eight content categories. §1102(1)(e).

Third, a “[v]iolations and remedies” (or “penalty”) provision, whereby social media companies may be liable to pay \$15,000 per violation per day and may be sued in court by the New York Attorney General for (a) violating the Terms of Service Requirement, (b) failing to timely submit a Terms of Service Report, or (c) “materially omit[ting] or misrepresent[ing] required information” in a Terms of Service Report. §1103.

Plaintiff’s lawsuit does not challenge the Terms of Service Requirement or even the Content Category Report in its entirety—only (i) the provisions of the Content Category Report that compel disclosures about hate speech, racism, extremism, radicalization, disinformation, misinformation, harassment, and foreign political interference⁴ and (ii) the violations and remedies provision as applied to those disclosures. *See* Compl. ¶¶24, 27.

A. Required Disclosure of Terms of Service (Terms of Service Requirement)

S895B’s Terms of Service Requirement mandates that social media companies publicly post, “in a manner reasonably designed to inform all users” of its “existence and contents,” the “terms of service for each social media platform” they own or operate. §1101(1). The terms of service must include (i) “contact information,” so users may “ask the social media company questions about” the terms, (ii) a “description of the process that users must follow to flag content, groups, or other users that they believe violate the terms of service,” as well as “the social media company’s commitments on response and resolution time,” and (iii) a “list of potential actions the social media company may take against an item of content or a user, including, but not limited to, removal, demonetization, deprioritization, or banning.” §1101(2).

⁴ Sections 1102(1)(c), 1102(1)(d)(i), and 1102(1)(e).

B. Terms of Service or Content Category Report

S895B’s Content Category Report Provisions require social media companies to submit semiannual Terms of Service Reports, on April 1 and October 1 of each year, to the New York Attorney General, who will then make them publicly available on the Attorney General’s website. *See* §1102(2)(a). The Terms of Service Report must include:

- A statement of whether, and if so how, the platform’s terms of use define hate speech, racism, extremism, radicalization, disinformation, misinformation, harassment, and foreign political interference, **§1102(1)(c)**⁵;
- The current version of the terms of service and a “complete and detailed description of any changes” to the terms since any previous report, **§1102(1)(a)–(b)**;
- A “*detailed description*” of the platform’s “content moderation practices,” including, but not limited to:
 - Any “existing policies *intended to address* the categories of content described in [§1102(1)(c)],” **§1102(1)(d)(i)**;
 - How “automated content moderation systems enforce” the terms of service and “when these systems involve human review,” **§1102(1)(d)(ii)**;
 - How the “company responds to user reports of violations of the terms of service,” **§1102(1)(d)(iii)**; and
 - How the “company would remove individual pieces of content, users, or groups that violate the terms of service, or take broader action against” individuals or groups of users “that violate the terms of service,” **§1102(1)(d)(iv)**;

⁵ For the Court’s convenience, the provisions challenged by Plaintiff are in red type.

- Information about “content that was flagged by the social media company as *content belonging to* any of the [eight content] categories,” §1102(1)(e)(i), including:
 - The “total number of flagged items of content,” §1102(1)(e)(i)(A);
 - The “total number of actioned items of content,” §1102(1)(e)(i)(B);
 - The “total number of actioned items of content that resulted in action taken by the social media company against the user or group of users responsible for the content,” §1102(1)(e)(i)(C);
 - The “total number of actioned items of content that were removed, demonetized, or deprioritized” by the company, §1102(1)(e)(i)(D);
 - The “number of times actioned items of content were viewed by users,” §1102(1)(e)(i)(E);
 - The “number of times actioned items of content were shared, and the number of users that viewed or heard the content before it was actioned,” §1102(1)(e)(i)(F); and
 - The “number of times users appealed social media company actions taken on” the platform and the “number of reversals of social media company actions on appeal disaggregated by each type of action,” §1102(1)(e)(i)(G);
- The information required by §1102(1)(e)(i) “*disaggregated*” by:
 - The “*category of content*, including any relevant categories described in [§1102(1)(c)],” §1102(1)(e)(ii)(A);
 - The “type of content” (e.g., “posts” or “comments”), §1102(1)(e)(ii)(B);
 - The “type of media of the content” (e.g., “text” or “videos”), §1102(1)(e)(ii)(C);

- How “the content was flagged” (e.g., “flagged by company employees,” “contractors,” “artificial intelligence software,” “community moderators,” “civil society partners,” or “users”), §1102(1)(e)(ii)(D); and
- How “the content was actioned” (e.g., “actioned by company employees,” “contractors,” “artificial intelligence software,” “community moderators,” “civil society partners,” or “users”), §1102(1)(e)(ii)(E).

C. Penalty Provision

Finally, S895B authorizes penalties of \$15,000 per violation per day and lawsuits for injunctive relief in any court of competent jurisdiction by Defendant if, according to Defendant, the company (i) “fails to post terms of service in accordance with” Section 1102, (ii) “fails to timely submit” a Terms of Service Report, or (iii) “materially omits or misrepresents required information in” a Terms of Service Report. §1103(1)(a)–(b).

II. Section 1102(1)(c)’s Content Categories

Hate speech, racism, extremism, radicalization, disinformation, misinformation, harassment, and foreign political interference comprise some of the most controversial and politically charged topics of the day. Compl. ¶28. They are taken word-for-word from California’s AB 587. *Id.*; see Dkt. 21-1 (S895B/AB 587 redline).

As AB 587’s legislative history makes clear, those content categories were selected largely *because* they are the most controversial and are “difficult to reliably define.” See Compl. Ex. 7 at 5 of 11. Their hard-to-define nature creates a “complex dilemma” for platforms attempting to moderate such content because—unlike “child pornography,” “depictions of graphic violence,” or other similar clear-cut topics—“assignment of their boundaries is often *fraught with political bias*,” such that “*both action and inaction*” as to moderation of those content categories is “*equally*

maligned: too much moderation and accusations of censorship and suppressed speech arise; too little, and the platform risks fostering a toxic, sometimes dangerous community.” *Id.*

Put differently, “[n]o matter how a social media company chooses to moderate such content, the company will face backlash from its users and the public.” *X Corp.*, 116 F.4th at 899 n.8. And that “is true even if the company decides not to define the enumerated categories, because they will draw criticism for under-moderating their community.” *Id.*

So, as the California legislature made clear, the thinking was, “if social media companies are forced to disclose” information about how they moderate these content categories, it may “*pressure them*” to “*eliminate hate speech and disinformation*” from their platforms. Compl. Ex. 8 at 5 of 8.

III. Plaintiff’s Current Content-Moderation Efforts

Plaintiff already undertakes substantial efforts to strike the right balance of content moderation on its platform. Compl. ¶¶32–33. But Plaintiff does so pursuant to its own protected content-moderation principles, not those of the State of New York. *Id.*

For instance, while Plaintiff does not currently regulate “hate speech,” “racism,” or “extremism,” §1102(1)(c), Plaintiff does regulate “hateful conduct”—Compl. Ex. 9 at 3, 7 of 10—a category of content that some could argue includes some speech that might qualify as “hate speech,” “racism,” or “extremism.” Compl. ¶32. Similarly, while Plaintiff does not currently regulate “misinformation” or “disinformation,” §1102(1)(c), Plaintiff does regulate “manipulated, or out-of-context media that may result in widespread confusion on public issues, impact public safety, or cause serious harm.” Compl. Ex. 10 at 10 of 15; Compl. ¶33.

The X Global Transparency Report and Meta Community Standards Enforcement Report cited by Defendant (*see* Mot. at 7) further demonstrate this point, because they show that neither

X’s nor Meta’s content-moderation policies utilize the same content terminology as S895B. *See* X Transparency, *Global Transparency Report* (H2 2024), available at <https://tinyurl.com/5cfvw7wu> (last visited Dec. 18, 2025) (showing that X moderates, e.g., “Hateful Conduct,” “Violent & Hateful Entities,” “Violent Content,” and “Misleading & Deceptive Identities”); Meta, *Community Standards Enforcement Report* (Q2 2025), available at <https://tinyurl.com/2vxuhvv4> (last visited Dec. 18, 2025) (showing that Meta moderates, e.g., “Violence and Incitement,” “Violent and Graphic Content,” “Hateful Conduct,” and “Fake Accounts”).

IV. S895B, AB 587, and the New York Legislature’s Refusal to Reconsider S895B Following the Ninth Circuit’s Decision in *X Corp.*

S895B is, again, for all intents and purposes, an exact replica of California’s AB 587. *See* Dkt. 21-1. When S895B was discussed on the floor of the New York Assembly on June 6, 2024, bill sponsor Assembly Member Grace Lee touted their “identical” nature and cited, as support for the passage of S895B, that X Corp.’s motion for a preliminary injunction of AB 587 had been denied by the district court. Compl. Ex. 3 at 221 of 272. On September 4, 2024, however, the Ninth Circuit issued its decision in *X Corp.*, reversing the district court. *See* 116 F.4th at 902 (noting that the “district court performed, essentially, no analysis” concerning whether the Content Category Report Provisions compelled commercial speech).

Not only did the New York legislature fail to revisit S895B after the Ninth Circuit’s decision, it “refuse[d]” to even have a “discussion” with X Corp. about “potential changes” to S895B after the decision. *See* Compl. Ex. 5 at 2 of 4. In a public September 9, 2024 letter to X Corp., S895B’s bill sponsors rejected an overture from the company to discuss changes to the law. The bill sponsors stated that they “decline the request on behalf of your client, X, to discuss amendments to [S895B]” because, in the view of the sponsors, X’s “owner, Elon Musk” used the

X platform to promote content that, in the view of the bill sponsors, “threatens the foundations of our democracy.” *Id.* at 2, 4 of 4.

ARGUMENT

I. The Content Category Report Provisions Violate the First Amendment as Applied to Plaintiff and On Their Face

A. *Volokh* and *X Corp.* Require Denial of Defendant’s Motion

As previewed above, the Second Circuit’s decision in *Volokh*, which endorsed the Ninth Circuit’s analysis and holding that the provisions of AB 587 that are identical to those challenged here violate the First Amendment, forecloses Defendant’s arguments and necessitates denial of Defendant’s Motion.

Volokh concerns a New York statute (the “Hateful Conduct Law”) that requires social media networks to “have a clear and concise policy readily available and accessible on their website and application which includes how such social media network will respond and address the reports of incidents of hateful conduct on their platform.” 148 F.4th at 80 (quoting N.Y. Gen. Bus. Law §394-ccc(3)). The statute defines “hateful conduct” as “the use of a social media network to vilify, humiliate, or incite violence against a group or a class of persons on the basis of race, color, religion, ethnicity, national origin, disability, sex, sexual orientation, gender identity or gender expression.” N.Y. Gen. Bus. Law §394-ccc(1)(a).

In *Volokh*, the Second Circuit held that whether the Hateful Conduct Law violates the First Amendment depends on “what the law actually *requires*.” 148 F.4th at 89 (emphasis in original). The *Volokh* Court offered two possible interpretations. If, on the one hand, the statute “mandates that social media networks disclose their content moderation policies—whatever they may be—without requiring those policies to specifically reference or otherwise encompass ‘hateful conduct,’” the law would be subject to and withstand review under *Zauderer*. *Id.* at 90. If,

however, the law requires disclosures that must “specifically address hateful conduct,” then “*Zauderer* would not apply, and the statute would not satisfy the intermediate or strict scrutiny that would follow.” *Id.* at 93.

The Second Circuit held that, under the latter interpretation, the statute would require a social media network to “disclose a content moderation policy that explains its procedures with respect to hateful conduct,” and thus “*would be analogous to*” AB 587, the “*California law considered by the Ninth Circuit in X Corp.*” *Id.* The *Volokh* Court unequivocally endorsed the Ninth Circuit’s decision and analysis in *X Corp.* and held that, if the Hateful Conduct Law were interpreted “analogously” to AB 587’s Content Category Report Provisions, it would trigger heightened scrutiny, not *Zauderer*, and would not survive. *See id.* at 93–94. The Court made clear that this was a holding, not dicta. *See id.* at 100 n.13 (rejecting the dissent’s argument that this “analysis is dicta”).

X Corp., again, concerned a California statute that is *identical* to S895B. Dkt. 21-1. There, the Ninth Circuit held that AB 587’s Content Category Report Provisions compelled *noncommercial* speech and thus triggered “strict scrutiny, under which they d[id] not survive,” rather than *Zauderer* or intermediate scrutiny. 116 F.4th at 898–903. The Ninth Circuit held that the law forced social media companies to “recast [their] content-moderation practices in language prescribed by the State, implicitly opining on whether and how certain controversial categories of content should be moderated.” *Id.* at 898, 901. In reaching this holding, the Ninth Circuit conducted a detailed commercial speech analysis, explaining that “a social media platform’s existing TOS and content moderation policies may be commercial speech,” but “opinions about and reasons for those policies,” like those compelled under the Content Category Report Provisions, “are different in character and kind,” because they “would require a social media

company to convey the company’s policy views on intensely debated and politically fraught topics,” and also “convey how the company has applied its policies.” *Id.* at 901–02.

Both the Second and Ninth Circuits squarely addressed—and rejected—the main argument in Defendant’s Motion—i.e., that *Zauderer* applies because the Content Category Report Provisions merely require social media companies to disclose their “terms of service,” which “constitutes commercial speech.” *Compare* Mot. at 9–10 (arguing that *Zauderer* applies to S895B) with *Volokh*, 148 F.4th at 93–94 (holding that *Zauderer* would not apply to the Hateful Conduct Law, and the law would fail heightened scrutiny if it were interpreted “analogously” to AB 587’s Content Category Report Provisions) and *X Corp.*, 116 F.4th at 898–903 (holding that *Zauderer* does not apply to AB 587 and that it fails to satisfy strict scrutiny).

Notwithstanding these clear holdings, Defendant argues that *Zauderer* should apply to S895B’s Content Category Report Provisions because “a social media company’s content moderation policies” are “commercial speech.” Mot. at 2. This argument is foreclosed by *X Corp.* and *Volokh*. As the Ninth Circuit correctly observed, forcing a company to disclose its “existing TOS and content moderation policies,” which “may be commercial speech,” is legally distinct and “different in character and kind” from forcing the company to express its “opinions about and reasons for those policies.” *X Corp.*, 116 F.4th at 901. The Second Circuit adopted this holding in *Volokh*. *See* 148 F.4th at 93 (holding that laws that are “analogous” to AB 587 are subject to heightened scrutiny, not *Zauderer*, and violate the First Amendment).

The Content Category Report Provisions trigger heightened scrutiny because they make companies disclose (i) “whether” they define the government-selected controversial content categories and, “if so,” the “definitions of those categories,” §1102(1)(c); (ii) a “detailed description” of how the company’s “content moderation practices” are “intended to address” those

eight categories, §1102(1)(d)(i); and (iii) statistics showing how often the company “flagged,” “actioned,” “removed,” “demonetized,” and “deprioritized” content “belonging to any of the [eight] categories,” §1102(1)(e)(i). The Content Category Report Provisions thus require social media companies to “specifically address,” “reference,” and “adopt” the *State’s* specific, controversial categories of content using the *State’s* terminology for such content. *See Volokh*, 148 F.4th at 93 (holding that such laws, which are “analogous” to AB 587, violate the First Amendment). And for these reasons, the Content Category Report Provisions are “different in character and kind” from neutral, content-agnostic disclosure laws—perhaps like S895B’s Terms of Service Requirement that Plaintiff does not challenge—that do not require companies to “specifically address” particular categories of content. *See id.* at 90 (holding that such laws do not violate the First Amendment).

Defendant’s reliance on *NetChoice, LLC v. Att’y Gen., Fla.* (“*NetChoice (Fla.)*”), 34 F.4th 1196 (11th Cir. 2022), and *NetChoice, LLC v. Paxton* (“*NetChoice (Tex.)*”), 49 F.4th 439 (5th Cir. 2022), is misplaced for this very reason. *See* Mot. at 9–10 (“A social media company’s terms of service, which encompasses rules governing the use of its platform, constitutes commercial speech.”). In reversing the district court on this exact point in *X Corp.*, the Ninth Circuit explained that neither *NetChoice* case “dealt with speech similar to the Content Category Reports” because, “[u]nlike” Texas HB 20 and Florida SB 7072, the “Content Category Report provisions compel social media companies to report whether and how they believe particular, controversial categories of content should be defined and regulated on their platforms.” 116 F.4th at 902. Specifically, neither law “require[d] a company to disclose the existence or substance of its policies addressing such categories,” and did not require “platforms to “mention [] controversial topics” or “particular content categories.” *Id.* at 902–903 (citing *NetChoice (Tex.)*, 49 F.4th at 446; *NetChoice (Fla.)*,

34 F.4th at 1206–07). The Ninth Circuit explained that while those *NetChoice* cases could be “relevant to an analysis of” AB 587’s Terms of Service Requirement—the analogous portion of S895B that Plaintiff does not challenge—they “are **unhelpful** on the issue of the Content Category Reports and **offer no compelling reason to apply *Zauderer***.” *Id.* at 903.

Finally, in the face of *Volokh*’s clear endorsement of *X Corp.*’s thorough analysis and holding regarding the very same law at issue in this case, Defendant resorts to attempting to distinguish *Volokh* on the basis that, while the Hateful Conduct Law defines “hateful conduct,” S895B does not provide a definition for any of the eight categories in Section 1102(1)(c). *See, e.g.,* Mot. at 9. But that argument ignores the fact that *Volokh* expressly held that a law “analogous” to AB 587 would be subject to heightened scrutiny, not *Zauderer*, and would violate the First Amendment. 148 F.4th. at 93 (citing *X Corp.*, 116 F.4th 888).

In any event, Defendant’s proposed distinction is unpersuasive because *Volokh* explained that the reason the Hateful Conduct Law would violate the First Amendment under an interpretation making it analogous to AB 587 is because a “disclosure requirement that incorporates the statute’s definition of hate speech would force social media networks to **develop and describe content moderation policies based on the specific categories identified by the State**.” *Id.* at 94. Even if S895B does not provide specific definitions for the eight content categories, those categories are framed under the State’s selected terminology. And for that reason, S895B requires the companies to “**recast [their] content-moderation practices in language prescribed by the State**, implicitly opining on whether and how certain controversial categories of content should be moderated.” *Id.* at 93 (quoting *X Corp.*, 116 F.4th at 901).

That is, the same impermissible end is achieved even though, in S895B, the State does not provide specific definitions for each content category. By forcing social media platforms to

explain whether they define and moderate particular controversial content *using the State’s preferred terminology*, S895B “force[s] the platforms to “to build and describe [their] framework for moderating content” in a way that “specifically address[es],” “reference[s]” and “adopt[s]” Section 1102(1)(c)’s eight controversial categories, “thereby burdening [their] expressive activity of curating content.” *Id.* at 93 (citing *Moody v. NetChoice, LLC*, 603 U.S. 707, 728 (2024)).

B. Neither Intermediate Scrutiny nor *Zauderer* Applies

Neither intermediate scrutiny nor *Zauderer* applies because S895B’s Content Category Report Provisions compel noncommercial speech. *Zauderer* also does not apply because the Content Category Report Provisions compel speech that is neither “purely factual” nor “uncontroversial.”

1. The Content Category Report Provisions Compel Noncommercial Speech, Precluding Intermediate Scrutiny and *Zauderer*

Because the Content Category Report Provisions compel noncommercial speech, neither intermediate scrutiny nor *Zauderer* may apply. *E.g., id.* at 85 (“Restrictions on *commercial* speech need only survive intermediate scrutiny[.]”); *id.* at 85–86 (“The *Zauderer* standard applies to regulations on *commercial* speech[.]”).

First, as discussed, even accepting Defendant’s repeated argument that social media terms of service and policies constitute commercial speech (*e.g.*, Mot. at 9–10), the disclosures compelled by the Content Category Report Provisions are “*opinions about and reasons for* those policies,” which “are different in character and kind” and are noncommercial speech. *Volokh*, 148 F.4th at 93 (emphasis in original) (quoting *X Corp.*, 116 F.4th at 901).

Second, the Content Category Report Provisions compel speech that falls outside of the “‘core notion’ of commercial speech”—i.e., “speech which does no more than propose a commercial transaction.” *Bad Frog Brewery, Inc. v. N.Y. State Liquor Auth.*, 134 F.3d 87, 97 (2d

Cir. 1998) (quoting *Bolger v. Youngs Drug Prods. Corp.*, 463 U.S. 60, 66 (1983)); *Volokh*, 148 F.4th at 85; *X Corp.*, 116 F.4th at 901. Specifically, the Content Category Report Provisions go beyond “communicat[ing] the terms of an actual or potential transaction” because “they express a view *about* those terms by conveying whether a company believes certain categories should be defined and proscribed.” *Id.* at 901 (emphasis in original).

Third, the Content Category Report Provisions compel disclosures that fail at least two of the three so-called “*Bolger* factors,” which apply in close cases where it is unclear whether or not speech is commercial. *See id.*; *Bad Frog*, 134 F.3d at 97 (explaining that “the presence of *all three* [*Bolger*] factors” supports a determination that the speech is commercial).

The compelled disclosures clearly “are not advertisements,” which Defendant does not even dispute, and the Content Category Report Provisions do not “merely disclose existing commercial speech, so a social media company has no economic motivation in their content.” *X Corp.*, 116 F.4th at 901. Defendant ignores this fact entirely in attempting to pigeonhole the disclosures into *Volokh*’s first interpretation of the Hateful Conduct Law, where it would *not* be “analogous” to AB 587 and would trigger *Zauderer*. *See* Mot. at 12, 17 (asserting that the challenged provisions merely require companies to disclose their terms of service and content-moderation policies as “they are”). But, again, compelling disclosure of a platform’s terms of service and policies *as they exist* (perhaps as required by S895B’s TOS Requirement) is “different in character and kind” from compelled disclosures, such as those required by the Content Category Report Provisions, that force platforms to (i) opine on “what constitutes hate speech or misinformation and whether to moderate such expression” and (ii) “convey [their] policy views” on “intensely debated and politically fraught topics” hand-selected by the State. *X Corp.*, 116 F.4th at 899, 901–02; *Volokh*, 148 F.4th at 94.

Finally, because the disclosures compelled by the Content Category Report provisions inherently “require companies to convey a policy view on ‘intensely debated and politically fraught topics, including hate speech [and] racism,’” *id.* (quoting *X Corp.*, 116 F.4th at 901–02), any commercial aspect of the compelled speech (and there is none) is, at the very least, “inextricably intertwined with otherwise fully protected speech,” such that it does not “retain[] its commercial character” and must receive full protection under the First Amendment. *Riley v. Nat’l Fed’n of the Blind of N.C., Inc.*, 487 U.S. 781, 796 (1988).

2. The Content Category Report Provisions Compel Disclosures that Are Neither “Purely Factual” Nor “Uncontroversial,” Precluding Application of *Zauderer*

Zauderer also cannot apply because the Content Category Report Provisions compel information that is not (i) “purely factual” or (ii) “uncontroversial.” *Zauderer*, 471 U.S. at 651.

Not Purely Factual. The Content Category Report Provisions compel speech that is not “purely factual” because they force platforms to express “opinions about” and “policy view[s]” on (i) what does or doesn’t constitute hate speech, racism, misinformation, disinformation, harassment, extremism, radicalism, and foreign political interference and (ii) whether, how, and to what extent such content should be moderated. *See* §1102; *Volokh*, 148 F.4th at 93–94; *X Corp.*, 116 F.4th at 901–02. These judgments are also bound up in a platform’s own protected editorial discretion and decision-making.

As just one example, Section 1102(1)(d) requires a “***detailed description*** of content-moderation practices used by” the platform, including its “existing policies ***intended to address*** the [eight] categories of content.” But whether a company has adopted a policy that is “intended to address” difficult-to-define categories of constitutionally protected speech requires the exercise of judgment and expression of opinions that could be the subject of disagreement among reasonable individuals. It is certainly not “purely factual” information.

The speech compelled by the Content Category Report Provisions is thus materially distinct from that compelled by the law in *Nat'l Retail Fed'n v. James* (“*NRF*”), 2025 WL 2848212 (S.D.N.Y. Oct. 8, 2025). *See* Mot. at 10, 13–15, 21. The law in *NRF* required merchants to make the following disclosure: “this price was set by an algorithm using your personal data.” *NRF*, 2025 WL 2848212, at *5. Judge Rakoff applied *Zauderer*, explaining that because the plaintiff “concede[d]” that “its members use algorithmic pricing to set prices and offer promotions,” the compelled disclosure was “plainly factual.” *Id.* The disclosures compelled under the Content Category Report Provisions are inherently distinct “opinions” and “policy views” that are not factual in the same way at all. *See X Corp.*, 116 F.4th at 901–02.

Not Uncontroversial. Nor are the disclosures compelled by the Content Category Report Provisions “uncontroversial,” as required to apply *Zauderer*. To be sure, Plaintiff’s argument is not that, because the eight content categories concern controversial topics, any compelled disclosure about those categories or topics is inherently controversial. *See* Mot. at 13. Nor is Plaintiff’s argument that the compelled information is controversial because Plaintiff would prefer not to communicate it. *See* Mot. at 10.

Rather, the disclosures compelled by the Content Category Report Provisions are controversial because they “require a social media company to convey the company’s policy views on intensely debated and politically fraught topics,” or force the company to publicly refuse to do so, which has the same effect. *X Corp.*, 116 F.4th at 899 n.8, 901–02. The controversial issue is the crux of the disclosure. Indeed, as AB 587’s legislative history makes clear, the statute is designed to generate public controversy about the actions of the social media companies by framing a divisive debate in such a way that will pressure the social media companies to change their content-moderation policies to align with those desired by the State. *See* pp.9–10 above.

That is why the statute focuses on what the legislative history concedes are the most controversial and politically charged categories of content that, no matter how they are moderated, will invariably leave some set of users dissatisfied. *Id.*

The Ninth Circuit explicitly acknowledged this structure in its decision: “No matter how a social media company chooses to moderate such content, the company will face backlash from its users and the public. That is true even if the company decides not to define the enumerated categories, because they will draw criticism for under-moderating their community.” *X Corp.*, 116 F.4th at 899 n.8 (“[W]e account for these effects in our analysis[.]”); *see also Nat’l Rifle Ass’n of Am. v. Vullo*, 602 U.S. 175, 190 (2024) (the State “cannot do indirectly what [it] is barred from doing directly”). Put another way, the very notion that the disclosures will result in “pressure” on the social media companies to “eliminate hate speech and disinformation” from their platforms (Compl. Ex. 8 at 5 of 8) presupposes that the disclosures are so controversial that they will result in the public demanding such changes.

Defendant’s citation to *CompassCare v. Hochul*, 125 F.4th 49 (2d Cir. 2025), actually illustrates the correctness of Plaintiff’s argument. *E.g.*, Mot. at 13. The law at issue in *CompassCare* required employers that issued employee handbooks to “include in the handbook notice of employee rights and remedies under” New York Labor Law §203-e (prohibiting discrimination based on reproductive health decision making). *CompassCare*, 125 F.4th at 54. In other words, the law required employers to include the “sort of disclosure requirement [that] is common and well-established in federal and state employment laws.” *Id.* at 67. And just because the topic of the state law was “controversial,” disclosing the “existence and contents” of Section 203-e was “not itself controversial.” *Id.* at 65.

Such a disclosure requirement is materially different from the Content Category Report Provisions. It would be as if the law at issue in *CompassCare* forced employers to inform their employees whether and how the employer defined “life-threatening pregnancies,” or whether and how the employer defined “pro-life” or “pro-choice,” and then to also provide a “detailed description” of what the employer was doing at the company “to address” those subjects, and statistics showing how the employer had addressed them in the past. Such disclosures would have been “controversial” and not subject to *Zauderer*. The Content Category Report Provisions cannot be subject to *Zauderer* for the same reasons.

C. Strict Scrutiny Applies Because the Content Category Report Provisions Are a Content- and Viewpoint-Based Regulation of Speech

The Content Category Report Provisions clearly target particular speech based on its content. They single out, for differential treatment, moderation of specific categories of speech and create additional burdens on social media companies moderating such content. *See also* Mot. at 14 (conceding that the Content Category Report Provisions are “content-based”). Under longstanding Supreme Court precedent, such content-based regulations of speech, those that “target speech based on ‘communicative content’ or ‘because of the topic discussed or the idea or message expressed’” are presumptively unconstitutional and may stand only if they survive “‘the most exacting scrutiny’—often called ‘strict scrutiny.’” *Volokh*, 148 F.4th at 84.

Not only do the Content Category Report Provisions target the particular categories of content set forth in Section 1102(1)(c), but they also discriminate against particular viewpoints within those categories, which makes them *per se* invalid. *See Pleasant Grove City, Utah v. Summum*, 555 U.S. 460, 469 (2009) (finding that content-based restrictions “must satisfy strict scrutiny,” but “restrictions based on viewpoint are prohibited”). The Content Category Report Provisions discriminate based on viewpoint because they do not compel disclosures regarding

content-moderation practices about categories that are “positive,” but do compel such speech for categories perceived by the State to be negative—i.e., about “hate speech,” but not “kind speech.” *See Iancu v. Brunetti*, 588 U.S. 388, 393 (2019) (explaining that such differential treatment “reflects the Government’s disapproval of a subset of messages it finds offensive” and is “the essence of viewpoint discrimination”). Such “ideologically driven attempts to suppress a particular point of view are [at least] presumptively unconstitutional[.]” *Rosenberger v. Rector & Visitors of Univ. of Va.*, 515 U.S. 819, 830 (1995).

The very selection of these categories strongly suggests that the State’s view is that they are types of speech that should be subject to content-moderation decisions. Indeed, the whole idea behind the statute is that, “if social media companies are forced to disclose” information about how they moderate these categories, it may “pressure them” to “eliminate hate speech and disinformation” from their platforms. Compl. Ex. 8 at 5 of 8. This design—which is evident whether or not the Court considers the legislative history (*see* Mot. at 16)—renders the Content Category Report Provisions viewpoint-based.

It also demonstrates that the law interferes with the protected “expressive activity” of social media companies, which they engage in by deciding whether and how to “publish others’ speech.” *Volokh*, 148 F.4th at 84 (citing *Moody*, 603 U.S. at 728). Accordingly, the Content Category Report Provisions are also a “content-based regulation subject to strict scrutiny” because they “dictate[] how to curate or compile others’ speech.” *Id.* In *Smith v. California*, 361 U.S. 147 (1959), the Supreme Court struck down an ordinance on First Amendment grounds that imposed liability on booksellers, because the law would “tend to restrict the public’s access to [content] which the State could not constitutionally suppress directly” and “self-censorship, compelled by the State, would be a censorship affecting the whole public, hardly less virulent for being privately

administered.” *Id.* at 154. The same is true as to laws, like S895B, that interfere with the speech rights of social media platforms. And, that the statute achieves this goal through public pressure rather than direct governmental pressure is of no moment, because the State “cannot do indirectly what [it] is barred from doing directly.” *Vullo*, 602 U.S. at 190.

D. The Content Category Report Provisions Fail Any Level of Scrutiny

1. Strict and Intermediate (*Central Hudson*) Scrutiny

The Content Category Report Provisions fail both strict scrutiny and intermediate scrutiny under *Central Hudson*. Defendant has not met its burden under either standard. The Content Category Report Provisions fail strict scrutiny because they are not (i) “narrowly drawn to serve” a “compelling government interest,” *Volokh*, 148 F.4th at 84, or (ii) the “least restrictive means to achieve” the interest, *Evergreen Ass’n, Inc. v. City of N.Y.*, 740 F.3d 233, 246 (2d Cir. 2014).⁶

The Content Category Report Provisions similarly fail intermediate scrutiny under *Central Hudson* because they do not “directly and materially advance[]” a “substantial” government interest and are “more extensive than necessary to serve that interest.” *Safelite Grp., Inc. v. Jepsen*, 764 F.3d 258, 264 (2d Cir. 2014). Under *Central Hudson*, the State has the “burden to ‘demonstrate that the harms it recites are real and that its restriction will in fact alleviate them to a material degree.’” *Ibanez v. Fla. Dep’t of Bus. and Pro. Regul., Bd. of Acct.*, 512 U.S. 136, 146 (1994). Defendant has not met its burden under any standard.

First, the governmental interest purportedly served by the Content Category Report Provisions—“empowering consumers to make informed choices about their use of social media,” Mot. at 16—is neither compelling nor substantial. There is no evidence, for example, that

⁶ Defendant has not even argued that the Content Category Report Provisions satisfy strict scrutiny, which is fatal, as in *X Corp.* 116 F.4th at 903.

consumers lack or desire more information about how content is regulated on X or other social media platforms. Defendant’s argument to the contrary rests largely on an assertion, supported by a citation to X’s “Terms of Service,” “Help Center,” and “X Rules,” that “attempting to discern what types of content or conduct may violate X’s rules is a confusing endeavor.” Mot. at 12 n.7. Yet, the very “X Rules” proffered by Defendant actually demonstrate that X does provide detailed information to the public about what content is and isn’t permitted on the platform. As just one example, while users may share “graphic media” that is “properly labeled,” “not excessively gory,” and that does not “depict[] sexual violence,” users may not share content that “explicitly threaten[s], incit[es], glorif[ies], or express[es] desire for violence.” *The X Rules*, X, <https://help.x.com/en/rules-and-policies/x-rules> (last visited Dec. 18, 2025). Thus, Defendant has failed to even show that “the harms it recites are real,” let alone that the law will “in fact alleviate them to a material degree.” *Ibanez*, 512 U.S. at 146.

Moreover, mandatory disclosure is not a compelling governmental interest in the context of regulating entities—such as newspapers, bookstores, and social media platforms—that make decisions about whether and how to display the speech of others, precisely because such laws implicate First Amendment concerns that do not exist with mandatory transparency laws outside of the speech context. While transparency laws outside of the speech context—e.g., a law requiring restaurants to disclose their Health Department inspection grades—may accomplish their goal of having businesses take certain desired non-speech conduct (e.g., clean up their kitchens) without any adverse impact on First Amendment rights, the same cannot be said of the Content Category Report Provisions, which are designed to pressure social media companies to change their content-moderation policies, and compliance with which will, “[n]o matter” what, generate

“backlash” for a social media company from its “users and the public.” *X Corp.*, 116 F.4th at 899 n.8.

Second, even if the State’s purported interest were compelling or substantial (and it is not), the Content Category Report Provisions are not narrowly tailored to serve that interest and are more extensive than necessary to further it. As the Ninth Circuit held regarding the very same transparency and consumer empowerment interest asserted by the California Attorney General, the Content Category Report Provisions are “more extensive than necessary” because consumers would still be “meaningfully informed” if the law required a company to disclose only “whether it was moderating certain categories of speech without having to define those categories in a public report,” or to disclose a “sample of posts that have been removed without requiring the company to explain why or on what grounds.” *Id.* at 903; *accord Volokh*, 148 F.4th at 93–94 (holding that, if the Hateful Conduct Law were interpreted in a manner that made it analogous to AB 587, it would fail even intermediate scrutiny as “more extensive than is necessary” because the “State might reasonably promote the goal of ensuring that users receive accurate information about a social media network’s content moderation policies with a generic and neutral disclosure requirement”). In fact, S895B already contains a “neutral” disclosure requirement of the type contemplated by *Volokh*—Section 1101’s Terms of Service Requirement, which, again, Plaintiff is not challenging.

2. *Zauderer*

The Content Category Report Provisions also do not survive under *Zauderer* because Defendant cannot demonstrate that the harms it recites are real and not “purely hypothetical.” *Nat’l Inst. of Fam. & Life Advocs. v. Becerra*, 585 U.S. 755, 776 (2018); *see pp.24–25 above*. Defendant has failed to show that consumers are unable to make informed choices now about

social media use or that current information about content-moderation policies is lacking in any way.

Nor is the governmental interest cited—providing transparency to allow consumers to make informed choices about social media use—sufficient to satisfy *Zauderer* review. As the Second Circuit held in *Int’l Dairy Foods Ass’n v. Amestoy*, 92 F.3d 67 (2d Cir. 1996), in concluding that a Vermont law requiring the labeling of milk from cows given rBGH likely violated the First Amendment, “consumer curiosity alone is not a strong enough state interest to sustain the compulsion of even an accurate, factual statement.” *Id.* at 74. Thus, under *Amestoy*, “transparency” is not a substantial interest in itself sufficient to satisfy *Zauderer* review, unless it serves some other important goal. Defendant has failed to demonstrate that, here, the government’s purported interest in transparency does anything more than satisfy consumer curiosity. The Content Category Report Provisions thus fail *Zauderer*. *Id.*

The Content Category Report Provisions also fail *Zauderer* because they are so “unduly burdensome” and “[u]njustified” that they will “chill[] protected speech.” *Milavetz, Gallop & Milavetz, P.A. v. United States*, 559 U.S. 229, 250 (2010). Specifically, the Content Category Report Provisions—and the threat of enforcement, investigation, and corresponding monetary and injunctive penalty authorized by Section 1103—pressure Plaintiff and the other platforms to moderate particular content disapproved of by the State, thereby interfering with their protected editorial speech rights. *See* Compl. ¶¶25, 29–30, 34–35; *NetChoice (Fla.)*, 34 F.4th at 1230–31 (finding SB 7072’s “explanation” requirement “unconstitutional under *Zauderer*” because it was “‘unduly burdensome’ and would ‘chill protected speech’—platforms’ exercise of editorial judgment”). In *X Corp.*, the Ninth Circuit acknowledged that the disclosure requirement would have “indirect chilling effects” by “generating public controversy about the actions of social media

companies,” given that, “[n]o matter how a social media company chooses to moderate” the eight content categories, the “company will face backlash from its users and the public,” thereby “*pressuring [the companies] to change their content moderation policies.*” *X Corp.*, 116 F.4th at 899 n.8.

The disclosures compelled by the Content Category Report Provisions are also “unduly burdensome” because they chill the protected speech of social media platform *users*—both as creators and viewers of speech. *See Volokh v. James*, 656 F. Supp. 3d 431, 445 (S.D.N.Y. 2023) (“The law requires that social media networks develop policies and procedures with respect to hate speech . . . This could have a profound chilling effect on social media users and their protected freedom of expression.”); *Volokh*, 148 F.4th at 98–99 (declining to opine on this point).

E. Plaintiff Meets the Requirements for a Facial Challenge

The Content Category Report Provisions are facially invalid under the First Amendment because “a substantial number of [the law’s] applications are unconstitutional, judged in relation to the statute’s plainly legitimate sweep.” *Americans for Prosperity Found. v. Bonta*, 594 U.S. 595, 615 (2021); *see also Moody*, 603 U.S. at 723 (noting that, “[i]n First Amendment cases,” the Supreme Court has “lowered” the “bar” for facial challenges).

In *X Corp.*, the Ninth Circuit conducted a detailed analysis of this very question and concluded that a facial challenge was “permissible” because “all aspects of the Content Category Report, in every application to a covered social media company, raise the same First Amendment issues”—specifically, “every Content Category Report must detail the company’s policies and actions concerning certain state-specified categories of content,” even “if only to detail the company’s decision not to define the enumerated categories.” 116 F.4th at 899.

Put differently, because, “[i]n effect, the Content Category Report provisions compel every covered social media company to reveal its policy opinion about contentious issues, such as what constitutes hate speech or misinformation and whether to moderate such expression,” they “raise the same First Amendment issues for every covered social media company.” *Id.* And the Ninth Circuit explained that because this was “*true from the face of the law*,” it was unnecessary to “speculate about ‘hypothetical’ or ‘imaginary’ cases.” *Id.* The Court should adopt that reasoning here.

II. The Content Category Report Provisions Violate and Are Preempted by 47 U.S.C. §230(c)(2)(A)

The Content Category Report Provisions facilitate governmental and public pressure to conform the content-moderation decisions of social media platforms to the State’s liking, subject to threats of enforcement. Accordingly, they violate and are preempted by Section 230(c)(2)(A), which prohibits liability “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.” In addition, under Section 230(e)(3), “[n]o cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section.”

The Content Category Report Provisions contravene this immunity by imposing monetary fines and suits for injunctive relief on ICS providers that “restrict access” to content on their platform in a way that, according to Defendant in its unfettered discretion, “misrepresents” or “materially omits” information in a Content Category Report. *See* §1103(1)(b)(iii). Put differently, the Content Category Report Provisions impose liability on ICS providers that take

action to restrict access to content without making the required disclosures to the State's satisfaction.

This directly contravenes the immunities afforded by Section 230, which grants ICS providers such as Plaintiff the autonomy to self-regulate content on their platforms. The Content Category Report Provisions strike at the core of that autonomy by “generating public controversy” about how the platforms moderate content, by “pressuring [platforms] to change their content moderation policies,” and by pressuring platforms to change how they moderate content. *See X Corp.*, 116 F.4th at 899 n.8; *Vullo*, 602 U.S. at 190 (the State “cannot do indirectly what [it] is barred from doing directly”). Indirect efforts by the State to influence content-moderation decisions through public “pressure,” *see* Compl. Ex. 8 at 5 of 8, violate Section 230, just as they violate the First Amendment.

The phrase “any action” in Section 230(c)(2)(A) covers actions taken by ICS providers in good faith to restrict access to content without making the disclosures required by the Content Category Report Provisions as the State wants them to be made. The Content Category Report Provisions thus violate and are preempted by Section 230(c)(2)(A).

CONCLUSION

For the foregoing reasons and those set forth in Plaintiff's Complaint, the Court should deny Defendant's Motion.

Dated: December 18, 2025

By: /s/ Joel Kurtzberg
CAHILL GORDON & REINDEL LLP
Joel Kurtzberg (SBN NY 2835007)
Floyd Abrams (SBN NY 1758184)
Jason Rozbruch (SBN NY 5753637)
32 Old Slip
New York, NY 10005
Phone: 212-701-3120
Facsimile: 212-269-5420
jkurtzberg@cahill.com
fabrams@cahill.com
jrozbruch@cahill.com

Attorneys for Plaintiff X Corp.

CERTIFICATION OF COMPLIANCE

The total number of words contained herein, exclusive of the cover page, certificate of compliance, table of contents, and table of authorities, is 8,736 words.

Dated: December 18, 2025

By: /s/ Joel Kurtzberg
CAHILL GORDON & REINDEL LLP
Joel Kurtzberg (SBN NY 2835007)
Floyd Abrams (SBN NY 1758184)
Jason Rozbruch (SBN NY 5753637)
32 Old Slip
New York, NY 10005
Phone: 212-701-3120
Facsimile: 212-269-5420
jkurtzberg@cahill.com
fabrams@cahill.com
jrozbruch@cahill.com

Attorneys for Plaintiff X Corp.