

In the
United States Court of Appeals
For the Second Circuit

August Term, 2023

(Argued: February 16, 2024 Decided: August 1, 2025)

Docket No. 23-356

EUGENE VOLOKH, LOCALS TECHNOLOGY INC., RUMBLE CANADA INC.,

Plaintiffs-Appellees,

–v.–

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

Defendant-Appellant.

Before: JACOBS, ROBINSON, and NATHAN, *Circuit Judges.*

Defendant-Appellant Letitia James, in her official capacity as Attorney General of New York, appeals from an order of the United States District Court for the Southern District of New York (Carter, *J.*) granting a preliminary injunction in favor of Plaintiffs-Appellees Eugene Volokh, Locals Technology Inc., and Rumble Canada Inc. (together, “Plaintiffs”) and enjoining enforcement of New York General Business Law § 394-ccc (“Hateful Conduct Law”).

The Hateful Conduct Law, broadly speaking, requires social media networks to (1) provide a “clear and easily accessible mechanism for individual users to report incidents of hateful conduct,” *id.* § 394-ccc(2), and (2) 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,” *id.* § 394-ccc(3). The statute defines “[h]ateful 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.” *Id.* § 394-ccc(1)(a). Plaintiffs maintain—as the district court held—that these provisions compel social media networks to engage in speech in violation of their First Amendment rights and chill their users from engaging in protected speech.

The constitutionality of the Hateful Conduct Law depends on how the statute is interpreted. If either substantive provision of the statute requires Plaintiffs to adopt or incorporate the State’s definition of “hateful conduct,” which includes constitutionally protected speech, then we review the statute under (at least) intermediate scrutiny, and it fails. But if Plaintiffs can comply with the Hateful Conduct Law by disclosing a content moderation policy that does not incorporate or affirmatively encompass the statute’s definition of “hateful conduct” and by providing a general mechanism for reporting content-related complaints, then we review the statute under the more relaxed standard set forth in *Zauderer v. Office of Disciplinary Counsel of Supreme Court of Ohio*, 471 U.S. 626 (1985), and its progeny, and the statute survives constitutional scrutiny. Whether the statute can support the latter, constitutional interpretation is a question best left to the New York Court of Appeals. We thus defer decision and **CERTIFY** to the New York Court of Appeals questions regarding the requirements of the statute.

Judge Jacobs dissents in a separate opinion.

SARAH COCO, Assistant Solicitor General (Barbara D. Underwood, Solicitor General, Judith N. Vale, Deputy Solicitor General, *on the brief*), for

Defendant-Appellant Letitia James, Attorney General of the State of New York, New York, NY.

JAMES MICHAEL DIAZ, Foundation for Individual Rights and Expression, Philadelphia, PA (Daniel Ortner, Foundation for Individual Rights and Expression, Philadelphia, PA; Barry Nelson Covert, Lipsitz Green Scime Cambria LLP, Buffalo, NY, *on the brief*), *for Plaintiffs-Appellees*.

Corbin K. Barthold, Andy Jung, TechFreedom, Washington, D.C., *for Amici Curiae* Prof. Eric Goldman and TechFreedom, *in Support of Plaintiffs-Appellees*.

Nicole Saad Bembridge, Chistopher J. Marchese, Carl Szabo, Paul D. Taske, NetChoice, Washington, D.C.; Jess Miers, Chamber of Progress, McLean, VA, *for Amici Curiae* NetChoice and Chamber of Progress, *in Support of Plaintiffs-Appellees*.

Catherine W. Short, Sheila A. Green, Life Legal Defense Foundation, Ojai, CA, *for Amici Curiae* Life Legal Defense Foundation and Young Americans for Freedom, *in Support of Plaintiffs-Appellees*.

Anastasia P. Boden, Cato Institute, Washington, D.C.; Joshua Robert Zuckerman, Brian C. McCarty, Gibson, Dunn & Crutcher LLP, Washington, D.C. *for Amicus Curiae* Cato Institute, *in Support of Plaintiffs-Appellees*.

David Greene, Electronic Frontier Foundation, San Francisco, CA; Brian Hauss, American Civil Liberties Union Foundation, New York, NY, *for Amici Curiae*, Electronic Frontier Foundation and American Civil Liberties Union, *in Support of Plaintiffs-Appellees*.

Bruce D. Brown, Katie Townsend, Gabe Rottman, Grayson Clary, Emily Hockett, Reporters Committee for Freedom of the Press, Washington, D.C., *for Amicus Curiae* Reporters Committee for Freedom of the Press, *in Support of Plaintiffs-Appellees*.

Talmadge Butts, Foundation for Moral Law, Montgomery, AL, *for Amicus Curiae* Foundation for Moral Law, *in Support of Plaintiffs-Appellees*.

Kaitlin D. Schiraldi, Richard Samp, New Civil Liberties Alliance, Washington, D.C., *for Amicus Curiae* New Civil Liberties Alliance, *in Support of Plaintiffs-Appellees*.

B. Tyler Brooks, Thomas More Society, Chicago, IL, *for Amicus Curiae* Thomas More Society, *in Support of Plaintiffs-Appellees*.

John J. Bursch, Alliance Defending Freedom, Washington, D.C.; James A. Campbell, Jonathan A. Scruggs, Bryan D. Neihart, Alliance Defending Freedom, Scottsdale, AZ; Jeremy Tedesco, Lansdowne, VA, *for Amicus Curiae* The Babylon Bee, *in Support of Plaintiffs-Appellees*.

ROBINSON, *Circuit Judge*:

Defendant-Appellee Letitia James, Attorney General of the State of New York, appeals from an order of the United States District Court for the Southern District of New York (Carter, *J.*), preliminarily enjoining enforcement of New York General Business Law § 394-ccc (“Hateful Conduct Law”) because it likely violates the First and Fourteenth Amendments to the United States Constitution.

For the reasons set forth below, we **CERTIFY** three questions of state law to the New York Court of Appeals.

BACKGROUND¹

I. 2022 Buffalo Shooting

On May 14, 2022, an 18-year-old white man drove to a supermarket in Buffalo with the stated intention of “killing as many blacks as possible.” J. App’x at 80.² With tragic success, the Buffalo shooter entered the supermarket, shot and killed ten innocent Black people, and injured several others.

The investigation that followed the Buffalo shooting revealed that social media played a key role in the shooter’s attack. First, investigators explained that “internet memes have been an effective means to mainstream white supremacist extremism and introduce it to new audiences” like the shooter because “memes can soften extremist ideas and make them more palatable to outsiders, while simultaneously creating an in-group—a community that understands the sometimes deeply encoded in-jokes.” *Id.* at 94. Because the Buffalo shooter wrote

¹ We draw this factual background from Plaintiffs’ Complaint and the parties’ submissions to the district court concerning Plaintiffs’ motion for preliminary injunction—including, in particular, the New York Attorney General’s “Investigative Report on the role of online platforms in the tragic mass shooting in Buffalo on May 14, 2022” (October 18, 2022). J. App’x at 70–118.

² In quotations from caselaw and the parties’ briefing, this opinion omits all internal quotation marks, footnotes, and citations, and accepts all alterations, unless otherwise noted.

a manifesto “peppered . . . with memes, in-jokes, and slang common on extremist websites and message boards,” investigators concluded the shooter found inspiration from the internet for his attack. *Id.* Investigators also determined the shooter found advice on online platforms about how to carry out his plans, including on how to obtain the weaponry needed for his attack. *See id.*

Second, the shooter broadcast a live video stream of his attack on the online video platform Twitch, which he then publicized using Discord, a different online platform. The video livestream lasted twenty-four minutes in total, the first twenty-two of which consisted of the shooter driving to the supermarket where his crimes took place. The last two minutes of the livestream showed the beginning of the shooter’s attack.

Twitch shut down the shooter’s livestream two minutes after the violence began, after at least one user reported the shooting to Twitch. Between twenty and twenty-eight Twitch users viewed at least a portion of the livestream. However, graphic videos and images taken while the shooter’s livestream was still publicly available “proliferated” on a fringe social media site in the days following the shooting, even though Twitch had terminated the livestream within two minutes after the violence began. *Id.* at 74.

II. New York's Hateful Conduct Law

In response to the Buffalo shooting, the New York State legislature enacted the Hateful Conduct Law, effective December 3, 2022. N.Y. Gen. Bus. Law § 394-ccc. The Hateful Conduct Law contains two key substantive requirements: the “Policy Disclosure Requirement” and the “Report Mechanism Requirement.”

The Policy Disclosure Requirement states:

Each social media network shall 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.

Id. § 394-ccc(3).

The Report Mechanism Requirement states:

A social media network that conducts business in [New York] shall provide and maintain a clear and easily accessible mechanism for individual users to report incidents of hateful conduct. Such mechanism shall be clearly accessible to users of such network and easily accessed from both a social media networks' application and website, and shall allow the social media network to provide a direct response to any individual reporting hateful conduct informing them of how the matter is being handled.

Id. § 394-ccc(2).

The statute defines “[h]ateful 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.” *Id.* § 394-ccc(1)(a). And it defines “[s]ocial media network[s]” as “service providers, which, for profit-making purposes, operate internet platforms that are designed to enable users to share any content with other users or to make such content available to the public.” *Id.* § 394-ccc(1)(b).³

The statute authorizes imposition of a civil penalty, up to \$1,000 per day, on any social media network that “knowingly fails to comply” with the Hateful Conduct Law, and it gives the New York Attorney General investigative, factfinding, and enforcement powers to carry out the statute’s mandate. *Id.* § 394-ccc(5).

Finally, the statute’s Savings Clause states that it shall not be construed “as an obligation imposed on a social media network that adversely affects the rights or freedoms of any persons,” including “the right of free speech pursuant to the first amendment to the United States Constitution.” *Id.* § 394-ccc(4)(a). The

³ The dissent substitutes the term “blogger” for the statutorily defined term, “social media network.” Dissent at 1. We doubt “the statute apparently applies to *any blog*,” *id.* (emphasis added), because some blogs aren’t fairly described as “service providers,” don’t have “profit-making purposes,” or don’t “enable users to share any content with other users or to make such content available to the public,” N.Y. Gen. Bus. Law § 394-ccc(1)(b). We accordingly stick to the defined term because “[t]he thing about a definition is that it is, well, definitional.” *Berman v. Neo@Ogilvy LLC*, 801 F.3d 145, 158 (2d Cir. 2015) (Jacobs, J., dissenting), *abrogated by Digital Realty Trust, Inc. v. Somers*, 583 U.S. 149 (2018); see also note 11, *below*.

Savings Clause similarly states that it shall not be construed “to add to or increase liability of a social media network for anything other than the failure to provide a mechanism for a user to report to the social media network any incidents of hateful conduct on their platform and to receive a response on such report.” *Id.* § 394-ccc(4)(b).

III. Plaintiffs

Plaintiffs are Eugene Volokh, Locals Technology Inc. (“Locals”), and Rumble Canada, Inc. (“Rumble”). Volokh is a law professor and the operator of *The Volokh Conspiracy*, a legal blog “dedicated to maintaining a free and open marketplace of ideas.” J. App’x at 9. Posts and open forums on *The Volokh Conspiracy* often receive dozens or sometimes hundreds of comments from readers. Commenters can respond to one another, and any member of the public can see *Volokh Conspiracy* comments.⁴ Volokh reports that some commenters have posted “controversial views on topics concerning one or more of the protected categories identified in” § 394-ccc(1)(a). *Id.* at 28. He also maintains full editorial control over *The Volokh Conspiracy* and contends that developing and maintaining the policies required by the New York Hateful Conduct Law would be time

⁴ Because of these interactive features, Volokh believes (and the State does not dispute) that *The Volokh Conspiracy* is a “social media network” as defined by the Hateful Conduct Law. See N.Y. Gen. Bus. Law § 394-ccc(1)(b).

consuming, burdensome, and “contrary to *The Volokh Conspiracy*’s ethos, purpose, and mission.” *Id.* at 27.

Plaintiff Rumble is an online service that allows users to upload and share video content (similar to YouTube, as Rumble says in the Complaint). *See id.* at 11. Users can comment on videos and reply to other comments, and all videos, comments, and replies are visible to the public. Rumble has “a mission to protect a free and open internet,” *id.* at 28–29, and considers its platform “immune to cancel culture,” *id.* at 28. Rumble maintains terms and conditions for its users; those terms prohibit content that is “grossly offensive to the online community, including but not limited to racism, anti-semitism and hatred.” *Id.* at 30. It likewise prohibits content that “supports or incites violence or unlawful acts” or that “supports groups that support or incite violence or unlawful acts.” *Id.* But the Complaint indicates that Rumble “does not prohibit content because it may ‘vilify’ or ‘humiliate’” others. *Id.* Rumble maintains an email address to receive complaints; though it “attempts to respond” to complaints, Rumble disclaims any responsibility to inform complainants about responsive actions it may have taken in response. *Id.* at 31 (alteration accepted).

Plaintiff Locals is a wholly owned subsidiary of Rumble and an online service that allows users to upload and share content with paid or unpaid

subscribers. Subscribers can comment on content and reply to other comments. Locals describes itself as “pro-free speech” and “committed to fostering a community that is safe, respectful, and dedicated to the free exchange of ideas.” *Id.* at 12. The site’s community guidelines prohibit “content that threatens violence against an individual or group of people” but otherwise allows its users to “independently police content.” *Id.* at 34. Locals maintains an email address to receive complaints but says “there is no requirement that complaints will receive a response from Locals.” *Id.*

IV. Procedural History

Plaintiffs sued Attorney General James on December 1, 2022, and moved for a preliminary injunction on December 6, 2022. The Complaint asserts five causes of action, all seeking to enjoin enforcement of the Hateful Conduct Law: (1) facial and as-applied First Amendment challenges to § 394-ccc as an unconstitutional content- and viewpoint-based regulation of speech, (2) facial and as-applied First Amendment challenges to § 394-ccc as unconstitutional compelled speech, (3) facial and as-applied First Amendment challenges to § 394-ccc as overbroad, (4) facial and as-applied Fourteenth Amendment challenges to § 394-ccc as void for vagueness, and (5) a challenge to § 394-ccc as preempted by the Communications Decency Act, 47 U.S.C. § 230.

In February 2023, following briefing and oral argument, the district court entered a preliminary injunction broadly prohibiting enforcement of § 394-ccc. *Volokh v. James*, 656 F. Supp. 3d 431, 447 (S.D.N.Y. 2023). The district court concluded that the Plaintiffs showed a substantial likelihood of success on their as-applied First Amendment challenges. *Id.* at 444. In particular, the district court reasoned that strict scrutiny applies because the Hateful Conduct Law compels the social media networks to engage in speech that is “inextricably intertwined” with protected speech that’s different in character from the “purely factual and uncontroversial information” subject to more relaxed review. *Id.* at 443. Because the regulation is not narrowly tailored to serve a compelling governmental interest, the district court concluded it likely does not survive strict scrutiny. *See id.* at 443–44.

In addition, the district court concluded that Plaintiffs showed a substantial likelihood of succeeding on their facial First Amendment vagueness and overbreadth challenges. *Id.* at 446. The court reasoned that even though the law does not require social media networks to remove content deemed to be “hateful conduct” from their websites and does not penalize social media *users* for purveying such content, it “could have a profound chilling effect” on social media users’ constitutionally protected expression, *id.* at 445, especially given the

ambiguity of the law's key terms such as "vilify" and "humiliate," *id.* at 446. Finally, the district court held that Plaintiffs' statutory preemption challenge under 47 U.S.C. § 230 was likely to fail. *See id.* at 446–47. Pending the State's interlocutory appeal of the preliminary injunction, the district court stayed proceedings below.

DISCUSSION

We have jurisdiction to hear this interlocutory appeal because we are reviewing a district court's grant of a preliminary injunction. *See* 28 U.S.C. § 1292(a)(1). "We review the grant of a preliminary injunction for abuse of discretion." *Evergreen Ass'n, Inc. v. City of New York*, 740 F.3d 233, 242 (2d Cir. 2014). "A district court abuses its discretion when (1) its decision rests on an error of law or a clearly erroneous factual finding, or (2) its decision—though not necessarily the product of a legal error or a clearly erroneous factual finding—cannot be located within the range of permissible decisions." *Id.*

"When a preliminary injunction will affect government action taken in the public interest pursuant to a statute or regulatory scheme, the moving party must demonstrate (1) irreparable harm absent injunctive relief, (2) a likelihood of success on the merits, and (3) public interest weighing in favor of granting the injunction." *Agudath Israel of America v. Cuomo*, 983 F.3d 620, 631 (2d Cir. 2020).

In a First Amendment challenge to a state or federal statute, “the likelihood of success on the merits is the dominant, if not the dispositive, factor.” *Id.* at 637. That’s because irreparable harm and the impact on the public interest track closely with the constitutionality of the challenged statute. If the statute is likely consistent with the Constitution, then there is no constitutional harm, and the public interest benefits from allowing the democratic process to function through the legislature’s voice; on the other hand, if the statute is likely unconstitutional, the public is irreparably harmed by the legislature’s infringement on First Amendment freedoms and any resulting chilling effect on expression. *See, e.g., New York Progress and Protection PAC v. Walsh*, 733 F.3d 483, 486 (2d Cir. 2013) (“The loss of First Amendment freedoms, even for minimal periods of time, unquestionably constitutes irreparable injury.”); *id.* at 488 (“[S]ecuring First Amendment rights is in the public interest.”). We therefore focus on the Plaintiffs’ likelihood of success.

We consider separately the district court’s ruling on Plaintiffs’ as-applied challenges to the Policy Disclosure Requirement and the Report Mechanism Requirement, and its judgment on Plaintiffs’ facial challenges. We conclude that the likelihood of success as to all of these challenges hinges on the proper interpretation of the New York statute—a matter of significant controversy.

Because this appeal presents questions best resolved by the New York Court of Appeals, and well-suited for certification, we certify the threshold, and likely dispositive questions to that court.

I. As-Applied Challenges

The district court analyzed Plaintiffs' contention that the Hateful Conduct Law was an unconstitutional content-based regulation of their speech as an as-applied challenge, and concluded that the law is subject to, and does not survive, strict scrutiny. In reviewing the district court's determination, we begin with general principles of First Amendment law that bear on our assessment, and then consider the Policy Disclosure Requirement and the Report Mechanism Requirement, respectively.⁵ We conclude that the constitutionality of each

⁵ "Under New York Law, a court should refrain from invalidating an entire statute when only portions of it are objectionable." *Evergreen*, 740 F.3d at 243. The touchstone of whether we can sever one provision of a statute from another is legislative intent: "[I]f partial invalidity had been foreseen," would the legislature "have wished the statute to be enforced with the invalid part excinded, or rejected altogether"? *Id.* The question is a functional one, and "separate commands, separately lettered" that "can operate independently" tend to be severable. *Centro de la Comunidad Hispana de Locust Valley v. Town of Oyster Bay*, 868 F.3d 104, 124 (2d Cir. 2017) (Jacobs, J., dissenting).

The Policy Disclosure Requirement can be severed from the Report Mechanism Requirement. Found in two different subsections, those requirements are "separate commands, separately lettered." *Id.* They also "operate independently." *Id.* The Report Mechanism Requirement could stand alone, requiring social media networks to provide a forum for complaints, even if we struck

provision turns on how one construes its requirements. Each provision is arguably subject to an interpretation that would render it constitutional, and one that would not. Whether the statutory text can support the constitutional interpretation is a question best answered by the New York Court of Appeals. Addressing the factors governing decisions to certify questions, we conclude that certification of three questions to the New York Court of Appeals is warranted.

A. Relevant First Amendment Framework

The First Amendment—which applies to New York by way of the Fourteenth Amendment Due Process Clause—prohibits states from abridging freedom of speech, expression, and the press. *See* U.S. Const. amend. 1; *Cantwell v. Connecticut*, 310 U.S. 296, 303 (1940); *Gitlow v. New York*, 268 U.S. 652, 666 (1925).

“As a general matter, government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.” *Brown v. Entertainment*

down the separate Policy Disclosure Requirement that might require them to explain what the social media networks do with those complaints. And making matters clearer, the legislators who shepherded the Hateful Conduct Law to passage stated the bill’s central purpose was to “require social media networks to provide and maintain mechanisms for reporting hate speech on their platform.” J. App’x at 268; *see also id.* at 174 (stating in the State Assembly, that “This is a bill that basically raises the standard for social media sites to have a[n] . . . easily accessible mechanism to report hateful material. That’s all the bill really does.” (emphasis omitted)). We take all these signs as a strong indication that the legislature prioritized the Report Mechanism Requirement and would want it to stand if a court were to strike down the Policy Disclosure Requirement. That doesn’t mean that we ignore the broader statutory scheme in assessing each challenged provision, but it does mean that we analyze each provision separately.

Merchants Ass’n, 564 U.S. 786, 790–91 (2011). Depending on both the type of expression and the nature of the government restriction, we employ different levels of judicial scrutiny to determine whether New York has overstepped constitutional lines. See, e.g., *Riley v. National Federation of the Blind of North Carolina, Inc.*, 487 U.S. 781, 796 (1988) (describing the nature of the speech taken as a whole and the effect of governmental regulation as “lodestars in deciding what level of scrutiny to apply” in a particular case).

i. Content-Based Regulations

Laws that regulate the content of a person’s speech—that is, laws that target speech based on “communicative content” or “because of the topic discussed or the idea or message expressed”—are the most constitutionally suspect. *Reed v. Town of Gilbert, Ariz.*, 576 U.S. 155, 163 (2015); see *Turner Broadcasting System, Inc. v. F.C.C.*, 512 U.S. 622, 642 (1994). This is because the central First Amendment freedom is the right of all people to decide for themselves which “ideas and beliefs [are] deserving of expression, consideration, and adherence,” and content-based restrictions on speech “raise the specter that the Government may effectively drive certain ideas or viewpoints from the marketplace.” *Turner*, 512 U.S. at 641. Accordingly, we apply “the most exacting scrutiny”—often called “strict scrutiny”—to content-based laws. *Id.* at 642, 653.

We treat a law compelling—rather than restricting—speech like any other content-based regulation because “[m]andating speech that a speaker would not otherwise make necessarily alters the content of the speech.” *Riley*, 487 U.S. at 795. That protection may apply as well to those who publish others’ speech, because for regulated entities that engage in their own expressive activity, “presenting a curated compilation of speech originally created by others” is likewise a form of expressive activity, and “ordering a party to provide a forum for someone else’s views” can thus implicate that party’s First Amendment rights. *Moody v. NetChoice, LLC*, 603 U.S. 707, 728 (2024) (considering constitutionality of laws regulating social media platforms’ content-moderation policies); *see also Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston*, 515 U.S. 557, 575 (1995) (observing that organizers of a St. Patrick’s Day parade could not be compelled to include a message of lesbian and gay pride). So, a law that dictates how to curate or compile others’ speech is generally a content-based regulation subject to strict scrutiny.

Finally, when a content-based law is subject to “strict scrutiny,” the state bears the burden of showing that the restriction on speech is “justified by a compelling government interest and is narrowly drawn to serve that interest.” *Brown*, 564 U.S. at 799. To satisfy this burden, “[t]he State must specifically identify

an actual problem in need of solving, and the curtailment of free speech must be actually necessary to the solution.” *Id.* Strict scrutiny is difficult to withstand, and the Supreme Court has observed that “[i]t is rare that a regulation restricting speech because of its content will ever be permissible.” *Id.*; see also *National Institute of Family and Life Advocates v. Becerra*, 585 U.S. 755, 766 (2018) (“NIFLA”) (describing strict scrutiny as a “stringent standard”).

ii. Regulation of Commercial Speech and Compelled Disclosures

Regulations involving some categories of speech are subject to a less-exacting standard of judicial scrutiny. “Commercial speech” — that is, speech that generally “does no more than propose a commercial transaction” — is one example. *Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, Inc.*, 425 U.S. 748, 762 (1976); see also *Central Hudson Gas & Elec. Corp. v. Public Service Commission of New York*, 447 U.S. 557, 562–63 (1980) (“The Constitution therefore accords a lesser protection to commercial speech than to other constitutionally guaranteed expression.”). Restrictions on commercial speech need only survive intermediate scrutiny to pass constitutional muster, meaning that the regulation must “directly advance[] a substantial governmental interest” and not be “overly restrictive.” *Safelite Group, Inc. v. Jepsen*, 764 F.3d 258, 261 (2d Cir. 2014); see also *Central Hudson Gas & Elec. Corp.*, 447 U.S. at 566. In other words, to regulate commercial speech,

the Constitution requires that there must be a “sufficient fit between the regulator’s ends and the means chosen to accomplish those ends.” *Vugo, Inc. v. City of New York*, 931 F.3d 42, 52 (2d Cir. 2019).

Under the umbrella of commercial speech, regulations requiring commercial *disclosure* of “purely factual and uncontroversial information about the terms under which . . . services will be available” may survive constitutional scrutiny if they are “reasonably related to the State’s interest in preventing deception of consumers,” and are not “unjustified or unduly burdensome.” *Zauderer v. Office of Disciplinary Counsel of Supreme Court of Ohio*, 471 U.S. 626, 651 (1985).⁶ The *Zauderer* standard applies to regulations on commercial speech that

⁶ This Court has often described scrutiny under *Zauderer* as “rational basis” review. See, e.g., *CompassCare v. Hochul*, 125 F.4th 49, 64–65 (2d Cir. 2025) (“[W]hile the [notice requirement] is a content-based regulation of speech, it is subject to the rational basis review standard articulated in *Zauderer*.”); *Safelite Group*, 764 F.3d at 259 (“We hold that the district court erred in applying rational basis review under *Zauderer*.”); *New York State Restaurant Ass’n v. New York City Bd. of Health*, 556 F.3d 114, 132 (2d Cir. 2009) (referencing “the rational basis test described in *Zauderer*”); *Connecticut Bar Ass’n v. United States*, 620 F.3d 81, 92 (2d Cir. 2010) (describing *Zauderer* as applying “rational basis review”). But some aspects of the *Zauderer* analysis are arguably more stringent than traditional rational basis review. For example, the universe of potentially cognizable state interests under the *Zauderer* framework has not been fully defined. See *Expressions Hair Design v. Schneiderman*, 877 F.3d 99, 103 (2d Cir. 2017) (“Despite the general rationale it offered in *Zauderer* for the lesser standard of review it articulated in that case, the Supreme Court has never clearly specified a governing framework that determines when *Zauderer*’s less-exacting standard should apply instead of *Central Hudson*’s intermediate scrutiny.”); *International Dairy Foods Ass’n v. Amestoy*, 92 F.3d 67, 74 (2d Cir. 1996) (“[C]onsumer

require a person to state “purely factual and uncontroversial information” about the goods or services the speaker may offer. *Id.*; see also *NIFLA*, 585 U.S. at 768. The Supreme Court’s rationale for applying a different framework to disclosure requirements is that “disclosure requirements trench much more narrowly on an advertiser’s interests than do flat prohibitions on speech.” *Zauderer*, 471 U.S. at 651. Commercial entities aren’t prevented from conveying the information they choose to convey; they are simply required “to provide somewhat more information than they might otherwise be inclined to present.” *Id.* at 650.

True, compelled speech can be just as violative of First Amendment freedoms as prohibitions on speech, *id.*, but requiring an advertiser to convey purely factual and uncontroversial information about the terms under which it will provide its services does not implicate the same weighty interests as forcing

curiosity alone is not a strong enough state interest to sustain the compulsion of even an accurate, factual statement in a commercial context.”); *American Meat Institute v. U.S. Dept. of Agriculture*, 760 F.3d 18, 20 (D.C. Cir. 2014) (concluding that permissible government ends under *Zauderer* extend sufficiently beyond preventing deception to encompass mandates requiring disclosure of country-of-origin information about meat products). Moreover, “rational basis” review does not typically require separate consideration of whether a regulation is “unjustified or unduly burdensome.” *Zauderer*, 471 U.S. at 651; see also *American Meat Institute*, 760 F.3d at 33–34 (“It is important to underscore that [the] *Zauderer* fit requirements are far more stringent than mere rational basis review.”) (Kavanaugh, J., concurring). We need not determine whether scrutiny under *Zauderer* is tantamount to traditional or perhaps more rigorous rational basis review, or whether it is better characterized as a special and more relaxed application of intermediate scrutiny; for purposes of our analysis, we refer to “*Zauderer*” scrutiny without assigning either label, other than to recognize that the scrutiny is more relaxed than ordinary intermediate or strict scrutiny.

individuals “to confess by word or act” their adherence to prescribed orthodoxy in “politics, nationalism, religion, or other matters of opinion,” *id.* at 651. “Because the extension of First Amendment protection to commercial speech is justified principally by the value to consumers of the information such speech provides,” an advertiser’s “constitutionally protected interests in *not* providing any particular factual information in [] advertising is minimal.” *Id.*

This Court echoed these sentiments in *National Elec. Mfrs. Ass’n v. Sorrell* (“NEMA”), explaining, “Commercial disclosure requirements are treated differently from restrictions on commercial speech because mandated disclosure of accurate, factual, commercial information does not offend the core First Amendment values of promoting efficient exchange of information or protecting individual liberty interests.” 272 F.3d 104, 113–14 (2d Cir. 2001). We thus concluded that “mandating that commercial actors disclose commercial information ordinarily does not offend the important utilitarian and individual liberty interests that lie at the heart of the First Amendment.” *Id.* at 114–15.

Applying this reasoning, we recently applied *Zauderer* in assessing a New York law requiring employers to include in their employee handbooks information about employees’ rights and remedies under a law prohibiting “discrimination based on an employee’s or a dependent’s reproductive health

decision making.” *CompassCare v. Hochul*, 125 F.4th 49, 53 (2d Cir. 2025). We reasoned that the notice requirement required disclosure only of “purely factual and uncontroversial information about the terms under which services will be available, specifically, the terms of employment under New York law.” *Id.* at 65. We recognized that aspects of the plaintiffs’ employee handbooks—such as the statement of the employer’s missions and values—were fully protected, but concluded that requiring inclusion of the required disclosures among other provisions involving the terms of employment, such as medical leave, payroll, information technology policies, and other legally mandated disclosures, did not compel speech that was “inextricably intertwined with otherwise fully protected speech.” *Id.* at 65 (citing *Riley*, 487 U.S. at 796).

Similarly, we subjected a New York City regulation requiring chain restaurants to post calorie information on menus to less exacting review in *New York State Restaurant Ass’n v. New York City Bd. of Health*, 556 F.3d 114 (2d Cir. 2009) (“NYSRA”). NYSRA argued that *Zauderer*’s relaxed scrutiny for required disclosures of purely “factual and uncontroversial information” didn’t apply to the calorie-disclosure requirement because, even though the regulation affected commercial speech and the required disclosures were factual, they were not “uncontroversial” because the restaurants did not want to communicate that

calorie amounts should be prioritized among other nutrients. *Id.* at 134. We disagreed: “[T]he First Amendment does not bar the City from compelling such under-inclusive factual disclosures, where . . . the City’s decision to focus its attention on calorie amounts is rational.” *Id.* at 134. In other words, the restaurants were still free to display other nutritional information, notwithstanding the regulation’s focus on calorie content.

We have likewise applied *Zauderer* scrutiny in upholding laws requiring disclosure that light bulbs contain mercury and should be disposed of as hazardous waste, *NEMA*, 272 F.3d at 107; requiring that debt relief agencies disclose specific information about the bankruptcy process to consumer debtors, *Connecticut Bar Ass’n v. United States*, 620 F.3d 81, 95–100 (2d Cir. 2010); and requiring as a matter of securities law that a disclaimer accompany hypothetical or simulated data, *Commodity Futures Trading Comm’n v. Vartuli*, 228 F.3d 94, 108 (2d Cir. 2000); *see also Expressions Hair Design v. Schneiderman*, 877 F.3d 99, 104 (2d Cir. 2017) (explaining, in certifying statutory interpretation question to New York Court of Appeals, that if the statute at issue required a merchant to disclose an item’s credit card price but did not otherwise bar the merchant from implementing a pricing scheme that differentiates between payments by cash and credit cards, the *Zauderer* framework might apply).

In sum, “at a minimum, *Zauderer* supplies the governing standard when evaluating the constitutionality of a law (1) designed to address misleading commercial speech (or, presumably, its equivalent, the non-disclosure of information material to the consumer), (2) which mandates only that the merchant make certain truthful statements, and (3) which does not prevent the merchant from conveying additional truthful information.” *Expressions Hair Design*, 877 F.3d at 104.

But there are limits to *Zauderer*’s applicability. In *Evergreen Ass’n, Inc. v. City of New York*, we considered the constitutionality of a New York City law relating to abortion and pregnancy services. 740 F.3d at 237–38. The law required pregnancy services centers to make certain disclosures at their entrances and waiting rooms, on ads, and during phone calls. The required disclosures included “whether or not they ‘provide or provide referrals for abortion,’ ‘emergency contraception,’ or ‘prenatal care.’” *Id.* at 238 (quoting New York City Administrative Code § 20-816(a)-(e)). We declined to apply *Zauderer* because the services disclosure mandate was not purely factual and uncontroversial; it “require[d] centers to mention controversial services that some pregnancy services centers, such as Plaintiffs in this case, oppose.” *Id.* at 245 n.6. We concluded that the provision did not survive either intermediate or strict scrutiny. *Id.* at 251; *see*

also *NIFLA*, 585 U.S. at 768–69 (California statute requiring licensed crisis pregnancy centers to disseminate notices about family planning and abortion services not subject to *Zauderer* review in part because disclosures related to the controversial topic of abortion).

iii. Unprotected Speech

Finally, some laws do not implicate the First Amendment at all and therefore receive no judicial scrutiny. For example, laws not fairly characterized as regulating expression but instead regulating conduct do not trigger First Amendment scrutiny. See *Rumsfeld v. Forum for Academic and Institutional Rights*, 547 U.S. 47, 65–70 (“*FAIR*”) (2006). In *FAIR*, the Supreme Court declined to apply any level of scrutiny to a law that required universities to provide military recruiters the same access to campuses that it provided to nonmilitary recruiters. See *id.* at 55. In doing so, the Court rejected the argument that the mere presence of military recruiters on campus compelled universities to express approval of the military by association; rather, that law regulated conduct only. *Id.* at 69–70.

We similarly do not scrutinize laws that regulate types of speech that have historically fallen outside First Amendment protection. “[T]he First Amendment has permitted restrictions upon the content of speech in a few limited areas,” including fraud, defamation, obscenity, incitement to violence, and fighting

words. *United States v. Stevens*, 559 U.S. 460, 468–69 (2010); *see also Brown*, 564 U.S. at 791 (collecting cases); *Virginia v. Black*, 538 U.S. 343, 359 (2003) (collecting cases). Those categories are “well-defined and narrowly limited classes of speech,” and “new categories of unprotected speech may not be added to the list by a legislature that concludes certain speech is too harmful to be tolerated.” *Brown*, 564 U.S. at 791. So we will uphold laws that regulate only those narrow categories of unprotected speech. *See, e.g., Brandenburg v. Ohio*, 395 U.S. 444, 449 (1969) (vacating conviction for advocating racial violence but noting that a state may prohibit “incitement to imminent lawless action”); *Beauharnais v. Illinois*, 343 U.S. 250, 261 (1952) (sustaining defamation law over First Amendment objection).

Critically important to this case, the Supreme Court has consistently held that expression motivated by bias, hatred, or bigotry falls within the First Amendment’s protection. *See Matal v. Tam*, 582 U.S. 218, 246 (2017); *R.A.V. v. City of St. Paul*, 505 U.S. 377, 391 (1992). In *Matal*, the Supreme Court invalidated a statute that prohibited registration of trademarks that “disparage or bring into contempt or disrepute any persons, living or dead.” 582 U.S. at 223. The challenge to the statute arose when the government denied trademark registration to a band that named itself “The Slants”—after a derogatory term for people of Asian descent—because the term was demeaning. *Id.* As the Court reasoned in striking

down that law, “Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express the thought that we hate.” *Id.* at 246.

*B. As-Applied Challenge to the Policy Disclosure Requirement*⁷

With these general principles in mind, we consider Plaintiffs’ as-applied challenge to the Policy Disclosure Requirement, which mandates that “[e]ach social media network shall have a clear and concise policy readily available and accessible on their website and application.” N.Y. Gen. Bus. Law § 394-ccc(3). That policy must “include[] how such social media network will respond and address the reports of incidents of hateful conduct on their platform.” *Id.*

Plaintiffs say that the Policy Disclosure Requirement compels Plaintiffs to speak about and take a position on speech the State characterizes as “hateful conduct.” Accordingly, the Hateful Conduct Law is subject to strict scrutiny,

⁷ The Plaintiffs’ Complaint raises four substantive theories of constitutional infirmity—that the Hateful Conduct Law (1) regulates content, (2) compels speech, (3) is overbroad, and (4) is vague. *See* J. App’x at 38–46. On each theory, the Complaint requests relief as applied to the Plaintiffs themselves and facially as to anyone anywhere. *See id.* The district court appears to have considered only as-applied relief for Plaintiffs’ content-regulation and compelled-speech theories, *see Volokh*, 656 F. Supp. 3d at 439–44, and it appears to have considered only facial relief for Plaintiffs’ overbreadth and vagueness theories, *see id.* at 444–46. We address only the theories and reasoning that the district court actually used to support its preliminary injunction. *See Sulzer Mixpac AG v. A&N Trading Co.*, 988 F.3d 174, 184 (2d Cir. 2021) (“We generally refrain from considering issues not decided by the district court.”).

which, Plaintiffs contend, it fails. Plaintiffs argue that the Policy Disclosure also fails intermediate scrutiny, and even scrutiny under the relaxed *Zauderer* standard.

The State, on the other hand, argues that the Policy Disclosure Requirement is a commercial disclosure requirement subject to review under the *Zauderer* framework. The State emphasizes that in its view, a network is not required “to adopt or even reference [the statute’s] definition of hateful conduct in its policy.” Appellant’s Br. at 44; *see also id.* at 45 n.5 (asserting that the statute “plainly does not require a network’s policy to reference the statute’s definition of hateful conduct”). And the State contends that the Policy Disclosure Requirement satisfies the *Zauderer* framework “because it is reasonably related to appropriate state interests and does not unduly burden [P]laintiffs’ speech.” *Id.* at 50. The State argues in the alternative that if *Zauderer* does not apply, the Policy Disclosure Requirement, as a regulation of commercial speech, is subject to intermediate scrutiny, which it survives. The State does not on appeal purport to advance a compelling state interest or the narrow tailoring required to overcome strict scrutiny, and it appears to have conceded at oral argument in the district court that it would have “a very difficult time” meeting the strict scrutiny standard. J. App’x at 314.

How we analyze the Policy Disclosure Requirement depends on what the law actually *requires*. If, as the State argues on appeal, it merely requires social media networks to publicly disclose their content moderation policies, whatever those policies may be, and contains no requirement that those policies actually specifically address “hateful conduct,” as defined by the statute, then the statute requires “purely factual and uncontroversial information about the terms under which [commercial] services will be available.” *Zauderer*, 471 U.S. at 651. Thus understood, the Policy Disclosure Requirement would pass constitutional muster under the *Zauderer* framework.

But if the law requires disclosures that reference “hateful conduct” or affirmatively encompass speech fitting within that definition, it would compel controversial speech and would not be eligible for relaxed *Zauderer* review. Whether evaluated under intermediate scrutiny or strict scrutiny, the Policy Disclosure Requirement would unconstitutionally burden the social media networks’ First Amendment rights.

The most natural reading is that the Policy Disclosure Requirement calls upon social media networks to develop and publish policies that specifically address hateful conduct, as that term is defined by statute. But the principle of constitutional avoidance weighs in favor of the narrower, constitutional

interpretation. Whether the text of the statute can support that construction is a close question best resolved by the New York Court of Appeals. We consider each of these points in turn.

i. A neutral disclosure requirement would survive *Zauderer* scrutiny

If the Policy Disclosure Requirement merely 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,” it would be more closely analogous to the calorie-content disclosure requirement in *NYSRA* or the mercury disclosure requirement in *NEMA* than the abortion-services disclosure requirement at issue in *NIFLA* and *Evergreen*. The content moderation policies require only that social media networks make “certain truthful statements” concerning “information material to the consumer.” *Expressions Hair Design*, 877 F.3d at 104. And because those statements bear on what social media users can expect social media companies to do after receiving certain complaints, they relate directly to “the terms under which . . . services will be available” from a social media network. *Zauderer*, 471 U.S. at 651. They are not, as the dissent claims, “different in kind” from the calorie content in food or the presence of mercury in light bulbs. Dissent at 12; *see NYSRA*, 556 F.3d at 134; *NEMA*, 272 F.3d at 114.

And though the policies themselves might be controversial, the fact that *they are what they are* is not. Cf. *CompassCare*, 125 F.4th at 65 (“[T]he policy judgment that motivated the [state law] may be ‘controversial’ in the same way that the policy judgments underlying Title VII, or minimum wage laws, are controversial. But the existence and contents of the [law]—and an employer’s obligation to comply with it—is not itself controversial.”). Moreover, if social media networks can comply with § 394-ccc(3) without specifically addressing “hateful conduct” and indicating whether it views such conduct as worthy of moderation, then the required disclosures are not particularly controversial. Accord *NetChoice, LLC v. Attorney General, Florida*, 34 F.4th 1196, 1227, 1230 (11th Cir. 2022) (Florida content moderation disclosure law was subject to *Zauderer* scrutiny because it required disclosure of “purely factual and uncontroversial information” about the terms of the commercial transaction between the platforms and “users”, that is, “consumers who engage in commercial transactions with platforms by providing them with a user and data for advertising in exchange for access to a forum”), *vacated on other grounds in Moody v. NetChoice, LLC*, 603 U.S. 707 (2024).

Importantly, the policies are chosen by social media networks themselves. They can adopt whatever policies they choose; they just have to disclose them. If we accept the Attorney General’s arguments, then even if a social media network

adopted *no formal policy at all* and applied its editorial discretion on a case-by-case basis, it could disclose that policy and satisfy its obligations under § 394-ccc. They may even say that they “generally avoid regulating or removing creator content or visitor comments from their sites” out of dedication to the marketplace of ideas and opposition to “cancel culture.” J. App’x at 9, 11.

A neutral requirement that social media networks disclose their content moderation policies would be easily distinguishable from requiring an anti-abortion crisis pregnancy center to mention abortion services to pregnant patients. *See Evergreen*, 740 F.3d at 245 n.6. For one, the disclosures here relate to the social media services Plaintiffs actually provide to consumers. *Cf. NIFLA*, 585 U.S. at 768–69 (noting that required disclosure about family planning services *available elsewhere*, including abortion, “in no way relates to the services that licensed [pregnancy crisis center] clinics provide”).

Moreover, if the Policy Disclosure Requirement is truly agnostic about the substance of the content moderation policy, social media networks are not required to adopt any particular policy or to address any particular category of speech—unless they want to. Social media networks like Plaintiffs may not *want* to discuss content moderation policies at all. But that wouldn’t remove this regulation from the *Zauderer* framework; after all, the chain restaurants in *NYSRA*

didn't want to talk about the calorie content of their meals. *See NYSRA*, 556 F.3d at 134 (rejecting argument that calorie disclosures were controversial because the restaurants did not want to communicate that calorie amounts should be prioritized among other nutrients); *Zauderer*, 471 U.S. at 650 (applying less stringent scrutiny to regulation requiring commercial entities "to provide somewhat more information than they might otherwise be inclined to present").

Under this view of the statute, the Policy Disclosure Requirement would be subject to *Zauderer*, and would pass constitutional muster if "reasonably related to the State's interest in preventing deception of consumers." *Zauderer*, 471 U.S. at 651. A neutral Policy Disclosure Requirement—compelling disclosure of content moderation policies without any requirements as to the scope or content of those policies—would fit this bill. In particular, it would ensure that users are fully informed about the terms of their engagement with a social media network, enabling them to make more informed choices about where they spend their screen time and how to interpret the content they find on a given social media network.

And Plaintiffs have not established a substantial likelihood that they would succeed in challenging the statute thus understood. *See Green Haven Prison Preparative Meeting of Religious Society of Friends v. New York State Department of*

Corrections and Community Supervision, 16 F.4th 67, 78 (2d Cir. 2021) (requiring party seeking injunction to bear burden of showing substantial likelihood of success on the merits).

The dissent’s contention—that *even if* the statute requires only a neutral disclosure that does not incorporate or affirmatively encompass the State’s definition of “hateful conduct,” it is still unconstitutional—rests on a misinterpretation of the statute and untested empirical assumptions. Dissent at 6.

As to the statute’s meaning: The premise underlying much of the dissent’s analysis on this point is that the statute subjects social media networks to enforcement actions and fines if the Attorney General concludes that they have failed to respond to specific user complaints in a way that is consistent with their published moderation policy. Dissent at 7–10. But the statute does no such thing. The statute specifies that none of its provisions shall be construed “to add to or increase liability of a social media network for anything other than the failure to provide a mechanism for a user to report to the social media network any incidents of hateful conduct on their platform and to receive a response on such a report.” N.Y. Gen. Bus. Law § 394-ccc(4)(b). In other words, the terms of the statute specifically refute the suggestion that the Attorney General can impose

consequences on a social media network for responding, or failing to respond, to a content-related complaint in any particular way.

Relying on a statement by counsel for the Attorney General during oral argument, the dissent suggests that the statute has far broader scope, subjecting a social media network to enforcement proceedings if it fails to comply with its own policies to the satisfaction of the Attorney General. *See* Dissent at 2, 7–10. But no comment of counsel can override the clear statutory language that belies the dissent’s description of the statute.

Moreover, even in the absence of the express statutory language limiting a social media network’s liability under the statute, the statute establishes civil liability only for “knowingly” violating the law’s requirements. *Id.* § 394-ccc(5). This state-of-mind requirement would preclude liability for a social media platform’s innocent or negligent deviation from its policies. At most, the statute would allow the Attorney General to impose liability for a social media network’s *knowing* propagation of *falsehoods* as to its moderation policy, rather than for well-intentioned mistakes flowing from a large platform’s insufficient management or an independent blogger’s inexperience. A social media site’s interest in withholding or misrepresenting accurate information is, to say the least, “minimal.” *Zauderer*, 471 U.S. at 651.

As to the dissent’s empirical assumption: the dissent argues that a neutral Policy Disclosure Requirement would impose disproportionate compliance costs on social media networks based on the content of discussions that tend to take place on different sites. *See* Dissent at 7–8 (citing *Simon & Schuster, Inc. v. Members of New York State Crime Victims Bd.*, 502 U.S. 105 (1991)). Maybe. It’s plausible that a network that discloses a policy to allow a more “free-wheeling discussion” of controversial topics may indeed generate more requests for moderation. *Id.* at 8. But it’s also plausible that disclosing moderation policies may instead allow consumers to choose social media networks that align with their individual preferences regarding content moderation. Those who prefer unmoderated discussions would be likelier to use social media networks like Locals and Rumble that “generally avoid regulating or removing creator content or visitor comments.” J. App’x at 9. Those who prefer an online discussion moderated more aggressively might simply use different social media networks. The result might be better-informed consumer choices and fewer complaints all around. *See Zauderer*, 471 U.S. at 651 (“[T]he extension of First Amendment protection to commercial speech is justified principally by the value to consumers of the information such speech provides.”). The dissent’s hypothesis is plausible—and so are many others. On this record, and at the preliminary injunction stage before

the law has even taken effect, we have no sound basis for embracing any particular supposition as to the empirical effect of a neutral disclosure requirement.

- ii. A requirement that the content moderation disclosure specifically address hateful conduct, as defined by the statute, would fail intermediate or strict scrutiny

On the other hand, if the Policy Disclosure Requirement requires social media networks to reference or adopt the State's definition of "hateful conduct" found in § 394-ccc(1)(a), then *Zauderer* would not apply, and the statute would not satisfy the intermediate or strict scrutiny that would follow.

As noted above, the statute's definition of "hateful conduct" encompasses a swath of protected speech by including speech to "vilify" or "humiliate . . . 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). Requiring a social media network to disclose a content moderation policy that explains its procedures with respect to hateful conduct, as defined by the statute, would force that network to build and describe its framework for moderating conduct around the *State's* definition, thereby burdening the networks' expressive activity of curating content. *See Moody*, 603 U.S. at 728 ("[E]xpressive activity includes presenting a curated compilation of speech originally created by others.").

Under this interpretation, the New York statute would be analogous to a California law considered by the Ninth Circuit in *X Corp. v. Bonta*, 116 F.4th 888 (9th Cir. 2024). There, plaintiffs sued to enjoin enforcement of a California statute that required social media companies to make public disclosures to the state attorney general concerning their content-moderation practices specifically with reference to certain types of speech, including “[h]ate speech or racism,” “[e]xtremism or radicalization,” “[d]isinformation or misinformation,” and “[h]arassment.” *Id.* at 896. Covered entities had to include in their public reports “whether the current version of the terms of service define[d]” those types of content, “[a] detailed description of content moderation practices” for those types of content, and metrics about how much of the amount of content falling into each category that users had flagged and that the social media company took action on. *Id.* at 896–97.

The Ninth Circuit concluded that *Zauderer* did not apply to the California law, and it subjected the law to strict scrutiny as a content-based regulation of non-commercial speech. The court explained that the statute “require[d] a company to recast its 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 901. It acknowledged that a social media platform’s

content moderation policies “may be commercial speech,” but said “its *opinions about and reasons for* those policies are different in character and kind.” *Id.* (emphases added). The disclosures required under the California law provided “[i]nsight into whether a social media company considers . . . a post citing rhetoric from on-campus protests to constitute hate speech,” or whether “posts about election fraud . . . constitute misinformation”—matters that constitute “sensitive, constitutionally protected speech that the State could not otherwise compel a social media company to disclose without satisfying strict scrutiny.” *Id.* at 902.

The New York Hateful Conduct Law imposes fewer disclosure requirements than the California law, but, if understood to require disclosure of content moderation policies that reference or encompass the State’s definition of hateful conduct, it would analogously require companies to convey a policy view on “intensely debated and politically fraught topics, including hate speech [and] racism.” *Id.* at 901–02. This is not the kind of “purely factual and uncontroversial information” subject to *Zauderer* review. *Zauderer*, 471 U.S. at 651. Instead, the law would essentially require social media networks to “provide a forum for someone else’s views”—in this case, the State’s. *Moody*, 603 U.S. at 728.

On this view of the statute, the Policy Disclosure Requirement would be subject to intermediate or strict constitutional scrutiny, and it cannot survive

either.⁸ Intermediate scrutiny would require the State to show that the Policy Disclosure Requirement “directly advances a substantial governmental interest and is not overly restrictive.” *Safelite Group*, 764 F.3d at 261. The State advances two interests to justify the Policy Disclosure Requirement: (1) providing social media users with accurate information about how reports of hateful conduct may be handled, and (2) facilitating reports of hateful conduct to prevent violence like the 2022 Buffalo shooting. *See* Appellants’ Br. at 56–57.

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. But 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. This requirement would be “more extensive than is necessary” to achieve the State’s goals. *Central Hudson Gas & Elec. Corp.*, 447 U.S. at 566; *cf. also X Corp.*, 116 F.4th at 901 (concluding, on a strict scrutiny standard, that California’s content moderation disclosure law was overly restrictive because “[c]onsumers would still be meaningfully informed if, for

⁸ Because we conclude that the Policy Disclosure Requirement would fail intermediate scrutiny under this reading, we need not separately evaluate whether it can survive strict scrutiny, which is even more demanding. *See Evergreen*, 740 F.3d at 245.

example, a company disclosed whether it was moderating certain categories of speech without having to define those categories in a public report”).

As for the second interest—preventing violence—the State’s goal is undoubtedly substantial, but the statute nonetheless fails even intermediate scrutiny because it is impermissibly “overly restrictive.” *Safelite Group*, 764 F.3d at 261. Of course, the State may regulate expression that meets the definition of incitement to violence, fighting words, and true threats. *See Brown*, 564 U.S. at 791; *see also Brandenburg*, 395 U.S. at 447–449 (incitement); *Chaplinsky v. New Hampshire*, 315 U.S. 568, 572 (1942) (fighting words); *Virginia*, 538 U.S. at 359 (true threats). But the Policy Disclosure Requirement is not effectively limited to constitutionally unprotected speech. That is, § 394-ccc(3) not only pertains to moderation of content that would “incite violence,” but it also affects moderation of content that might “humiliate” and “vilify” others. As Plaintiffs point out, humiliation and vilification could encapsulate expression like satire and campaigning that are within the First Amendment’s core protections.

And finally, the connection between the Policy Disclosure Requirement and facilitating reports that might prevent future horrific acts like the Buffalo shooting is tenuous at best. The Policy Disclosure Requirement does not require social

media networks to adopt policies disfavoring hateful speech—even unprotected threats and incitement. It simply requires them to disclose the policies they have.

iii. Weighing Competing Interpretations

The text of the statute is at least theoretically susceptible to both of the above interpretations. The content moderation policy required to be disclosed under the Hateful Conduct Law need only “*include[]* how such social media network will respond and address” reported “hateful conduct” on its platform. N.Y. Gen. Bus. Law § 394-ccc(3) (emphasis added). Arguably, a disclosed content moderation policy can be as broad and general as social media networks would like, as long as they are broad enough to cover reports of what the State has defined as “hateful conduct.” The State concedes that Plaintiff Volokh’s proposed policy of simply responding to *all* complaints “on [his] own discretion” would comply with the law. Appellant’s Br. at 44–45 (quoting J. App’x at 120). By the State’s interpretation, any policy that applies to *all* complaints, by logical necessity, also applies to complaints of what the statute would call “hateful conduct.” That also leaves open the possibility that a social media network could properly describe its moderation policies using whatever categories it chose, with a catch-all provision governing anything not specifically described.

The State urges us to accept this reading of the Policy Disclosure Requirement because, applying the doctrine of constitutional avoidance, courts should “strive to save a statute when confronted with a Free Speech challenge.” *People v. Marquan M.*, 24 N.Y.3d 1, 10 (2014); *see also People v. Dietze*, 75 N.Y.2d 47, 52 (1989) (“[A] statute ought normally to be saved by construing it in accord with constitutional requirements.”).

But constitutional avoidance has its limitations. “[O]ur *primary* consideration is to ascertain and give effect to the intention of the Legislature,” not to find a conceivable way to read a statute so that it survives judicial review. *Matter of DaimlerChrysler Corp. v. Spitzer*, 7 N.Y.3d 653, 660 (2006) (emphasis added). Interpreting a statute to avoid constitutional infirmity cannot give way “to wholesale revision of the Legislature’s enactment, rather than prudent judicial construction.” *Dietze*, 75 N.Y.2d at 53.

Interpreting the Policy Disclosure Requirement to require nothing more than disclosure of *any* content moderation policy, without tying that policy to hateful conduct as defined by the statute, renders much of the language of the Policy Disclosure Requirement, and the accompanying definition of “hateful conduct,” superfluous. This interpretation would read out of the statute the specific definition of “hateful conduct” that is arguably central to the intended

operation of the Hateful Conduct Law. Allowing a statute to say one thing but “stand for something entirely different” would be “especially intolerable in a statute regulating speech.” *Id.*; see also *Marquan M.*, 24 N.Y.3d at 10 (“[D]eparture from a textual analysis is appropriate only if the statutory language is fairly susceptible to an interpretation that satisfies applicable First Amendment requirements.”).

Moreover, while a generic requirement to disclose a content moderation policy would promote transparency, it would do little to address the State’s goal of reducing the risk of future atrocities like the Buffalo shooting. Ordinary principles of statutory interpretation disfavor such an interpretation. See, e.g., *Matter of Carver v. Nassau County Interim Finance Authority*, 142 A.D.3d 1003, 1008 (N.Y. App. Div. 2d Dep’t 2016) (“[A] court should avoid a statutory interpretation rendering the provision meaningless or defeating its apparent purpose.”). As to whether the interpretation of the Policy Disclosure Requirement that renders it constitutional is a permissible construction of the statute, or an impermissible “wholesale revision,” *Dietze*, 75 N.Y.2d at 53, we think the New York Court of

Appeals is better positioned to decide.⁹ “[T]he Supreme Court has warned against premature adjudication of constitutional questions when a federal court is asked to invalidate a State’s law, for the federal tribunal risks friction-generating error where it endeavors to construe a novel state Act not reviewed by the State’s highest court.” *Expressions Hair Design*, 877 F.3d at 106. “Because we believe that a state court is bound to have a better idea of the elasticity of the state’s statutes than a federal court, we leave this task to the New York Court of Appeals.” *Kuhne v. Cohen & Slamowitz, LLP*, 579 F.3d 189, 199 (2d Cir. 2009).

C. As-Applied Challenge to Report Mechanism Requirement

The Report Mechanism Requirement requires social media networks to “provide and maintain a clear and easily accessible mechanism for individual users to report incidents of hateful conduct,” which can be reached from “both a

⁹ These are not, as the dissent would say, “rhetorical questions.” Dissent at 6. The Court of Appeals has managed to devise limiting constructions for phrases as nebulous as “no legitimate purpose.” *People v. Stuart*, 100 N.Y.2d 412, 428–29 (2003); *see also People v. Shack*, 86 N.Y.2d 529, 538 (1995). It has similarly glossed statutes with implicit constitutional limitations not found in statutory language. *See Matter of Bell v. Waterfront Comm’n of N.Y. Harbor*, 20 N.Y.2d 54, 62 (1967) (holding that an anti-riot statute should be construed to prohibit “only that type of advocacy (1) which is intended to indoctrinate or incite to action in furtherance of the defined doctrine and (2) which is accompanied by a ‘clear and present danger’ of success”). And many of the constitutional-avoidance cases—including one the dissent relies on—were decided by a divided court, further demonstrating that these interpretive questions are not rhetorical, but close calls over which reasonable minds exercising sound judgment can differ. *See Dietze*, 75 N.Y.2d at 54–61 (Wachtler, C.J., concurring) (two judges concurring, arguing for the possibility of a saving construction); *Marquan M.*, 24 N.Y.3d at 12–15 (same) (Smith, J., dissenting).

social media network[’s] application and website, and shall allow the social media network to respond” to any complaints.¹⁰ N.Y. Gen. Bus. Law § 394-ccc(2).

The State reads the Report Mechanism Requirement (like the Policy Disclosure Requirement) to not require reference to the State’s definition of “hateful conduct” and urges us to conclude that the Report Mechanism Requirement is simply a regulation of conduct rather than speech. And the State contends that, although the reporting mechanism must *allow* the social media network to respond to individual complaints, there is no requirement that the social media network actually do so. The State does not attempt, in the alternative, to defend the Report Mechanism Requirement under either intermediate or strict scrutiny.¹¹

¹⁰ At the outset, we reject the State’s argument that Plaintiffs have no standing to challenge § 394-ccc(2). To have standing under Article III in these circumstances, Plaintiffs are required to show that they have engaged in conduct “arguably proscribed” by the challenged law (among other requirements). *Picard v. Magliano*, 42 F.4th 89, 98 (2d Cir. 2022). As we explain below, the Report Mechanism Requirement is at least ambiguous, and Plaintiffs’ argument about why their existing reporting mechanisms are insufficient are plausible. *See* Appellee’s Br. at 66–69.

¹¹ We agree with the dissent that if the Report Mechanism Requirement includes a mandate that social media networks review or respond to reports of objectionable conduct, then it is likely unconstitutional. Unlike the dissent, however, we are open to the possibility that the words “shall allow” suggest a technical capability to get in touch with a reporting user, rather than a requirement to review or respond to incoming reports. *Cf.* Dissent at 13. Although the statute

In contrast, Plaintiffs suggest that the statute calls for a reporting mechanism to “be created specifically ‘for’ reporting ‘hateful conduct,’” thus compelling social media networks to adopt the State’s view of what speech should be reported and potentially addressed by social media networks. Appellee’s Br. at 24. And Plaintiffs contend that the Hateful Conduct Law requires that social media networks respond to complaints conveyed through the reporting mechanism. In Plaintiffs’ view, the Report Mechanism Requirement is subject to strict scrutiny, and it fails.

As with the Policy Disclosure Requirement, whether the Report Mechanism Requirement is constitutional depends on how we interpret that section of the statute. If the State’s interpretation is right, then the Report Mechanism Requirement does not regulate or compel speech, and is constitutional. If the Plaintiffs are right, it is a content-based regulation of speech and fails constitutional scrutiny. As with the Policy Disclosure Requirement, the New York Court of Appeals is best positioned to tell us what the statute requires.

may seek to make it easier for social media networks to respond to complaints by assuring that they have the technical ability to do so, we see nothing in the statute that requires that the social media networks respond to complaints in any particular way. Moreover, the dissent’s suggestion that the Report Mechanism Requirement impermissibly forces bloggers “to open themselves to criticism or mandatory dialogue” makes no sense. Dissent at 13. The only bloggers who even qualify as social media networks under the statute are those who allow readers to post their own comments in response to the bloggers’ (or fellow readers’) posts. *See* N.Y. Gen. Bus. Law § 394-ccc(1)(b).

“[T]he First Amendment does not prevent restrictions directed at commerce or conduct from imposing incidental burdens on speech.” *Sorrell v. IMS Health Inc.*, 564 U.S. 552, 567 (2011); *see also FAIR*, 547 U.S. at 67. If a social media company can comply with the Report Mechanism Requirement by maintaining a reporting mechanism where consumers may direct complaints—whether or not they are complaints about content meeting the State’s definition of hateful conduct—and if the statute does not require social media networks to *respond* to individual complaints, then the Report Mechanism Requirement does not regulate speech at all. The requirement of a generic avenue for consumer complaints is no more an expressive requirement than requiring universities to allow military recruiters to access their campuses. *FAIR*, 547 U.S. at 65–70.

On the other hand, if social media networks are required to provide dedicated reporting mechanisms for “hateful conduct,” as defined by the State, or if the statute requires the networks to respond to reports, then the Report Mechanism Requirement would force Plaintiffs to promote the State’s view as to categories of speech that should be discouraged, reported, and responded to. That is, social media networks would be required to indicate to consumers “what kinds of speech, including what viewpoints” ought to be reported and “are not worthy of promotion.” *Moody*, 603 U.S. at 736 n.5.

Thus understood, the Report Mechanism Requirement would compel speech, and would be subject to intermediate or strict scrutiny. Because the burden is on the State to justify its regulation of expressive conduct, *Brown*, 564 U.S. at 799, and for the same reasons set forth above, the Report Mechanism Requirement would not meet this burden, and would run afoul of the First Amendment.

For different reasons than the Policy Disclosure Requirement, but with similar effect, the constitutionality of the Report Mechanism Requirement turns on whether the statute can support the narrower, constitutional interpretation. That is, can the statutory requirement of a “mechanism for individual users to report incidents of hateful conduct,” N.Y. Gen. Bus. Law § 394-ccc(2), be satisfied by a generic reporting avenue? And does the requirement that the social media network’s disclosure describe “how such social media network will respond and address the reports of incidents of hateful conduct on their platform,” *id.* § 394-ccc(3), combined with a requirement that the reporting mechanism “allow the social media network to provide a direct response to any individual reporting hateful conduct informing them of how the matter is being handled,” *id.* § 394-ccc(2), suggest that the statute *requires* the social media network to respond to reports of hateful conduct as defined by the statute?

For the reasons set forth above, the New York Court of Appeals is best positioned to provide an authoritative answer to these questions.

II. Facial Challenges

In addition to ruling that enforcement of the Hateful Conduct Law would likely violate the First Amendment rights of the Plaintiffs, the district court concluded that the law is *facially* invalid and enjoined *any* application of the law.¹² *See Volokh*, 656 F. Supp. 3d at 444–46. The district court’s reasoning appears to be rooted primarily in vagueness and overbreadth doctrines, and it focuses on the Hateful Conduct Law’s downstream effect on social media users who are not parties to this case and who are not themselves regulated by the express terms of the Hateful Conduct Law. *See id.*

The district court reasoned that the Hateful Conduct Law “could have a profound chilling effect on social media users” because “[e]ven though the law . . . does not impose liability on users for engaging in ‘hateful conduct,’ the state’s targeting and singling out of this type of speech for special measures certainly could make social media users wary about the types of speech they feel free to engage in without facing consequences from the state.” *Id.* at 445. And it observed

¹² We do not decide today whether the district court’s injunction is impermissibly “universal” under the Supreme Court’s recent decision in *Trump v. CASA, Inc.*, 145 S. Ct. 2540, 2551 (2025). We leave the question for the parties to litigate before the district court in the first instance.

that any chilling effect would be “exacerbated by the indefiniteness of some of the Hateful Conduct Law’s key terms,” namely “vilify” and “humiliate.” *Id.* at 446.

“[T]he overbreadth doctrine instructs a court to hold a statute facially unconstitutional even though it has lawful applications, and even at the behest of someone to whom the statute can be lawfully applied.” *United States v. Hansen*, 599 U.S. 762, 769 (2023). It applies when a restriction on speech “prohibits a substantial amount of protected speech relative to its plainly legitimate sweep.” *Id.* at 770. “Invalidation [of a statute] for overbreadth is strong medicine that is not to be casually employed.” *Id.* “To justify facial invalidation, a law’s unconstitutional applications must be realistic, not fanciful, and their number must be substantially disproportionate to the statute’s lawful sweep.” *Id.*

The vagueness doctrine requires a court to prohibit enforcement of a law that “is so fatally indefinite that it cannot constitutionally be applied to anyone,” *Copeland v. Vance*, 893 F.3d 101, 110 (2d Cir. 2018), and in turn has “a substantial chilling effect on protected conduct,” *Farrell v. Burke*, 449 F.3d 470, 497 (2d Cir. 2006) (Sotomayor, J.); *see also id.* at 496–97 (“Facial vagueness challenges may go forward only if the challenged regulation reaches a substantial amount of constitutionally protected conduct.”).

We need not decide whether vagueness and overbreadth—applied from the perspective of non-party social media users whose speech is not regulated by the Hateful Conduct Law—provide a fitting framework for analyzing this statute and invalidating it on its face. The district court’s analysis implicitly rests on the view that the statute requires social media networks to incorporate the State’s definition of hateful conduct into its content moderation policy disclosures and reporting mechanism. *See, e.g., Volokh*, 656 F. Supp. 3d at 441 (“[T]he Hateful Conduct Law requires a social media network to endorse the state’s message about ‘hateful conduct.’”); *id.* (“Implicit in this language is that each social media network’s definition of ‘hateful conduct’ must be at least as inclusive as the definition set forth in the law itself.”); *id.* (“[Plaintiff Locals’ policies] would need to be modified to be brought into compliance with the law by including speech that potentially vilifies or humiliates a group or individual.”).

But as noted above, the State disavows this interpretation of the law and contends that the statute merely requires disclosure of a social media network’s content moderation policy and provision of a reporting mechanism, without requiring the social media network to incorporate the State’s definition of “hateful conduct.” If the State is right, then the suggestion that the Hateful Conduct Law would chill the speech of social media users would lose its force. And we need

not at this juncture assess whether, if the *Plaintiffs'* view of the statute's requirements is correct, the statute is facially unconstitutional on account of its impact on social media users.

III. Certification

The threshold questions about what the Policy Disclosure Requirement and the Report Mechanism Requirement actually require satisfy the criteria for certification. Our decision whether to certify is guided by the following factors: “(1) whether the New York Court of Appeals has addressed the issue; (2) whether the question is of importance to the state and may require value judgments and public policy choices; and (3) whether the certified question is determinative of a claim before us.” *Government Employees Insurance Company v. Mayzenberg*, 121 F.4th 404, 420 (2d Cir. 2024) (“*GEICO*”).

No New York court considered how the Hateful Conduct Law may be interpreted before it was enjoined by the district court shortly after it became effective. *See Nicholson v. Scoppetta*, 344 F.3d 154, 167 (2d Cir. 2003) (“Normally this Court ought not to consider the Constitutionality of a state statute in the absence of a controlling interpretation of its meaning and effect by the state courts.”).

The question as to whether the Hateful Conduct Law can support the constructions that would render the contested provisions constitutional, while a

legal question, involves consideration of the value judgments and public policy choices of New York lawmakers—questions best evaluated by New York courts. *Kuhne*, 579 F.3d 199 (allowing the Court of Appeals to determine the “elasticity” of a local statute).

Finally, an authoritative interpretation of the Hateful Conduct Law may be “determinative” of this appeal. *See GEICO*, 121 F.4th at 420. As we have explained, the likely constitutionality of the statute depends on what its challenged terms actually require. Answers to the questions below would be dispositive with respect to the as-applied constitutional challenges, and they may be dispositive as to the facial challenge.¹³

We thus **CERTIFY** the following questions to the New York Court of Appeals pursuant to our Local Rule 27.2 and 22 N.Y.C.R.R. § 500.27(a):

(1) Does a social media network comply with N.Y. Gen. Bus. Law § 394-ccc(3)’s requirement to publish a “clear and concise policy . . . which includes how such social media network will respond and address the reports of incidents of hateful conduct on their platform” if its policy does not explicitly reference or address the

¹³ The dissent suggests that the entirety of our analysis is dicta because we have not resolved the appeal but have merely asked the Court of Appeals of New York to accept and answer certified questions. Dissent at 4 n.3. But the canon of constitutional avoidance requires that we construe statutes to avoid constitutional defects if “there is another reasonable interpretation available.” *Kennedy v. Braidwood Management, Inc.*, 145 S.Ct. 2427, 2451 (2025). So our determination of the constitutionality of the hypothetical interpretation of Section 394-ccc proffered by the State is necessary to our disposition of this appeal. Thus, the only “hypothetical”—that is, yet undecided—aspect of the opinion is the true meaning of the statute.

content encompassed by the statute's definition of "hateful conduct" and does not otherwise address content that encompasses this defined category?

(2) Does a social media network comply with N.Y. Gen. Bus. Law § 394-ccc(2)'s requirement to "provide and maintain a . . . mechanism for individual users to report incidents of hateful conduct" if that mechanism does not explicitly reference or address the content encompassed by the statute's definition of "hateful conduct" and does not specifically state that content meeting the statute's definition of "hateful conduct" may be reported using that mechanism?

(3) Does N.Y. Gen. Bus. Law § 394-ccc require a social media network to provide a direct response to any individual reporting hateful conduct informing them of how the matter is being handled?

By so formulating the questions, we do not intend to limit the scope of the Court of Appeals' analysis, and we invite the Court of Appeals to reformulate or expand upon these questions as it deems appropriate. This panel retains its jurisdiction to decide this appeal once we have had the benefit of the New York Court of Appeals' views, or it declines to accept certification.

It is therefore **ORDERED** that the Clerk of this Court transmit to the Clerk of the Court of Appeals of the State of New York a Certificate, as set forth below,

together with this opinion, a complete set of briefs and appendices, and the record filed in this Court by the parties.

CERTIFICATE

The foregoing is hereby certified to the Court of Appeals of the State of New York pursuant to Second Circuit Local Rule 27.2 and New York Codes, Rules, and Regulations Title 22, section 500.27(a), as ordered by the United States Court of Appeals for the Second Circuit.

DENNIS JACOBS, *Circuit Judge*, dissenting:

I respectfully dissent.

To whack the moles of hate, the State of New York proposes to put regulation of internet content in the hands of the state’s elected Attorney General. The district court has temporarily enjoined this “Hateful Conduct Law,”¹ and the panel today unanimously confirms that the First Amendment forbids such a measure.

However, rather than affirm the district court’s preliminary injunction, the majority certifies questions to the New York Court of Appeals, asking whether the statute might be creatively read to reduce the infringement on speech. It cannot. I dissent from the certification of questions because no answer that the Court of Appeals could give to the questions posed would advance the inquiry. I would instead affirm; Judge Carter got it right.

I.

New York’s Hateful Conduct Law governs all “service providers, which, for profit-making purposes, operate internet platforms that are designed to enable users to share any content with other users or to make such content available to the public,” and which “conduct[] business in [New York]” from anywhere in the world. N.Y. Gen. Bus. Law § 394-ccc(1)(b) and (2). While the statute refers to these businesses by the defined term “social media network,” that label is grossly underinclusive; the statute apparently applies to any blog. *Id.*

The statute requires all bloggers to “have a clear and concise policy readily available and accessible on their website and application which includes how [they] will respond and address [] reports of incidents of hateful conduct on their platform.” *Id.* § 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.” *Id.* § 394-ccc(1)(a). Bloggers must as well “provide and maintain a clear and easily

¹ I adopt the Majority’s defined terms.

accessible mechanism for individual users to report incidents of hateful conduct,” so that the platform can duly comply with its stated policy. *Id.* § 394-ccc(2). Each knowing violation of the statute subjects the blogger to “a civil penalty . . . by the attorney general not to exceed one thousand dollars” per day. *Id.* § 394-ccc(5). This is supposedly a consumer protection measure.

The Hateful Conduct Law mandates that each blog’s disclosed policy be *accurate*. In the words of New York’s counsel, a “consumer” must “be able to look at that policy and say, ‘okay, I understand what the network is going to do if I report hateful conduct as defined by the statute.’” Oral Arg. at 2:47-2:55; *see id.* at 16:28-16:35. So, “if [a network’s] policy was a lie,” “that would lead to liability under the statute.” *Id.* at 17:16-20. In that event, “[t]he Attorney General would be able to enforce the statute through her enforcement powers in the statute in Section 5 with civil fines and so forth,” *id.* at 17:02-17:10--i.e., through daily penalties of up to one thousand dollars per violation, N.Y. Gen. Bus. Law § 394-ccc(5).²

II.

I agree with the majority that “[t]he most natural reading” of this statute is that it requires blogs to “develop and publish policies that specifically address hateful conduct, as that term is defined by statute.” Majority at 30. Indeed, I see no other reading. I further agree that requiring moderation policies and reporting mechanisms that speak to the state’s definition of hateful conduct is unconstitutional. *See* Majority at 30, 48. In my view, the whole thing is about as constitutional as dukedoms.

New York acknowledges the attendant “‘constitutional difficulties’” raised by a straightforward reading of the words in the statute; so the state’s backpedaling lawyers urge this Court to construe the Hateful Conduct Law so that it does not “require a network’s policy to reference the statute’s definition of hateful conduct.” Appellant’s Br. at 45 n.5 (quoting *Skilling v. United States*, 561 U.S. 358, 406 (2010)). Accordingly, the only seriously contested question on

² The Hateful Conduct Law is part of a wider regulatory program. *See generally* New York Senate Bill S895B, codified at Article 42 (Sections 1100–1104) (New York statute, signed December 21, 2024, requiring disclosure of whether and how social media companies’ terms of service address “hate speech”).

appeal is whether the Hateful Conduct Law is amenable to some kind of a saving construction.

New York asks us to save the statute by reading it so that a regulated platform is not “required to adopt or even reference GBL § 394-ccc(1)(a)’s definition of hateful conduct in its policy” or “in making a report mechanism available.” Appellant’s Br. 26, 44. But before we can rely on an interpretation to save a statute, the statute must be “readily susceptible to such a construction.” *United States v. Stevens*, 559 U.S. 460, 481 (2010) (citation modified). This one is not. A policy that does not “even reference” the “definition of hateful conduct” would hardly constitute the “*clear and concise* policy” on moderation that the text of the statute demands. See N.Y. Gen. Bus. Law § 394-ccc(3) (emphasis added). Among other objections, allowing such a counterintuitive limiting construction would “sharply diminish” the New York legislature’s “incentive to draft a narrowly tailored law in the first place.” *Stevens*, 559 U.S. at 481 (citation modified).

Regardless of any judicial salvage job, every lay reader would understand this statute to forbid protected speech--not least because it is *titled* “Social media networks; hateful conduct prohibited.” The clear statutory wording would therefore “have a substantial chilling effect on protected conduct.” *Farrell v. Burke*, 449 F.3d 470, 497 (2d Cir. 2006) (explaining that such speech restrictions are unconstitutionally vague on their face). In addition to the chill on those who read it, the words on the books empower litigious censors to invoke the statute in “hate speech” strike suits. It is no comfort that bloggers might defeat such litigation by hiring constitutional lawyers at going rates.

III.

All agree that the statute’s words are an unconstitutional speech restriction. New York’s Court of Appeals has told us not to resuscitate such statutes through judicial opinion. See *People v. Dietze*, 549 N.E.2d 1166, 1169 (N.Y. 1989) (rejecting a saving construction of a speech regulation as “unacceptably vague” where “the statutory language would signify one thing but, as a matter of

judicial decision, would stand for something entirely different”). I would therefore affirm.³

³ The questions asked of the New York Court of Appeals posit a statute that is radically rewritten and reconceived. Since no such statute has been presented to us, the hypothesis that it would be constitutional is dicta. Cf. *United States v. Johnson*, No. 22-1289, 2025 WL 1921533, at *1 (2d Cir. July 14, 2025) (op. of Lohier, J., joined as to this part by Robinson and Nathan, JJ.) (“[T]he question . . . remains an open one in this Circuit because the panel opinion’s statements bearing on a hypothetical . . . are clearly dicta.”).

A question is not a holding; and a certification is not a disposition of the case or the issues presented. So the majority’s reasoning would remain dicta whether or not the New York Court of Appeals blesses the majority’s redrafting. After all, “[a] panel opinion’s assertion is decidedly *not* a holding of this Court” unless it necessarily disposes of “the case before it.” *Id.* at *2 (op. of Lohier). Our opinions certifying questions to the Court of Appeals make clear that we do not yet reach a “*disposition* of the case before [us].” See, e.g., *Ferreira v. City of Binghamton*, 975 F.3d 255, 291 (2d Cir. 2020) (“[T]he question certified will control our *disposition* of the case.” (emphasis added)); *Beck Chevrolet Co. v. Gen. Motors LLC*, 787 F.3d 663, 682 (2d Cir. 2015) (“The *disposition* of this case could have a substantial impact . . . in New York State. . . . We accordingly certify the following two questions” (emphasis added)); see also *Adelson v. Harris*, 774 F.3d 803, 811 & n.5 (2d Cir. 2014) (“Two unresolved questions of state law may be determinative of the instant appeal and are hereby respectfully CERTIFIED Costs shall abide final *disposition* of the case.” (emphasis added)).

The hypothetical nature of such a certification is easily demonstrated. A certification can run by the New York court more than one alternative reading; and the New York court is expressly invited to reframe the certified questions (and thus the statute) as it wills. See Majority at 56 (“[W]e invite the Court of Appeals to reformulate or expand upon these questions as it deems appropriate.”).

If (as seems correct to me) the New York court advises that the statute will not bear the reading posited by the majority here, our certification will not bind a

IV.

Nevertheless, the majority agrees with New York that the Hateful Conduct Law might be read to require no more than a virtual complaint box, plus a disclosed “policy of simply responding to *all* complaints” at the platforms’ “own discretion.” Majority at 43 (quoting Appellant’s Br. at 44-45); *see* J. App’x at 120. Contrary to fact, I will assume that the majority is correct.

The majority holds that New York’s proposed saving construction would be constitutional, even though the natural interpretation of the statutory wording would not. So the majority enlists the New York Court of Appeals to say which interpretation is correct, with the idea that the answer would decide this appeal.

That is itself error. Before we certify questions to a state’s highest court, we first ask “whether state-law principles of statutory interpretation and related precedents permit [us] to predict how the forum state’s highest court would decide the point at issue.” *Reyes v. City of New York*, 141 F.4th 55, 68 (2d Cir. 2025) (citation modified). We certify when “we cannot confidently predict how the [state’s highest court] will interpret the relevant statute, and no controlling precedent from [that court] resolves” the case. *Loomis v. ACE Am. Ins. Co.*, 91 F.4th 565, 569 (2d Cir.) (Robinson, J., joined by Nathan, J.), *certified question answered*, 244 N.E.3d 908 (Ind. 2024).

This case is not certifiable because the majority’s resolution of the constitutional issue leaves no lingering questions of statutory interpretation for the New York court. “Under the federal and New York State canons of constitutional doubt, a statute must be construed, if fairly possible, so as to avoid not only the conclusion that it is unconstitutional, *but also grave doubts upon that score.*” *1256 Hertel Ave. Assocs., LLC v. Calloway*, 761 F.3d 252, 260-61 (2d Cir. 2014)

future federal court to the constitutionality (or not) of some future enactment that conforms to the text of the hypothetical statute offered up for consideration by the Court of Appeals.

(emphasis added) (citation modified). We accordingly apply a “presumption that legislatures are cognizant and respectful of constitutional limitations.” *Id.* at 261. If there were any doubt that this canon controls, it is codified in the text of the Hateful Conduct Law: “Nothing in this section shall be construed [] as an obligation imposed on a social media network that adversely affects the rights or freedoms of any persons, such as exercising the right of free speech pursuant to the first amendment to the United States Constitution.” N.Y. Gen. Bus. Law § 394-ccc(4)(a).

Since the majority concludes that there is one interpretation that is constitutional and one that is not, and since the statute must be construed to avoid even a doubt about its constitutionality, the majority necessarily should “confidently predict” (without certification) that the Court of Appeals would adopt the reading that is supposedly constitutional. The majority’s choice to certify anyway is not without cost. An appeal from a preliminary injunction is a remedy “granted under the theory that there is an urgent need for speedy action.” *Citibank, N.A. v. Citytrust*, 756 F.2d 273, 276 (2d Cir. 1985). Certification will “inevitably delay[] proceedings,” thwarting “prompt resolution” of this appeal. *Cruz v. Banks*, 134 F.4th 687, 697 (2d Cir.), *certified question accepted*, 259 N.E.3d 1122 (N.Y. 2025).

V.

Even if I were willing to indulge the state’s rewrite of its statute and to certify rhetorical questions, I would affirm, not certify. The Hateful Conduct Law, when read as the state urges, is still unconstitutional. It is immaterial which unconstitutional reading is correct.

Under the state’s reading, the Hateful Conduct Law requires bloggers to commit in advance to how they will curate speech. It threatens bloggers with punishing fines if they fail to police content in precise accordance with their stated moderation policies. And it grants the state enforcement discretion over whether bloggers’ content moderation decisions live up to those policies. Through compliance costs and fines, the statute burdens speech--and the weight of the burden depends on the content of the speech. The Hateful Conduct Law is thus unconstitutional as applied and on its face, even if (in the words of the majority) it “merely requires social media networks to publicly disclose their content moderation policies, whatever those policies may be, and contains no

requirement that those policies actually specifically address ‘hateful conduct,’ as defined by the statute.” Majority at 30.

A.

The business end of the Hateful Conduct Law is New York’s confessed plan for enforcing it: imposing “liability” of up to one thousand dollars daily, per violation, on any platform that (New York contends) disclosed a “policy [that] was a lie.” Oral Arg. at 17:16-20. In other words, the statute levies fines for falling short of the blogger’s announced policy--or, as the statute puts it, “liability” for “the failure to provide a mechanism for a user to report to the social media network any incidents of hateful conduct on their platform and to receive a response on such report.”⁴ N.Y. Gen. Bus. Law § 394-ccc(4).

The Hateful Conduct Law’s constitutional deficiencies thus become manifest. It discriminates based on content by (1) deterring blogs from speaking on incendiary (or even interesting) topics; (2) forcing bloggers to commit in advance to how they will moderate future posts, thus restraining their future speech; and (3) deterring blogs from employing certain content moderation policies. Meanwhile, the law is too vague to obey, or to constrain arbitrary enforcement.

This is all elementary: the First Amendment, applicable to the States through the Fourteenth Amendment, prohibits the enactment of laws “abridging the freedom of speech.” U.S. Const. amend. 1. A state has “no power to restrict expression because of its messages, its ideas, its subject matter, or its content.” *Nat’l Inst. of Fam. & Life Advocs. v. Becerra* (“NIFLA”), 585 U.S. 755, 766 (2018) (quoting *Reed v. Town of Gilbert*, 576 U.S. 155, 163 (2015)). Laws that “target speech based on its communicative content” are therefore “presumptively unconstitutional.” *Id.* (quoting *Reed*, 576 U.S. at 163).

Because “[l]awmakers may no more silence unwanted speech by burdening its utterance than by censoring its content,” *Sorrell v. IMS Health Inc.*, 564 U.S. 552, 566 (2011), the presumption of unconstitutionality applies when the state “imposes a financial burden on speakers because of the content of their

⁴ The statute obligingly disclaims adding or increasing liability “for anything other than” such failure. N.Y. Gen. Bus. Law § 394-ccc(4).

speech.” *Simon & Schuster, Inc. v. Members of N.Y. State Crime Victims Bd.*, 502 U.S. 105, 115 (1991). For example, *Simon & Schuster* struck down New York’s restriction on the ability of criminals to earn royalties on their memoirs because it “establishe[d] a financial disincentive to create or publish works with a particular content.” 502 U.S. at 118.

Like the royalty restriction in *Simon & Schuster*, the Hateful Conduct Law disincentivizes speech based on its content. As the Supreme Court recently confirmed, a website’s “presenting a curated and edited compilation of third party speech is itself protected speech.” *Moody v. NetChoice, LLC*, 603 U.S. 707, 744 (2024) (citation modified). “When the government interferes with [a website’s] editorial choices,” it “alters the content of the compilation.” *Id.* at 731-32. The Hateful Conduct Law does that by requiring blogs, under pain of fines, to comply with their disclosed content moderation policies. Blogs’ cost of compliance varies with: (1) the subjects on which they speak; (2) their moderation decisions; and (3) their moderation policies. Bloggers may accordingly “avoid or mitigate the effects of the [a]ct by altering their speech”; so the act “impose[s] a . . . burden by reason of content.” *TikTok Inc. v. Garland*, 145 S. Ct. 57, 67 (2025) (citation modified).

Subjects. The more a blog allows free-wheeling discussion of “race, color, religion, ethnicity, national origin, disability, sex, sexual orientation, gender identity or gender expression,” the more complaints it will field (through the Section 394-ccc(2) report mechanism) about contributions that are said to “vilify, humiliate, or incite violence” based on these categories--i.e., statutory “hateful conduct.” Since, as explained, a blog must comply with its stated policy whenever a user reports hateful conduct, more complaints means more time and cost spent on compliance with the Hateful Conduct Law’s obligation to moderate in conformity with the platform’s stated policy. And it means more pitfalls.

Moderation Decisions. The Hateful Conduct Law compels bloggers to speak in accordance with their previously announced policies, even when intervening events would lead them to speak otherwise. Bloggers who discover from experience that their content moderation policies are too lax face fines if they revise their policy and apply it retroactively to third-party comments that they would rather not host; the policy in place when the offending comments were posted would then be “a lie.” Oral Arg. 17:16-20. Ironically, a blogger who

belatedly decides that a third party's "hateful conduct" might inspire a hate crime may remove that "hateful conduct" *only if* the blogger previously announced an intention to do so.

Moderation Policies. The Hateful Conduct Law also incentivizes blogs to drain any nuance from their moderation policies. If a blog insists on including some third-party speech and excluding other speech, it can never be sure if the Attorney General will deem the blog's disclosed moderation policy to be "a lie." Oral Arg. 17:16-20. Each moderation decision requires a subjective judgment with which the Attorney General is enabled to disagree, with ruinous consequences. Accordingly, it is safer--and thus cheaper--to choose a content moderation policy that requires no subjective judgments, such as a prohibition on all third-party comments using the word "race," or "gender," or even a prohibition on third-party commenting altogether.

Even the very moderation policies that the Hateful Conduct Law appears designed to encourage--those applying the state's definition of "hateful conduct"--invite liability. Again, speech is "hateful conduct" if it "vilif[ies], humiliate[s], or incite[s] 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). And whether language "vilifies" or "humiliates" will often vary with the sensitivities of the listener.⁵

New York's (telling) advice to platforms is to simply adopt a policy of "*no formal policy at all* and appl[y] its editorial discretion on a case-by-case basis."

⁵ Not even the law's sponsors can be sure that their speech is not "hateful conduct." For example, one sponsor, then-Assemblywoman Patricia Fahy, posted to a social media platform about the bill: "Social media companies aren't doing enough to combat the spread of hate, racism, and white supremacy on their platforms. Authorities say the #Buffalo shooter published a document online that included the Great Replacement Theory." Senator Patricia Fahy (@PatriciaFahy46), X (May 16, 2022, at 11:27 AM ET), <https://x.com/PatriciaFahy46/status/1526222894748925952>. "White supremacy" and the "Great Replacement Theory" are terms laden with pejorative, race-based animus.

Majority at 33. That is a confession that the Hateful Conduct Law “impose[s] a . . . burden by reason of content”: New York concedes that bloggers can “avoid or mitigate the effects of the [a]ct by altering” their curation of speech. *TikTok*, 145 S. Ct. at 67 (citation modified). Moreover, bloggers who take New York’s advice are still not safe: if there *is* a logic underlying a blog’s curation decisions, then a non-policy is “a lie.” The Attorney General thus bears a lethal weapon: she may claim that any blogger who disclaims a hateful conduct policy secretly has one, and has failed to disclose it.

No blogger can be free of jeopardy for the further reason that the regulation (enforced by ruinous fines) is intolerably opaque. A lawful speech restriction must “provide meaningful notice of the scope of [its] prohibition” and “meaningful limits on an enforcing officer’s discretion.” *Farrell*, 449 F.3d at 498. New York’s recast of the Hateful Conduct (as I have explained, *supra* Part II) requires bloggers to consult judicial opinions in addition to the statute’s wording in order to figure out their obligations. But identifying their obligations is still not enough for confident compliance: bloggers cannot know whether they are *complying* with their obligations, or--per the Attorney General--just *lying*.

B.

The Hateful Conduct Law thus violates the First Amendment. The statute chills disfavored speech through multiple, mutually-reinforcing components: (1) it requires blogs to accept reports of hateful conduct; (2) those reports trigger the blog’s obligation to moderate in compliance with a disclosed moderation policy; (3) and every moderation decision entails jeopardy to investigation or fines. When, as here, a state “establishes a financial disincentive to create or publish works with a particular content,” the state must satisfy strict scrutiny, i.e. “must show that its regulation is necessary to serve a compelling state interest and is narrowly drawn to achieve that end.” *Simon & Schuster*, 502 U.S. at 118 (citation modified). The majority and I agree that New York cannot satisfy strict scrutiny--or, for that matter, even intermediate scrutiny, which would require New York to show that its law “directly advances a substantial governmental interest and is not overly restrictive,” *Safelite Group, Inc. v. Jepsen*, 764 F.3d 258, 261 (2d Cir. 2014). See Majority at 40-41, 41 n.8.

New York insists that it need satisfy only *Zauderer* scrutiny, a standard that is less restrictive than intermediate scrutiny, and that applies only to mandatory

commercial disclosures of “purely factual and uncontroversial information about the terms under which services will be available.” *CompassCare v. Hochul*, 125 F.4th 49, 65 (2d Cir. 2025) (citation modified) (citing *Zauderer v. Off. of Disciplinary Couns. of Sup. Ct. of Ohio*, 471 U.S. 626, 651 (1985)). Under *Zauderer*, a state “has the burden to prove” that a commercial disclosure mandate “is neither unjustified nor unduly burdensome,” albeit without the obligation to show that the statute is otherwise tailored. *NIFLA*, 585 U.S. at 776 (citing *Zauderer*, 471 U.S. at 651).

Zauderer is no fit for the Hateful Conduct Law. Considered as a whole, the Hateful Conduct Law looks nothing like the compelled commercial disclosures we have analyzed under *Zauderer*. For example, we applied *Zauderer* in *New York State Restaurant Ass’n v. New York City Board of Health* (“NYSRA”), which rejected a First Amendment challenge to a law requiring certain restaurants to disclose calorie counts on the menu. *See* 556 F.3d 114, 134 (2d Cir. 2009). That law did not (for instance) increase compliance costs or legal jeopardy depending on the restaurant’s choice of menu items. True, the idea was that informed consumers would voluntarily buy fewer high-calorie dishes, *see, e.g., id.* at 134-36, so the law disproportionately harmed restaurants selling decadences. But NYSRA still applied *Zauderer* scrutiny because mandatory disclosure of calorie counts “further[ed], rather than hinder[ed], the First Amendment goal of the discovery of truth and contribute[d] to the efficiency of the ‘marketplace of ideas.’” *Id.* at 132 (quoting *Nat’l Elec. Mfrs. Ass’n v. Sorrell* (“NEMA”), 272 F.3d 104, 114 (2d Cir. 2001)).

The Hateful Conduct Law is not such a commercial disclosure measure. Though it requires a blog to disclose a “hateful conduct” moderation policy, that commercial disclosure is just the hook for the state to police *non-commercial* content that the blog curates. Such “a financial disincentive to create or publish works with a particular content” *distorts* the marketplace of ideas, and must therefore satisfy strict scrutiny, not merely *Zauderer*. *See Simon & Schuster*, 502 U.S. at 118 (citation modified).

The majority sidesteps this analysis because the majority never evaluates the Hateful Conduct Law as a whole. Instead, the majority dissects the statute and assesses in isolation the constitutionality of (New York’s reading of) [i] its policy disclosure requirement and [ii] its report mechanism requirement. *See*

Majority at 30, 48. The majority concludes that under New York's reading, the policy disclosure requirement must satisfy only *Zauderer*, see *id.* at 34, while the report mechanism requirement "does not regulate speech at all," *id.* at 49.

The majority asks the wrong questions. We do not decide whether a discrete component of a regulatory regime "could survive constitutional scrutiny if it existed separately." *Church of Lukumi Babalu Aye, Inc. v. City of Hialeah*, 508 U.S. 520, 540 (1993). If a component "functions, with the rest of the enactments in question, to suppress" First Amendment rights, "it must be invalidated." *Id.*

That said, even bifurcation cannot save the Hateful Conduct Law. Its policy disclosure requirement still compels speech in violation of the First Amendment--even when that provision is assessed in isolation from the report mechanism requirement or the Attorney-General's enforcement powers--because it is not a true *Zauderer* disclosure.

Zauderer governs only compelled disclosures of "purely factual and uncontroversial information," like the caloric content of food, see *NYSRA*, 556 F.3d at 134, or the presence of mercury in light bulbs, see *NEMA*, 272 F.3d at 107, 113-14. A mandate to disclose a content moderation policy is different in kind. A light bulb either contains mercury, or it does not, whereas virtually all "hateful conduct" moderation policies are inherently subjective. Even a straightforward policy of removing *all* statutory hateful conduct requires subjective judgments, because it is inherently contestable whether a statement "vilifies" or "humiliates" within the statutory definition. N.Y. Gen. Bus. Law *Id.* § 394-ccc(1)(a). The Attorney General is empowered to evaluate a blog's actual moderation decisions and to disagree that the blog's policy is as the blog disclosed. Thus, "the fact that [the policies] are what they are," Majority at 32, is itself controversial and subject to challenge. *Contra id.*

It follows that *Zauderer* does not apply. And if it does not, then--as the majority acknowledges--New York must satisfy at least intermediate scrutiny. See Majority at 40-41. And--as the majority agrees--New York cannot do that. See *id.*

Even if *Zauderer* governed, the Hateful Conduct Law would violate the First Amendment nevertheless. New York "has the burden to prove" that any *Zauderer* disclosure mandate "is neither unjustified nor unduly burdensome."

NIFLA, 585 U.S. at 776. Such a disclosure mandate must be “no broader than reasonably necessary,” to avoid “chilling protected speech.” *Id.* (citation modified). Here, New York cannot discharge even that burden.

As I have explained, *supra* Section V.A., the policy disclosure requirement is the vehicle that allows the state to disincentivize speech on certain topics. This mandatory disclosure “chill[s] protected speech,” and is thus unduly burdensome. *See NIFLA*, 585 U.S. at 776. Accordingly, under any level of scrutiny that might apply, New York’s strained reading of the Hateful Conduct Law is still unconstitutional.

C.

Since the policy disclosure requirement renders New York’s reading of the Hateful Conduct Law unconstitutional, I need not reach the majority’s contention that the report mechanism requirement, as New York would construe it, “does not regulate speech at all.” Majority at 49. But I cannot help myself: yes, it does. That provision mandates that platforms “shall provide and maintain a clear and easily accessible mechanism for individual users to report incidents of hateful conduct.” N.Y. G.B.L. § 394-ccc(2). The mechanism “shall allow the social media network to provide a direct response” to the reporter. *Id.* The requirement that the mechanism “shall allow” a response means that the platform cannot send reports to a black box: they must be reviewed. Accordingly, bloggers must make themselves available for criticism.

The First Amendment protects against just this sort of regulation. In general, “individuals are not required to welcome unwanted speech into their own homes.” *Frisby v. Schultz*, 487 U.S. 474, 485 (1988). Forcing bloggers to open themselves to criticism or mandatory dialogue is itself impermissible: “[i]t is uncontroversial that the First Amendment protects the right to speak anonymously.” *Antonyuk v. James*, 120 F.4th 941, 1003 (2d Cir. 2024). The State has no legitimate interest in punishing bloggers who do not wish to read their hate mail.

VI.

This case is neither the first nor last iteration of states’ eroding constitutional rights through legal uncertainty. The facts here evoke those of

Whole Woman's Health v. Jackson, 595 U.S. 30 (2021). There, Texas enacted the "Texas Heartbeat Act," also known as "S. B. 8," which created state law causes of action to sue for "statutory damages awards against those who perform or assist prohibited abortions." *Id.* at 36. The Texas statute predated *Dobbs v. Jackson Women's Health Org.*, 597 U.S. 215 (2022); so the "prohibited abortions" were therefore constitutionally protected. Accordingly, the statute's enforcement mechanism depended on private citizens bringing meritless lawsuits under the then-governing law.

The Court did not consider the constitutionality of S. B. 8 on its merits. *Id.* at 38. But several Justices observed that, even though S. B. 8 was "[u]nenforceable," the "threat of its punitive measures create[d] a chilling effect." *Id.* at 65 (Sotomayor, J., concurring in the judgment in part and dissenting in part). The law thus succeeded in "substantially suspend[ing]" (what was then) "a constitutional guarantee." *Id.* at 62.

Even if, years from now, every platform subject to enforcement proceedings for inaccurate policy disclosures is vindicated on the merits, the uncertainty of that outcome under the Hateful Conduct Law will have chilled speech. "It is not merely the sporadic abuse of power by the censor but the pervasive threat inherent in its very existence that constitutes the danger to freedom of discussion." *Thornhill v. Alabama*, 310 U.S. 88, 97 (1940). The uncertainty today is the point. It is the power to suppress that every government craves.

The biggest players in the marketplace of ideas may be able to weather this chill. But the "persons of modest means or limited mobility" who start online communities as "an unusually cheap and convenient form of communication," *City of Ladue v. Gilleo*, 512 U.S. 43, 57 (1994), are naked to their enemies. The costs of compliance (if compliance is even possible) and insurance (if it is even available), and fines of a thousand dollars per day for any misstep, will be ruinous.

VII.

The Hateful Conduct Law is unconstitutional under any plausible interpretation. I would not ask the New York Court of Appeals which unconstitutional interpretation is correct.

Besides, I already know the answer: that court has foreclosed saving constructions like the state's here. "[C]onstruing the statute as limited to certain constitutionally proscribable speech would likely result in transforming an otherwise overbroad statute into an impermissibly vague one; the statutory language would signify one thing but, as a matter of judicial decision, would stand for something entirely different." *Dietze*, 549 N.E.2d at 1169. "Under those circumstances, persons of ordinary intelligence reading" the Hateful Conduct Law "could not know what it actually meant." *See id.*

Moreover, I doubt that the Court of Appeals will accept the certified questions' unstated premise: that the supposed saving construction is viable under *New York's* laws and constitution. At the risk of being obvious, a "guarantee of free speech" is secured by the state constitution. *Holmes v. Winter*, 3 N.E.3d 694, 698 (N.Y. 2013); *see* N.Y. Const., art. I, § 8. "The drafters" of the state guarantee "chose not to model [it] after the First Amendment, deciding instead to adopt *more expansive* language: 'Every citizen may freely speak, write and publish his or her sentiments on all subjects . . . and no law shall be passed to restrain or abridge the liberty of speech or of the press.'" *Holmes*, 3 N.E.3d at 698 (emphasis added) (quoting N.Y. Const., art. I, § 8). It cannot be that the New York Court of Appeals will construe a state statute so that it offends the state constitution; I for one would not ask.

I would affirm Judge Carter's preliminary injunction.