

AN UNSETTLING AGREEMENT: META, JAWBONING, AND THE FIRST AMENDMENT

Evelyn Douek[*]

The settlement agreement that Meta reached with state Attorneys General in August 2026 is constitutionally unsound. The changes that the AGs pressured Meta to make to its platforms as part of the settlement burden the rights of third parties—Meta's users—in various ways (some content-neutral, some content-based, and some even viewpoint-based), and yet these terms have so far escaped First Amendment scrutiny.

Because these changes are being done under government pressure, the agreement violates the rule against jawboning; Meta's users should therefore be able to bring a First Amendment challenge seeking to enjoin government actors from enforcing the agreement's speech-restrictive terms, or seeking to have a court declare that those terms are unenforceable. Settlement agreements should not be allowed to become a loophole in the First Amendment's protections.

[*] Associate Professor, Stanford Law School. Many thanks to Alex Abdo, Easha Anand, Natasha Hjerrild, Daphne Keller, Genevieve Lakier, and Eugene Volokh for helpful comments and suggestions. Thanks also to Nina Raj and Ursula Neuner for excellent editorial assistance.

Introduction 2
I. The Settlement Agreement and Its Burdens on User Speech 5
A. Age Assurance 8
B. Overnight and Daily Usage Limits 15
C. Age-Appropriate Content Provisions 19
II. The Settlement Agreement Is Jawboning 25
A. The Capacious Rule Against Jawboning 29
B. The Standard of Review 37
C. Government Will Continue to Function 40
III. The Thwarted Ambitions of Platform Regulation 42
Conclusion 44

Introduction

According to the settlement agreement that Meta reached with 52 Attorneys General in August 2026, it is the parties' belief that "the public interest is best served when platforms and regulators work as partners."[1] This may well be true in some, or perhaps even many, cases. But the settlement creates a system of pervasive and invasive speech regulation, and in that context such cooperation might be said to be "startling and dangerous."[2]

First Amendment doctrine stands for the idea that we should generally be suspicious of government power over speech,[3] and specifically ensures that government actors cannot pressure, or partner with, private actors to "do indirectly what [they are] barred from doing directly."[4] That is: the government cannot evade constitutional constraints by enlisting private actors to impose its favored speech rules on its behalf. A settlement agreement is not a First Amendment-free zone, and the

---

[1] Settlement Agreement, Exh. A to Meta and State Attorneys General Consent Judgment at § IX, *In re* Social Media Adolescent Addiction/Personal Injury Products Liability Litigation, No. 4:23-cv-05448-YGR, MDL No. 3047 (N.D. Cal. Aug. 26, 2026), ECF No. 572-1.

[2] *Cf.* United States v. Stevens, 559 U.S. 460, 470 (2010).

[3] *See, e.g.*, Citizens United v. Fed. Election Comm'n, 558 U.S. 310, 340 (2010) (describing the First Amendment as "[p]remised on mistrust of governmental power").

[4] Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175, 190 (2024).

terms of Meta's deal with the Attorneys General at the very least need to be subjected to First Amendment scrutiny. This Article argues that because the Settlement Agreement attempts to evade such scrutiny, it is constitutionally unsound.

To make the stakes of this argument clear, imagine an alternate settlement agreement in which Meta, instead of promising changes like age assurance measures and time limits on kids' usage of its platforms, agreed to prohibit all users from seeing political content or (to make the point even more acute) all pro-Democrat content. The dangers of allowing such an agreement to escape constitutional scrutiny seem obvious. But because concern about kids' use of social media is bipartisan—and no doubt in part because kids themselves are often excluded from conversations about their rights—the constitutional infirmities of the Settlement Agreement have generally flown under the radar.[5] But kids have speech rights too,[6] as do the adult users of Meta's platforms who will also be affected by this Agreement, and politicians and platforms cannot constitutionally agree to forgo those rights on others' behalf.

I will argue that the third parties whose speech rights are burdened by the Settlement Agreement can bring a First Amendment challenge seeking to enjoin government actors from enforcing the Agreement's speech-restrictive terms, or seeking to have a court declare that those terms are unenforceable. This is true for settlement agreements that create such burdens on third-party rights in general, because any other conclusion would allow settlement agreements to become a loophole officials could use to regulate speech without facing constitutional scrutiny. And, as applied to this Settlement Agreement in particular, a large number of its provisions may well fail to survive constitutional scrutiny. Many of the content-neutral requirements in the Agreement mirror mandates that states have attempted

---

[5] For some notable exceptions, *see* Jeff Kosseff, *The Meta Settlement: Jawboning by Another Name?*, 8 J. FREE SPEECH L. 57 (2026); Jeff Kosseff, *The First Amendment Problems with Meta's $17 Billion Deal*, LAWFARE (Aug. 31, 2026), https://perma.cc/F5AX-NH7S; Mike Masnick, *Meta Just Paid Nearly $17 Billion to Make Sure It Gets to Write the Kid Safety Rules for Every Other Social Media Platform*, TECHDIRT (Aug. 26, 2026), https://perma.cc/455H-2728.

[6] Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 794 (2011) ("No doubt a State possesses legitimate power to protect children from harm, but that does not include a free-floating power to restrict the ideas to which children may be exposed.") (citation omitted); Erznoznik v. City of Jacksonville, 422 U.S. 205, 212–13 (1975) ("[M]inors are entitled to a significant measure of First Amendment protection, and only in relatively narrow and well-defined circumstances may government bar public dissemination of protected materials to them.") (citation omitted).

to impose through legislation in recent years, and while these laws are still working their way through the courts and the case outcomes are mixed, the majority of decisions have enjoined similar requirements as violating the First Amendment.[7] There are also content- and viewpoint-based requirements in the Agreement that are aimed at influencing what kids see online that would almost certainly fail First Amendment scrutiny if enacted as law.[8]

Considering the Settlement Agreement against this backdrop throws into sharp relief exactly *why* the First Amendment has rules to ensure that the government cannot avoid constitutional scrutiny by regulating speech through means other than legislation. This fact may be deeply frustrating to some. We are now around a decade into the "techlash" against social media platforms.[9] Despite all these years of concern and debate, the law and politics of platform regulation have made legislative solutions elusive. Attempts to reimagine First Amendment doctrine to give states more leeway to regulate the internet have failed.[10]

There might be a temptation, therefore, to think of the Agreement as an overdue reckoning that cleverly overcomes these obstacles and finally brings Big Tech's worst offender to heel. Indeed, it may only be the start of the reckoning: A month after Meta agreed to settle, TikTok entered into a settlement with Alabama that is clearly modeled on the Meta Settlement Agreement.[11] States are bringing lawsuits against artificial intelligence companies using the same theories as in the Meta case that led to the Settlement.[12] States may think they have found a new way of achieving their regulatory objectives that does not require passing any laws at all.

---

[7] *See* discussion of these cases *infra* Part I.

[8] *See Ent. Merchs. Ass'n*, 564 U.S. at 794 (rejecting the State's attempt to "create a wholly new category of content-based regulation that is permissible only for speech directed at children"). *See* further discussion *infra* Part I.C.

[9] Rana Foroohar, *Year in a Word: Techlash*, FIN. TIMES (Dec. 16, 2018), https://www.ft.com/content/76578fba-fca1-11e8-ac00-57a2a826423e?syn-25a6b1a6=1.

[10] Evelyn Douek & Genevieve Lakier, *Lochner.Com?*, 138 HARV. L. REV. 100, 126–29 (2024) (describing these efforts).

[11] Cecilia Kang, *TikTok to Pay Alabama $100 Million to Settle Social Media Addiction Claims*, N.Y. TIMES (Sept. 25, 2026) ("The obligations largely echo platform changes that Meta agreed to last month.").

[12] Cassandre Coyer, *States Bring Lawsuits Against OpenAI Over Kids Online Safety (1)*, BLOOMBERG LAW (Oct. 14, 2026), https://perma.cc/TD6U-5EAP.

But the First Amendment does not permit the government to evade review by simply changing the instrument of regulation. Otherwise, its protections would be paper-thin. And as frustrating as the government may find the commands of the First Amendment, they are a feature and not a bug of the American constitutional system. It is imperative that courts make clear that settlement agreements are not one neat trick to evade these constraints, lest states conclude they can abandon the hard and important work of building political will and legislative capacity for more comprehensive legal solutions, and instead regulate through one constitutionally infirm settlement at a time.

This Article proceeds as follows. Part I shows how the Settlement Agreement creates burdens on the speech rights of third parties: namely, Meta's users. Part II, the crux of the Article, argues that given these burdens, the Settlement Agreement violates the broad rule against what is known as jawboning—governmental pressure on private actors to restrict the speech of others. Part III takes a step back from the doctrinal landscape and considers what the Settlement Agreement suggests about the reduced ambitions and disappointments of debates about platform regulation.

How to effectively regulate new and fast-changing technologies continues to be one of the most pressing challenges of our time. Our regulatory solutions must respect the rights of society at large and not merely reflect the interests of the tech companies and the politicians that regulate them. The First Amendment is one essential safeguard in ensuring this is the case.

I. The Settlement Agreement and Its Burdens on User Speech

In exchange for settling a variety of claims against it, including allegations of deceptive and unlawful conduct in violation of state law as well as violations of federal privacy laws related to children's data, Meta agreed to pay the states at least $12.1 billion over ten years—and up to $17.1 billion if its major competitors adopt comparable obligations.[13] The Attorneys General billed the Settlement Agreement

[13] Settlement Agreement, *supra* note 1, at § VI.

as a "monumental victory,"[14] "a transformative framework"[15] and "groundbreaking."[16] Some outside observers agreed, calling it "the largest accountability moment in social media history,"[17] "a watershed moment,"[18] and a "historic step."[19] These dramatic claims were not made simply because of the money involved, however. Twelve billion dollars is obviously nothing to sniff at, but as commentators were quick to point out, this large sum looks less impressive when put in context—it represents roughly a single quarter of profits for the $1.45 trillion company.[20] Florida's Attorney General, who sat out of the settlement, called the sum "a rounding error."[21]

Far more significant, then, are the various changes that Meta agreed to make to its platforms under the Settlement Agreement. Indeed, many Attorneys General focused on these changes as the true goal of the entire case'. As Kentucky Attorney General Russell Coleman put it, "this was never just about the money."[22] Tennessee Attorney General Jonathan Skrmetti said, "[w]e brought these suits because . . . we

---

[14] Press Release, Phil Weiser, Colorado Attorney General, Attorney General Weiser Announces Historic Settlement with Meta Platforms (Aug. 26, 2026), https://perma.cc/6YNG-HHUC.

[15] Press Release, Letitia James, New York Attorney General, Attorney General James Secures Up to $17.1 Billion and Groundbreaking Reforms from Meta to Protect Children on Social Media (Aug. 26, 2026), https://perma.cc/8SVB-PD5U.

[16] Press Release, Jonathan Skrmetti, Tennessee Attorney General & Reporter, Attorney General Skrmetti Announces Largest Big Tech Settlement in History (Aug. 26, 2026), https://www.tn.gov/attorneygeneral/news/2026/8/26/pr26-33.html.

[17] Jonathan Haidt, LINKEDIN (Aug. 26, 2026), https://perma.cc/AC6K-PUA3.

[18] Rachael Myrow, *Meta Agrees to $17 Billion Settlement and New Protections for Teens*, NPR (Aug. 26, 2026), https://www.npr.org/2026/08/26/nx-s1-5945010/meta-agrees-to-17-billion-settlement-and-new-protections-for-teens (quoting Jim Steyer of Common Sense Media).

[19] Justin Hendrix, *Assessing the Landmark Meta Settlement*, TECH POLICY PRESS (Aug. 27, 2026), https://perma.cc/T4WC-A234 (quoting Meetali Jain of Tech Justice Law).

[20] *See, e.g.*, Emmy Martin, *What Does $17.1 Billion Mean for Meta?*, N.Y. TIMES (Aug. 26, 2026).

[21] Alex Ebert, *Florida Top Cop Pushes Meta to Trial, Rips 'Rounding Error' Deal*, BLOOMBERG LAW (Aug. 27, 2026), https://perma.cc/JJ9X-EXVQ (quoting Florida Attorney General James Uthmeier).

[22] McKenna Horsley, *Meta Settles with Kentucky, Other States, Including Changes for Kids' Safety*, KENTUCKY LANTERN (Aug. 26, 2026), https://perma.cc/3P94-WHHJ (quoting Kentucky Attorney General Russell Coleman).

wanted to reform the platforms."[23] California Attorney General Rob Bonta said that it was Meta's agreement to make these changes that "addresse[d] the concerns at the core of our lawsuit."[24] Colorado Attorney General Phil Weiser also called them "[t]he focus of this case."[25] The changes Meta agreed to were numerous; those most relevant for my argument here about third-party rights included:

- implementing age assurance measures;[26]
- limiting minor users' access to Meta's platforms late at night;[27]
- imposing overall daily maximum usage limits for minor users;[28]
- certain assurances about the kind of content that minor users will see on the platforms.[29]

In celebrating these reforms, several Attorneys General specifically emphasized that they went beyond the changes that courts had ordered in the past, or likely would order.[30] The Attorneys General were right to be concerned that a court would not order similar changes. Many of the terms of the Settlement Agreement reflect changes that states have tried to impose through legislation in the past few years, and which courts have held are unconstitutional.

---

[23] Cecilia Kang & Eli Tan, *How Meta's $17.1 Billion Social Media Settlement Came Together*, N.Y. TIMES (Aug. 29, 2026) (quoting Tennessee Attorney General Jonathan Skrmetti).

[24] Press Release, Rob Bonta, California Attorney General, Attorney General Bonta Secures Transformative $17 Billion Settlement with Meta, Proposed Settlement Includes Fundamental Changes to Instagram and Facebook (Aug. 26, 2026), https://perma.cc/76FM-39PL.

[25] Colorado Attorney General, *supra* note 14.

[26] Settlement Agreement, *supra* note 1, at § II.A, discussed *infra* Part I.A.

[27] *Id.* at § II.B.2, discussed *infra* Part I.B.

[28] *Id.* at § II.B.3, discussed *infra* Part I.B.

[29] *Id.* at § II.E, discussed *infra* Part I.C.

[30] *See, e.g.*, Colorado Attorney General, *supra* note 14 (quoting the Colorado Attorney General saying the settlement was "well beyond what any court has ordered or is likely to order"); Max Zahn et al., *Meta Settles with States in Landmark Social Media Addiction Case*, ABC NEWS (Aug. 26, 2026), https://perma.cc/AL2R-4ZLP (quoting the New Jersey Attorney General saying the settlement is "not something that any court has ever ordered"); Tennessee Attorney General & Reporter, *supra* note 16 (describing the changes Meta agreed to in the settlement as "more significant and comprehensive than any previously ordered by a court").

This Part surveys some of the terms of the Agreement that affect the speech rights of Meta's users, as well as the recent case law about similar requirements imposed legislatively. This is not intended to be a comprehensive analysis of the First Amendment issues raised by the Agreement; rather, my aim in this Part is simply to show that the Settlement Agreement implicates third-party speech rights that are plausibly and even likely protected by the First Amendment. I focus on three of the most burdensome terms for third parties: Meta's commitments to adopt age assurance measures, to limit the times of day and the aggregate amount minors can use their platforms, and to restrict the kinds of content minors will see.

A. Age Assurance

Under the Settlement Agreement, Meta has committed to adopting an "age assurance framework"[31] to determine if a user accessing their platforms from any of the jurisdictions party to the Agreement is either under 13 or between the ages of 13 and 17. The Settlement Agreement then specifies how Meta must treat these users, and I discuss some of the specific requirements further below. In essence, once Meta determines that a user is under 18 (or under 13), Meta must restrict their access to speech on the platform in a variety of ways. If Meta cannot or does not perform age assurance on a particular account, then that account is defaulted into a more restrictive experience.[32]

The Agreement specifies the maximum "U18 False Positive Rate"—the share of 13-to-17-year-olds wrongly classified as adults—that the tools can have in order to comply with the Agreement,[33] but does not specify a particular maximum false

---

[31] A note on terminology: As Eric Goldman explains, the terminology in this area is confusing and not standardized. But generally speaking, "age assurance" involves determining that a user is at least a certain age, whereas "age verification" involves determining a user's precise age. *See* Eric Goldman, *The "Segregate-and-Suppress" Approach to Regulating Child Safety Online*, 28 STAN. TECH. L. REV. 173, 179 (2025).

[32] If Meta cannot or does not perform age assurance on a particular account, then after fourteen days it must treat that account as a Teen User—unless the user claims to be an adult, in which case Meta must instead restrict the account's contact with teens and base its recommendations on "Age Appropriate Experiences." Settlement Agreement, *supra* note 1, at §§ II.A.10(b), II.A.10(a)(ii).

[33] This rate varies over the life of the agreement and depending on whether Meta uses a commercially available or proprietary tool but is between 10–14% for minors aged 16–17 and 3–7% for minors aged 13–15. *Id.* at § II.A.6.

negative rate.[34] That is, Meta can violate the Agreement if it lets too many minors use its platforms without restriction, but not if it restricts the use of too many adults because it accidentally determines they are too young. The incentive this creates is obvious: Meta is likely to err on the side of caution and set the confidence thresholds on its age assurance tools to over-classify people as minor users to avoid liability. While Meta might generally benefit from having more adult users on its platforms to serve ads to, the marginal benefit of the accounts wrongly classified as minors in this way is unlikely to outweigh the potential liability for being found to be in breach of the Settlement Agreement. These features of the age assurance framework as set out in the Agreement—the default to a more restrictive experience when a user's age cannot be assessed, and the legal incentives to err on the side of over-classifying users as under 18—are constitutionally significant given the way the Court has thought about age-gating online.

The Supreme Court has taken surprisingly few cases about how the First Amendment applies to internet regulation.[35] But because age-gating measures were an early and common regulatory response to concerns about the impacts of the internet on young people, a large proportion of the cases that the Court *has* taken have involved various measures that encouraged the use of age verification tools. For nearly three decades, prevailing precedent was clear that such requirements were presumptively unconstitutional.[36] In two early internet cases, the Court made clear that requiring age verification would burden not only the speech rights of minors but also those of adults, either because it discouraged adults from accessing

---

[34] Although it does require a "Clear and Conspicuous" appeal for users misidentified as minors, with timely decisions (*id.* at § II.A.9), and also requires Meta to take "reasonable steps" to lower both false positive and false negative rates (*id.* at § II.A.11(a)).

[35] Douek & Lakier, *supra* note 10, at 129 (noting that for years the Court "assiduously avoid[ed] the question of how the First Amendment applied to internet regulation").

[36] *See* Meg Leta Jones, *Age Verification Beyond Porn: What Paxton v. Free Speech Coalition Could Mean for the Future of the Internet* 7 (Apr. 3, 2026) (unpublished article) (noting that until recently, "the takeaway [of the Court's cases] seemed clear that online age verification requirements would face strict scrutiny, and they would likely fail"), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6516238; Free Speech Coal., Inc. v. Paxton, 606 U.S. 461, 500 (2025) (Kagan, J., dissenting) ("Cases raising that question have reached this Court on no fewer than four prior occasions—and we have given the same answer, consistent with general free speech principles, each and every time. Under those principles, we apply strict scrutiny, a highly rigorous but not fatal form of constitutional review . . . .").

sites that required it or because the unreliability of age verification would mean adult users would be mistakenly prevented from accessing content they had a constitutional right to see.[37]

This prevailing doctrinal framework was altered slightly in 2025 in *Free Speech Coalition v. Paxton*.[38] Texas had imposed an age verification mandate for commercial websites that published sexually explicit content.[39] *Free Speech Coalition* affirmed the lesson from the Court's previous cases that age verification requirements can burden the speech rights of adult visitors to the websites that require it.[40] But the Court now held that because the traditional powers of the state include the power to protect minors from speech that is obscene as to them, when the state regulates websites that disseminate this kind of speech, the burden on adults is merely "incidental" to the regulation of unprotected speech.[41] The Court therefore applied intermediate scrutiny, rather than strict scrutiny, and found that the law survived review.[42]

But the Court's reasoning was explicitly grounded in the unique First Amendment status of speech that is obscene as to minors, which the Court called "partially protect[ed]" speech because it is protected for some (adults) but unprotected for others (minors).[43] The key move in the opinion—that age verification measures count as incidental rather than direct burdens on adult speech—therefore depended on this distinctive constitutional feature of obscenity, which the Court has never found to be true of other categories of speech. Certainly, the vast majority of

---

[37] Reno v. ACLU, 521 U.S. 844, 874 (1997) (holding that age verification "effectively suppresses a large amount of speech that adults have a constitutional right to receive and to address to one another"); *id.* at 856 (acknowledging that age verification requirements burdened adult speech rights by "discourag[ing] users" from accessing sites that required such verification); Ashcroft v. ACLU (*Ashcroft II*), 542 U.S. 656, 665 (2004) (affirming *Reno*'s holding that age verification requirements burden the speech rights of adults).

[38] 606 U.S. 461 (2025).

[39] *Id.* at 466–68.

[40] *Id.* at 466 (noting that Texas's law "also burdens adult visitors of these websites, who all agree have a First Amendment right to access at least some of the content that the websites publish").

[41] *Id.* at 483.

[42] *Id.* at 496.

[43] *See, e.g.*, *id.* at 493 n.12.

speech on social media is not partially protected speech, but fully protected speech for both kids and adults.

Perhaps states thought *Free Speech Coalition* was a harbinger of further shifts in the Court's approach to online speech regulation in general and age verification in particular. Nothing in the opinion suggested it should be read this way, but laws imposing age verification or assurance mandates for social media platforms, and not just material obscene as to minors, have proliferated in recent years. As courts have considered these requirements, though, the majority have continued to find them unconstitutional.[44] In doing so, courts have noted not only the burden on minors' speech rights in having their access to social media restricted, but also the burden on *adults'* speech rights in being forced to comply with age verification.[45] Judges have correctly read the holding in *Free Speech Coalition* as a product of the

---

[44] Note, *Content Neutrality for Kids: Intermediate Scrutiny for Social Media Age-Verification Laws*, 139 HARV. L. REV. 942, 950–51 (2026) (noting that "the district courts that have addressed [recent age verification laws restricting minors' access to social media] have almost uniformly enjoined them for violating the First Amendment").

[45] NetChoice, LLC v. Reyes, 748 F. Supp. 3d 1105, 1129 n.169 (D. Utah 2024) ("[T]he Act's age assurance provision, which applies to all users, broadly burdens adult users' ability to 'access a broad range of protected speech on a broad range of covered websites.'" (quoting NetChoice, LLC v. Fitch, 738 F. Supp. 3d 753, 774 (S.D. Miss. 2024))); NetChoice v. Carr, 789 F. Supp. 3d 1200, 1224 (N.D. Ga. 2025) ("[I]n addition to curbing children's speech rights, the law's Age Verification Provision would also impose an invasive burden on *all* Georgians' online speech by requiring their submission of personal information to access any social media platforms."), *aff'd in part, vacated in part sub nom.* NetChoice v. Att'y Gen., Georgia, No. 25-12436, 2026 WL 3079563 (11th Cir. Oct. 15, 2026); *Netchoice v. Carr*, 789 F. Supp. 3d at 1225 (finding Georgia's Age Verification Provision "imposes a sweeping burden on adults' access to speech by requiring 'commercially reasonable' age verification mechanisms for *all* users, creating a broad-reaching chilling effect"); NetChoice, LLC v. Griffin, No. 5:23-CV-5105, 2025 WL 978607, at *8 (W.D. Ark. Mar. 31, 2025) (holding that "the age-verification requirement will deter adults from speaking or receiving protected speech on social media" and citing *ACLU v. Gonzales*, 478 F. Supp. 2d 775, 806 (E.D. Pa. 2007), and *PSINET, Inc. v. Chapman*, 167 F. Supp. 2d 878, 889 (W.D. Va. 2001), *aff'd*, 362 F.3d 227 (4th Cir. 2004)); NetChoice v. Murrill, 812 F. Supp. 3d 594, 649 (M.D. La. 2025) ("[T]he age-verification and parental-consent requirements . . . trigger First Amendment scrutiny because of their indirect but substantial burdens to (accessing and engaging in) speech . . . ."); NetChoice v. Jones, 822 F. Supp. 3d 656, 676 (E.D. Va. 2026) (holding "age verification burdens adults'—as well as minors'—rights to access protected speech"); NetChoice v. Hilgers, 840 F. Supp. 3d 777, 790 (D. Neb. 2026) (noting the "incidental burden on adults who must verify their age before making an account").

uniqueness of obscenity as a First Amendment category, and thus of limited relevance in other contexts.[46]

There are some exceptions. The Eleventh Circuit recently read *Free Speech Coalition*'s reasoning with respect to obscenity as not "establish[ing] a constitutional outer limit" for age verification requirements; "[t]o the contrary," the Eleventh Circuit reasoned, "everything the Court said in *Free Speech Coalition* seems to point in the other direction," including the Court's comments that age verification is common when laws draw lines based on age, including for alcohol, tobacco, or online gambling (three things, it should be noted, that no one has a constitutional right to access, which is obviously not the case with respect to protected speech).[47] But this is an outlier opinion, which has not been followed by other circuits. Moreover, even the Eleventh Circuit did not, in this opinion, dispute that age-gating measures impose *some* burden on speech and went on to apply intermediate scrutiny to the Florida law at issue in the case.[48] The Eleventh Circuit held that the burdens in that particular case required only intermediate scrutiny because the definition of the platforms to which the law applied was content-neutral.[49] Whether social media bans can be characterized as content-neutral continues to be contested in lower courts, and the issue deserves more space than I can give it here. For now,

---

[46] *Netchoice v. Murrill*, 812 F. Supp. 3d at 651 (finding Defendants could not "persuasively compare the instant case to *Free Speech Coalition*" because "the Court recognized repeatedly that pornography is unprotected speech for minors"); *Netchoice v. Hilgers*, 840 F. Supp. 3d at 792 (citing *Free Speech Coalition* to note the law's ability to "prevent children from accessing unprotected speech"); Comput. & Commc'ns v. Uthmeier, No. 25-11881, 2025 WL 3458571, at *18 (11th Cir. Nov. 25, 2025) (Rosenbaum, J., dissenting) ("Restricting a person's ability to access obscene, pornographic material is quite different from burdening a person's ability to engage in any speech, of any kind, at all, on a social media platform.").

[47] Computer & Commc'ns Indus. Ass'n v. Uthmeier, 194 F.4th 144, 167–68 (11th Cir. 2026). The Eleventh Circuit had similarly read *Free Speech Coalition* more broadly at the preliminary injunction stage. *See* Computer & Commc'ns Indus. Ass'n v. Uthmeier, No. 25-11881, 2025 WL 3458571, at *8 n.11 (11th Cir. Nov. 25, 2025) (acknowledging that "age verification requirement imposes some burden on adult speech (thus implicating First Amendment scrutiny in the first place)" but reading *Free Speech Coalition* as having "recently clarified" that age verification does not always trigger strict scrutiny because it constituted a "'more modest burden' on adult speech" (quoting Free Speech Coal., Inc. v. Paxton, 606 U.S. 461, 497 (2025))).

[48] *Computer & Commc'ns Indus. Ass'n*, 194 F.4th at 160 ("We have little trouble concluding that HB3's restrictions implicate the First Amendment . . . .").

[49] *Id.* at 162–64.

I will just note that the matter is disputed,[50] the Eleventh Circuit did not engage with any of the Court's precedents that protect the right to engage in speech anonymously, and in any event the Meta Settlement Agreement includes explicitly content-based terms.

Therefore, were a legislature to pass a statute that required Meta to adopt the age verification requirements written into the Agreement, that statute would likely be held unconstitutional under the prevailing authority in most circuits, and at the very least would require the application of constitutional scrutiny in all circuits. Further supporting this conclusion—and undermining the approach taken by the recent Eleventh Circuit case—is the fact that at least one member of the majority in *Free Speech Coalition* has since reaffirmed that that case's reasoning should be understood as limited to the context of obscenity. In a concurrence on the shadow docket in a case that remains ongoing in lower courts, Justice Kavanaugh noted that a Mississippi law requiring age verification and banning minors from using social media without parental consent was "likely unconstitutional."[51] In stating that it was "no surprise" that social media age verification schemes around the country had been enjoined on First Amendment grounds,[52] he favorably cited many of the lower court decisions that had found as much, and cited *Free Speech Coalition* with a "cf.," suggesting he thought it was materially different.

To be sure, many of the laws in these recent cases are not direct analogs to the age verification provisions in the Settlement Agreement. The burdens those laws seek to impose on minors are in some ways greater than under the Settlement Agreement—indeed, many states have sought to ban minors' access to social media entirely. Under the Settlement Agreement, by contrast, users under 13 are prevented from having accounts, but Teen Users determined to be between 13 and 17

---

[50] *Compare, e.g.*, *Netchoice v. Reyes*, 748 F. Supp. 3d at 1123 (finding a social media age restriction requirement is likely "content based" because it "facially distinguishes between 'social' speech and other forms of speech") *with* Comput. & Commc'ns Indus. Ass'n v. Uthmeier, No. 25-11881, 2025 WL 3458571, at *10 (11th Cir. Nov. 25, 2025) (finding an age-restriction on platforms that use addictive features to be content neutral because "[i]t regulates social-media platforms based on their functional capabilities and features, not the content they communicate").

[51] NetChoice, LLC v. Fitch, 145 S. Ct. 2658, 2658 (2025) (Kavanaugh, J., concurring in the denial of application to vacate stay).

[52] *Id.*

are defaulted into a more restrictive platform experience, not completely prevented from using the platform.

This may make a difference to the constitutional bottom line. In one case, for example, the Ninth Circuit rejected a facial challenge to a California law that required age estimation but did not necessarily bar minors from accessing content,[53] and remanded the case to the district court to further develop the factual record. Crucial to the Ninth Circuit's reasoning, though, was that it read the statute to require only that a platform adopt greater privacy and data protections for users the platform estimated to be minors, and would not in any way affect the content they could see.[54] The Settlement Agreement is different from the California law because, as I will shortly discuss, it does impose restrictions on access to content for those who are found to be minors. It may not impose a full ban, but some of the restrictions are severe. This suggests that even its content-neutral provisions may be unconstitutional. Furthermore, importantly, the Agreement also includes content-based conditions that generally require even greater scrutiny under the First Amendment.

In summary, courts have made clear that at least *some* First Amendment scrutiny is warranted for age verification and assurance schemes, and nothing in *Free Speech Coalition* says otherwise. That is enough to establish that the Settlement Agreement implicates third-party speech rights and for that reason is constitutionally perilous. This is true even though lower courts are continuing to wrestle with the appropriate level of scrutiny for age verification requirements, and whether these laws can survive whatever scrutiny applies.

The landscape is shifting, and the almost unanimous trend of district courts striking down social media age-gating laws has become more mixed at the appeals

---

[53] NetChoice, LLC v. Bonta, 170 F.4th 744, 761–62 (9th Cir. 2026) (noting that "it is unclear that the age estimation requirement prevents access to content" and that "businesses that do not wish to conduct age estimation may publish any content they would like, as long as they default to data and privacy protections for all users. We are therefore not convinced that the age estimation provision necessarily burdens protected expression on its face by requiring covered businesses to determine 'what content is appropriate' for children").

[54] *Id.* at 761–63.

level.[55] This no doubt affected Meta's analysis of its litigation risk in a way that made it more likely to settle.

But Meta's risk tolerance is not a reliable measure of constitutionality. Under current Supreme Court authority—and as understood by the majority of lower courts at the time of the signing of the Settlement Agreement—laws that impose age-gating measures to access protected speech remain presumptively unconstitutional. Certainly, the government would have a heavy burden to prove otherwise, given the mountain of case law against it. The extent of some of the restrictions on users who are age-gated out of the regular platform under the Settlement Agreement would make this argument especially difficult. The remainder of this Part examines the two most burdensome restrictions: usage limits and content moderation commitments.

B. Overnight and Daily Usage Limits

Once Meta determines an account belongs to someone aged 13–17 (a "Teen User"), the Settlement Agreement places severe restrictions on that person's use of Meta's platforms by default. Teen Users can't access the platforms at all overnight (although direct messaging is excluded),[56] and have an overall daily usage limit—

---

[55] It is important not to over-generalize the relevance of these appellate holdings for our analysis of the Settlement Agreement, however. The specifics of each law matter greatly and differ from the Settlement Agreement in material ways. For example, the Sixth Circuit upheld a social media law, but the law did not require age verification. NetChoice, LLC v. Yost, 180 F.4th 268, 302 (2026) (Batchelder, J., concurring) ("The statute is silent on age verification, but at oral argument on summary judgment, counsel for Yost took the position that '[t]his is a self-reporting statute.'"). The Fifth Circuit stayed a preliminary injunction of Mississippi's social media law but did not explain its reasoning. NetChoice, L.L.C. v. Fitch, No. 25-60348, 2025 WL 2078435, at *1 (5th Cir. July 17, 2025). Other circuits have taken issue with whether challengers have satisfactorily discharged their burden to bring a facial challenge. NetChoice, LLC v. Bonta (*Bonta II*), 170 F.4th 744, 750 (9th Cir. 2026); NetChoice v. Att'y Gen., Georgia, No. 25-12436, 2026 WL 3079563, at *8 (11th Cir. Oct. 15, 2026). Thus, each of these decisions has limited relevance for the constitutional analysis of the Settlement Agreement.

[56] Initially defined as midnight–6 a.m., but with the potential to expand to 10 p.m.–7 a.m. in the event that other industry members become subject to similar obligations as Meta. Settlement Agreement, *supra* note 1, at § II.B.2.

two hours initially.[57] Only a linked "Supervising Parent" can relax these restrictions,[58] which seem intended to address concerns about addictive use of the platform, including the impact on teen sleep if they use social media overnight.[59]

These requirements clearly burden speech. Limiting users to accessing a speech platform for only two hours a day, and not at all overnight, restricts their ability to engage in expressive activity and to receive the speech of others. This is enough to establish that the Settlement Agreement would be subject to First Amendment scrutiny if it were a law.

And it is not clear that these requirements could survive that scrutiny once applied. Even assuming these restrictions are content-neutral (which, as I note shortly, is not at all obvious), they would be subject to at least intermediate scrutiny.[60] Under intermediate scrutiny, a speech restriction is constitutional if it "advances important governmental interests unrelated to the suppression of free speech and does not burden substantially more speech than necessary to further

---

[57] *Id.* at § II.B.3. This too excludes messaging and video or audio content of at least 22 minutes. It also has the potential for the limit to fall to one hour per platform, still capped at two hours overall, if Snap, TikTok, and YouTube become subject to similar obligations.

[58] *Id.* at §§ II.B.2(a)(v), II.B.3(a)(iv), II.B.4(b), II.B.5(a). The Court in *Brown* rejected the state's argument in that case that the fact that parents could still give permission for their children to access content the state has placed off limits justifies or lessens the First Amendment violation of restricting minors' access to that content. Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 804 (2011) ("While some of the legislation's effect may indeed be in support of what some parents of the restricted children actually want, its entire effect is only in support of what the State thinks parents *ought* to want. This is not the narrow tailoring to 'assisting parents' that restriction of First Amendment rights requires.").

[59] Complaint for Injunctive and Other Relief at 83, *In re* Social Media Adolescent Addiction/Personal Injury Products Liability Litigation, No. 4:23-cv-05448, MDL No. 3047 (N.D. Cal. Aug. 26, 2026) ("Young users who use social media for more than five hours per day are three times more likely than non-users to not sleep enough, contributing to associated physical and mental health impacts."). *See also* U.S. Surgeon General's Advisory, *Social Media and Youth Mental Health* 5 (2023), https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf ("The influence of social media on youth mental health is shaped by many complex factors, including, but not limited to, the amount of time children and adolescents spend on platforms, . . . and the degree to which it disrupts activities that are essential for health like sleep and physical activity.").

[60] Free Speech Coal., Inc. v. Paxton, 606 U.S. 461, 471 (2025) (citing Turner Broadcasting System, Inc. v. FCC, 512 U.S. 622, 642 (1994)).

those interests."[61] This is a much more forgiving standard than strict scrutiny, but intermediate scrutiny is not a free pass, and the burden remains on the government.

The recent case most directly on point found such a requirement unconstitutional. *NetChoice v. Jones* involved a challenge to a Virginia law that required platforms to implement age verification and limit the access of users under 16 to one hour per day.[62] The law defined social media platforms by reference to their content, and so the court applied strict scrutiny.[63] Importantly, though, the court also found that the law would likely fail even intermediate scrutiny, given the extent of the burden it would impose on both adults' and minors' rights to access protected speech.[64] As the court noted, under the law, minors would no longer be able to access online church services or history and science programming lasting over an hour without parental consent,[65] and adults also wouldn't be able to access this content without undergoing age verification.

Somewhat quixotically, the Settlement Agreement may permit the user to access the content that the court was concerned minors might not be able to access in *Jones* because it might fall within the exception for "longform content" over 22 minutes. But this raises other questions about whether the restrictions are truly content-neutral, given the fine distinctions they draw between types of online content. It also raises questions about whether the restrictions are tailored to addressing the harms they target, given that minors can still spend significant time every day across Meta's platforms provided they engage only in messaging and watching videos of 22 minutes or more.[66]

---

[61] *Id.*

[62] NetChoice v. Jones, 822 F. Supp. 3d 656, 664 (E.D. Va. 2026).

[63] *Id.* at 672.

[64] *Id.* at 676.

[65] *Id.* at 675.

[66] Settlement Agreement, *supra* note 1, at § I.EE ("'Longform Content' shall mean a piece of video or audio content with a duration of at least 22 minutes that Meta has determined with a high degree of reliability to not have been artificially extended."); *id.* at § II.B.3(a)(i) ("This daily limit shall not include time spent watching Longform Content, engaging in messaging, or accessing settings.").

Another court found that a much less restrictive provision—one that simply limited notifications overnight, rather than platform use entirely—also failed intermediate scrutiny.[67] The court held that because platforms would effectively have to silence overnight notifications for everyone who was not an age-verified adult account holder, the law was not narrowly tailored.[68]

These few cases hardly settle the matter, but they show that even content-neutral restrictions are not simply waved through by courts. Limiting access to a whole medium of expression to only one or two hours a day is a significant restriction on speech; the government's evidentiary burden to defend such a restriction is significant even under intermediate scrutiny.[69] Even if excessive social media use is responsible for a youth mental health crisis (a topic on which experts sharply disagree[70]), the government would have to show that restricting every minor's use every

---

[67] NetChoice v. Griffin, 831 F. Supp. 3d 816, 831 (W.D. Ark. 2026) (finding a restriction on overnight notifications "burdens platforms' speech by silencing them for a third of the day without any indication that the burden will reduce nighttime social media use or otherwise serve the State's asserted interest at all").

[68] *Id.* ("[P]latforms would therefore have to silence notifications between 10 p.m. and 6:00 a.m. for everyone in Arkansas unless they have become an age-verified adult account holder.").

[69] *See, e.g.*, Packingham v. North Carolina, 582 U.S. 98, 108 (2017) (finding that a law banning sex offenders from using social media failed even intermediate scrutiny because "to foreclose access to social media altogether is to prevent the user from engaging in the legitimate exercise of First Amendment rights").

[70] *See, e.g.*, Matti Vuorre & Andrew Przybylski, *Estimating the Association Between Facebook Adoption and Well-Being in 72 Countries*, ROYAL SOCIETY OPEN SCIENCE app. A (2023) ("These data did not indicate strong support for or against the idea that countries' Facebook adoption was associated with those countries' levels of mental health problems . . . ."); NAT'L ACAD. OF SCIENCES, ENGINEERING & MEDICINE, SOCIAL MEDIA AND ADOLESCENT HEALTH 2024, at xvi (Sandro Galea et al. eds., 2024) ("While some users, using social media in particular ways, may have their mental health adversely affected, for many others there will be no such harm, and for others still the experience will be helpful."); Candice Odgers, *The Great Rewiring: Is Social Media Really Behind an Epidemic of Teenage Mental Illness?*, 628 NATURE 29 (2024) ("When associations over time are found, they suggest not that social-media use predicts or causes depression, but that young people who already have mental-health problems use such platforms more often or in different ways from their healthy peers . . . . [A]ge-based restrictions and bans on mobile devices, are unlikely to be effective in practice—or worse, could backfire given what we know about adolescent behaviour."); Amy Orben & Andrew Przybylski, *The Association Between Adolescent Well-Being and Digital Technology* Use, 3 NATURE HUM. BEHAV. 173 (2019) ("The association we find between digital technology use and adolescent well-being is negative but small, explaining at most 0.4% of the variation in well-being.").

single day—including during an election season, during student government elections, or during any other protected activity that minors engage in, sacred or profane—actually helps alleviate that crisis. Even with respect to minors, this would be a hard burden to carry.

This is also where the infirmities of the age assurance scheme described in the previous section bite. It is a burden on adult speech to have to complete age assurance in order to access protected content in an unrestricted way, and Meta is incentivized to over-classify adult users into these restrictive use conditions. Social media platforms include much trivial speech, to be sure, but they also contain much important democratic discourse, and many people visit them to get their news, or engage in culture-making or community organizing.[71] Severely restricting people's access to such speech demands at least some form of First Amendment review, and courts have shown that this review is meaningful.

C. Age-Appropriate Content Provisions

The Settlement Agreement also contains explicitly content-based speech restrictions.[72] The most significant is Meta's representation "that it has already taken measures to promote age-appropriate content for Teen Users" and that it "agrees to maintain" these measures, or others "at least as effective."[73] It also commits to "continuously improve" these measures, and "take steps designed to reduce the frequency of [] exposure over time" as well as make changes "to address where [Meta's Community Standards] are not capturing content or experiences that are causing harm to Teen Users."[74]

---

[71] *See, e.g.*, *Packingham*, 582 U.S. at 107 (describing social media as "the principal sources for knowing current events, checking ads for employment, speaking and listening in the modern public square, and otherwise exploring the vast realms of human thought and knowledge. These websites can provide perhaps the most powerful mechanisms available to a private citizen to make his or her voice heard").

[72] Beyond those discussed in this Section, s*ee, e.g.*, restrictions on contact with "Age Inappropriate Accounts" (Settlement Agreement, *supra* note 1, at §§ I.E, II.E.1(f)); parental notification of repeated searches for suicide/self-harm/eating-disorder terms and of first contacts with adults (*id.* at § II.G.1–2); "'Age Appropriate Experiences"' default tied to a 13+ setting "inspired by movie ratings" (*id.* at §§ I.C, II.E.1(d)); and a ban on cosmetic-filters (*id.* at § II.D.2).

[73] *Id.* at § II.E.1–2.

[74] *Id.* at § II.E.2–3.

The careful wording of this provision suggests that the parties are all too aware of the First Amendment perils here. Phrasing the obligation as Meta maintaining measures it already had in place seems to be an attempt to insulate this provision from First Amendment challenge by minimizing the coercive role of the government counterparties. I discuss further below the reason I think the attempt is not successful.[75] In short, the Settlement Agreement transforms Meta's relevant Community Standards from voluntary rules into binding obligations—enforceable through contempt proceedings—and there's a very big practical and legal difference between the two. For now, it's sufficient to note that the drafting suggests the parties saw the First Amendment implications of the Settlement Agreement, even if they were wrong in assuming they only had to worry about these explicitly content-based provisions.

The measures Meta commits to maintain and improve include "implementing content policies designed to help limit Teen Users' exposure to Age Inappropriate Content." This means Meta is committing to continue trying to prevent certain users from hearing certain speech, and to reduce users' exposure to that speech over time.

That speech, "Age Inappropriate Content," is defined as content "that is generally perceived by U.S. parents, youth experts, and teens as not being appropriate for Teen Users."[76] This definition is extremely vague, if not incoherent. Anyone who has ever been a teen or a parent, or has known either, will be aware that these groups can have very different conceptions of what kind of content might be "appropriate" for teens.

Mitigating this incoherence somewhat, the definition goes on to list a number of Meta's existing Community Standards that should be taken to represent Age Inappropriate Content for the purposes of the Agreement. But the Agreement acknowledges that the Community Standards will be "amended from time to time"[77] and, as already noted, requires Meta to revise its policies if they are later found to not be "capturing content or experiences that are causing harm to Teen

---

[75] *See infra* Part II.A.

[76] Settlement Agreement, *supra* note 1, at § I.F.

[77] *Id.* at § I.S.

Users." Thus, tying the definition to existing Community Standards does not provide certainty as to the meaning of "Age Inappropriate Content" over the life of the Agreement.

Meta also agrees it has and will maintain "product interventions designed to limit or interrupt repeated exposure to Sensitive Aggregate Content" for Teen Users. This is "content that, in the aggregate, may be inappropriate for Teen Users"—another vague definition. The term is defined by a list of broad categories, which is worth reproducing in full to emphasize the scope of the kinds of speech that the Settlement Agreement requires Meta to reduce minors' access to:

> nutritional discussion, personal weight and physical fitness accomplishments, muscular physiques, promotion or depiction of non-invasive aesthetic procedures for maximizing physical attractiveness, depiction of certain body parts commonly associated with body idealization, and depiction or discussion of personal experience living with mental or emotional distress.[78]

Together, these provisions have several First Amendment problems in common.

First, as I have already suggested, these definitions are vague. This would be a constitutional problem if enacted as law because legal obligations must give fair notice of what is forbidden or required, and this rule is applied with extra rigor in the First Amendment context.[79] As the Court has explained, "[w]hen speech is involved, rigorous adherence to those requirements [of clear notice] is necessary to ensure that ambiguity does not chill protected speech."[80] Thus, the lack of clarity about what content might be "generally perceived by U.S. parents, youth experts, and teens as not being appropriate" or "in the aggregate, may be inappropriate" for teens raises serious First Amendment questions.

Importantly, this is not just because it makes it difficult for Meta to understand its obligations under the Agreement. As I discuss further below, I am not interested in defending Meta's rights in this Article. But it is also a problem for Meta's users—

---

[78] *Id.* at § I.BBB.

[79] Smith v. California, 361 U.S. 147, 151 (1959) ("[S]tricter standards of permissible statutory vagueness may be applied to a statute having a potentially inhibiting effect on speech; a man may the less be required to act at his peril here, because the free dissemination of ideas may be the loser.").

[80] F.C.C. v. Fox Television Stations, Inc., 567 U.S. 239, 253–54 (2012).

vague definitions will incentivize Meta to err on the side of caution and over-moderate in order to avoid even the potential for liability.[81] Indeed, for exactly this reason, vagueness concerns are especially acute in the context of speech intermediaries that host or distribute the speech of others.[82] As First Amendment cases recognize, intermediaries are particularly prone to chilling speech as a result of risk aversion because "intermediaries will often be less invested in the speaker's message and thus less likely to risk the regulator's ire."[83] Even though intermediaries have a general profit motive in hosting content, the marginal gain from any individual piece of borderline content is usually minimal and thus their commitment to hosting it is weak.

Courts have cited concerns about vagueness as a reason to strike down laws that attempt to limit minors' access to harmful online content. In *CCIA v. Paxton*, for example, a district court struck down as unconstitutionally vague a law requiring platforms to develop strategies to prevent a minor's exposure to "harmful material and other content that promotes, glorifies, or facilitates (1) suicide, self-harm, or eating disorders; (2) substance abuse; (3) stalking, bullying, or harassment; or (4) grooming, trafficking, child pornography, or other sexual exploitation or abuse."[84] "Given the substantial liability companies face for failing to comply," the court explained, "it is reasonable to expect that companies will adopt broad definitions" of the categories listed in the law; elsewhere, the court warned that the law's

---

[81] This is the Court's rationale in *Smith*, 361 U.S. at 151. Empirical work confirms these incentives operate in the social media context, if not even more acutely given the sheer scale of content on their services. *See, e.g.*, Daphne Keller, *Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List*, STANFORD CIS (Feb. 8, 2021), https://perma.cc/5BPM-6A5E (collecting studies).

[82] Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 29 (2006) ("Intermediaries are peculiarly susceptible to chill, for they often face cost and revenue structures quite different from those of first-party speakers."); Genevieve Lakier, *Enforcing the First Amendment in an Era of Jawboning*, 93 U. CHI. L. REV. 781, 791 (2026).

[83] Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175, 198 (2024) (quoting Brief for First Amendment Scholars as Amici Curiae Supporting Petitioner at 8, *Vullo*, 602 U.S. 175 (No. 22-842)).

[84] Computer & Commc'ns Indus. Ass'n v. Paxton, 747 F. Supp. 3d 1011, 1040 (W.D. Tex. 2024), *aff'd*, 184 F.4th 420 (5th Cir. 2026). On appeal, the Fifth Circuit held that these requirements were preempted by federal law and did not reach the constitutional issues. Computer & Commc'ns Indus. Ass'n v. Paxton, 184 F.4th 420, 444 (5th Cir. 2026), *vacated*, 2026 WL 3231536, at *1 (5th Cir. Oct. 28, 2026).

vague terms "will result in wide-ranging censorship of speech."[85] Worse, the vague definitions made it "easy to see how an attorney general could arbitrarily discriminate in his enforcement of the law,"[86] especially in light of the "potential wide breadth and politically charged nature"[87] of the kinds of content involved. "[P]ro-LGBTQ content might be especially targeted for 'grooming,'" the court suggested as an example, or "[c]ontent related to marijuana use might be prosecuted as 'glorifying' 'substance abuse,' even if cigarette and alcohol use is not."[88]

Another district court held that NetChoice had a plausible case that a Maryland law requiring platforms to act in the "best interests of children" in their data use and platform design choices was unconstitutionally vague.[89] Unlike in the family law context where courts apply this standard in fact-specific circumstances, the court noted that the Maryland law would require platforms to follow the standard on a much broader scale, thereby requiring them to make "subjective determinations."[90] In a different case, the Ninth Circuit similarly held that "what is 'materially detrimental' to any single child's 'well-being' is not clear"[91] and therefore found that a California law prohibiting covered businesses from using a child's personal information in a way that the business knew, or had reason to know, was "materially detrimental to the . . . well-being of a child" was likely unconstitutionally vague.[92]

As these cases demonstrate, courts tend to be wary—for good reason—of broad mandates that place obligations on platforms to make sure kids don't see harmful content. The Settlement Agreement seems to try to avoid some of these problems by relying on Meta's existing standards as supplying some of the relevant definitions, but the Agreement also makes clear that these definitions are not exhaustive and that Meta's rules must continue to change over time to cover new

---

[85] Computer & Commc'ns Indus. Ass'n v. Paxton, 747 F. Supp. 3d at 1040.

[86] *Id.* at 1041.

[87] *Id.* at 1036.

[88] *Id.* at 1041.

[89] NetChoice v. Brown, No. CV RDB-25-0322, 2025 WL 3267786, at *19 (D. Md. Nov. 24, 2025).

[90] *Id.*

[91] NetChoice, LLC v. Bonta, 170 F.4th 744, 766 (9th Cir. 2026).

[92] *Id.* at 767.

forms of content that causes harm. And exactly how that "harmful content" is to be defined, and who gets the final say, remains uncertain.

These concerns about vagueness would be a problem even if the Agreement only targeted unprotected speech.[93] But, in fact, the list of content that Meta needs to limit to comply with the Agreement includes primarily protected expression.

Some of the categories of content are such that it is perhaps unsurprising that there has been little outcry about the attempt to limit kids' access to them.[94] Doctrinally, however, the Court has emphatically rejected the idea that there is some "free-floating power to restrict the ideas to which children may be exposed,"[95] even to content that few might find valuable. For this reason, the Court struck down a California law that restricted minors' access to graphically violent video games.[96] Normatively, there are good reasons to worry about giving the government the power to define what content is or is not appropriate for children. The politically diverse Attorneys General may all agree at a high level about the need to protect kids, but one only needs to glance at debates about sexual and gender identity, or reproductive care, or a myriad of other topics, in this country today to know that the settling states may have very different understandings of what protecting kids means in practice.[97]

Even beyond politically fraught topics, the list of content that Meta must reduce teen access to is stunningly broad. It may be difficult to work out exactly how much of this content a teen should see, but it's not obvious that less is always preferable.

---

[93] *See, e.g.*, Smith v. California, 361 U.S. 147 (1959).

[94] The definition covers "content prohibited by Meta's Community Standards concerning bullying and harassment; nudity and sexual activity; child sexual exploitation, abuse, and nudity; sexually explicit language; suicide, self-harm and eating disorders; graphic violence and incitements to violence; gambling; and restricted substances or goods (including illegal drug use), as well as policies specifically focused on protections for Teen Users, including those regarding high-risk viral challenges and risky stunts." Settlement Agreement, *supra* note 1, at § I.F.

[95] Brown v. Ent. Merchants Ass'n, 564 U.S. 786, 794 (2011).

[96] *Id.*

[97] *See, e.g.*, Mike Masnick, *Heritage Foundation Says That of Course GOP Will Use KOSA to Censor LGBTQ Content*, TECHDIRT (May 24, 2023), https://perma.cc/34MM-FEGN; Mike Masnick, *KOSA Won't Just Silence LGBTQ Voices; It Will Also Be Used to Hide Abortion Info from the Internet*, TECHDIRT (Oct. 20, 2023), https://perma.cc/7AMC-F56Q.

Some teens may be in communities where discussions about such topics are essential, such as sports teams or ballet dancers discussing physical accomplishments, or the vast expanse of teen literature and music that describes the emotional distress of adolescence. Recall, too, that this list of topics applies apparently indiscriminately to Teen Users from 13 years old to 17 years and 11 months old—a diverse group that has different capacities and needs. These provisions of the Settlement Agreement are no doubt well-intentioned, but they are not well-defined or well-founded.

II. THE SETTLEMENT AGREEMENT IS JAWBONING

The Settlement Agreement purports to impose the burdens on third-party speech rights outlined in Part I while bypassing constitutional scrutiny. The First Amendment does not tolerate this kind of constitutional evasion, however, and this Part argues that the Settlement Agreement is therefore constitutionally infirm.

Before proceeding, I want to be clear about what I am *not* arguing. My goal in this Part is not to make any argument about the general pros and cons of using litigation and settlement agreements as a means of regulating.[98] This is not a new model, and we can accept for the purposes of this Part that there will be some circumstances in which these tools can form an important part of the regulatory toolbox.[99]

Note, too, that nothing in my argument should be read as a statement one way or the other on the strength of the claims in the Attorneys General's lawsuit. This was a novel lawsuit which, to my mind (as Part I might suggest), raised some difficult First Amendment questions. But it reflected a legal theory that has been gaining traction in the courts—that platforms can and should be held liable for the manipulative and addictive ways in which they design their products, and that this can be distinguished from holding them liable for the content on their services—and there is no reason to doubt that it was brought in good faith rather than, say, as a means of retaliation, a form of harassment, or purely to coerce a settlement.

---

[98] *See generally* PAUL NOLETTE, FEDERALISM ON TRIAL: STATE ATTORNEYS GENERAL AND NATIONAL POLICYMAKING IN CONTEMPORARY AMERICA (2015). *See also* sources cited *infra* note 180.

[99] *See, e.g.*, David Pozen, *Balkinization: Regulation by Deal Comes to Higher Ed*, BALKINIZATION (Jul. 23, 2025) ("Regulation by deal may have a useful role to play under certain conditions."), https://perma.cc/6L9H-WG4S; Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014) (recounting how settlement agreements became the primary form of privacy regulation in the United States).

Finally, and most importantly, nothing I say here should be read to reflect a concern about *Meta*'s constitutional rights, as opposed to those of its users. Parties can generally choose to waive their own constitutional rights, although the conditions under which a court will enforce such waivers are not entirely settled. Courts will enforce non-disparagement clauses or non-disclosure agreements even though they prevent a party from engaging in what would otherwise be protected speech.[100] This can be the case even where the party enforcing such a clause is the government. For example, in *Snepp v. United States*, the Supreme Court enforced a former CIA employee's contractual promise, made when hired, to submit any writings about the Agency for prepublication review.[101] The Court cursorily dismissed a First Amendment argument against such an obligation because the employee "voluntarily signed the agreement" that included this term.[102]

Snepp, however, squarely implicated national security concerns, and it is unclear if it would apply more broadly.[103] Some courts have refused to enforce non-disparagement clauses in settlement agreements between government actors and individuals when they find them against the public interest.[104] Further complicating the question, the doctrine of unconstitutional conditions also restricts the government's ability to use contractual or other benefits to induce parties to waive their

---

[100] *See, e.g.*, Cohen v. Cowles Media Co., 501 U.S. 663, 671 (1991) (allowing damages for breach of promise of confidentiality because under promissory estoppel, "[t]he parties themselves, as in this case, determine the scope of their legal obligations, and any restrictions that may be placed on the publication of truthful information are self-imposed"); Lothamer Tax Resol., Inc. v. Kimmel, No. 1:25-CV-579, 2025 WL 2490380, at *12 (W.D. Mich. Aug. 29, 2025) (summarizing precedent that "largely supports the principle that a person can waive their free speech rights via contract, including their right to not face prior restraints from a court"); Democratic Nat. Comm. v. Republican Nat. Comm., 673 F.3d 192, 205 (3d Cir. 2012) ("Court enforcement of a private agreement to limit a party's ability to speak or associate does not necessarily violate the First Amendment.").

[101] 444 U.S. 507 (1980).

[102] *Id.* at 508–09 n.3.

[103] *Id.* (concluding that "[t]he Government has a compelling interest in protecting both the secrecy of information important to our national security and the appearance of confidentiality so essential to the effective operation of our foreign intelligence service. The agreement that Snepp signed is a reasonable means for protecting this vital interest").

[104] *See, e.g.*, Note, Overbey v. Mayor of Baltimore, 133 HARV. L. REV. 1460, 1465–66, 1465 n.56 (2020).

constitutional rights in some circumstances.[105] Exactly how the First Amendment constrains government enforcement of contractual terms restricting a counterparty's speech is therefore thorny and doctrinally uncertain.

I can set that issue aside here, though. Meta waived any right to challenge the Agreement, including on the ground that it is unconstitutional,[106] and we can assume that Meta itself will not seek to renege on this promise. In this way, calling the Settlement Agreement social media's "Big Tobacco moment"[107] might be especially apt. Under the 1998 Master Settlement Agreement with states, the four major tobacco companies agreed to stop billboard ads and marketing targeted at minors.[108] This provision was never challenged, even though similar but less restrictive Massachusetts regulations were later struck down in part as violating the First Amendment.[109]

In the same way, provisions of the Settlement Agreement that restrict Meta's speech rights (for example, limits on sending push notifications, requiring Meta to provide users an option to select a non-personalized feed, or mandating that Meta implement "productive pauses and notices . . . designed to reduce or prevent excessive, mindless or unintended teen usage" including "a Clear and Conspicuous notice upon any 15-minute session of continuous usage") are likely to withstand constitutional scrutiny even if they would raise First Amendment issues were they

---

[105] Agency for Int'l Dev. v. All. for Open Soc'y Int'l, Inc., 570 U.S. 205, 214 (2013) ("In some cases, a funding condition can result in an unconstitutional burden on First Amendment rights.").

[106] Consent Judgment at § IX.G, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-md-03047-YGR (N.D. Cal. Aug. 26, 2026), ECF No. 3451.

[107] *See, e.g.*, Nathan Borney & Mary Helen Moore, *Meta Settlement Gives Social Media Its Big Tobacco Moment*, AXIOS (Aug. 26, 2026), https://www.axios.com/local/raleigh/2026/08/26/north-carolina-meta-settlement-teen-safety; Queenie Wong & Nilesh Christopher, *Meta's "Big Tobacco Moment" Leaves Parents Asking if Social Media Is Too Dangerous for Kids*, L.A. TIMES (Aug. 28, 2026).

[108] *Master Settlement Agreement* at §§ III(a), III(d)(1)–(2), NAT'L ASS'N OF ATT'YS GEN (Nov. 23, 1998), https://www.naag.org/wp-content/uploads/2020/09/2019-01-MSA-and-Exhibits-Final.pdf; Nora Freeman Engstrom & Robert L. Rabin, *Pursuing Public Health Through Litigation*, 73 STAN. L. REV. 285, 342 n.285 (2021).

[109] Lorillard Tobacco Co. v. Reilly, 533 U.S. 525, 561–66 (2001).

implemented through legislation.[110] (Indeed, Meta is a member of the trade association NetChoice, which is—successfully—challenging many similar legislative requirements on First Amendment grounds.)

There is a caveat to this conclusion, though. The First Amendment also protects a right to listen that in certain circumstances allows people to challenge speech restrictions on others on the basis that they wanted to hear what that other had to say.[111] In many cases of waiver, listeners will not have a right to challenge someone else's waiver of their rights and compel someone else to speak. After all, "[f]reedom of speech presupposes a willing speaker," and the First Amendment protects listener rights only "where [such] a speaker exists."[112] Thus, a listener rights theory will generally not succeed when a speaker waives only their own speech rights directly—as in, for example, individual plea agreements or agreements like the Tobacco settlement.[113]

But the analysis may be more complicated in the case of speech platforms like Meta, where their First Amendment-protected decisions (their algorithms, affordances and content moderation) affect the listener rights of so many third parties, and not only affect Meta's own capacity to engage in expression but also the rights of users to communicate with each other. We might especially worry about allowing listener rights to be so easily waived in this context because of the intermediary incentives to avoid resisting government pressure discussed earlier. The scope of listener interests here is unclear, and I bracket this question for now.

But if the law around the extent to which the government can require a party to waive *their own* constitutional rights is uncertain, the same cannot be true of the

---

[110] *See, e.g.*, NetChoice v. Weiser, 808 F. Supp. 3d 1223, 1240 (D. Colo. 2025) (finding a similar notice requirement was likely unconstitutional compelled speech); NetChoice v. Griffin, 831 F. Supp. 3d 816, 831 (W.D. Ark. 2026) (finding that a law that restricted notifications overnight was likely unconstitutional).

[111] *See, e.g.*, Murthy v. Missouri, 603 U.S. 43, 75 (2024) ("[W]e have recognized a 'First Amendment right to 'receive information and ideas.'").

[112] Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, Inc., 425 U.S. 748, 756 (1976).

[113] *But see* Overbey v. Mayor of Baltimore, 930 F.3d 215, 229 (4th Cir. 2019) (finding that a local newspaper "has plausibly alleged that its legally protected interest in newsgathering has, at some point, been invaded by the City's use of non-disparagement clauses in settlement agreements with police brutality claimants").

extent to which the government can use agreements to restrict the First Amendment rights of *third parties* directly. Here the law is unambiguous: The government cannot use agreements with private parties to side-step constitutional constraints.

A. The Capacious Rule Against Jawboning

As the Supreme Court recently unanimously reaffirmed, "a government official cannot do indirectly what she is barred from doing directly: A government official cannot coerce a private party to punish or suppress disfavored speech on her behalf."[114] This rule against what is often called "jawboning"[115] is an important safeguard—it prevents government actors from evading the constraints of the First Amendment by simply enlisting private actors to do their speech suppression for them. By requiring Meta to restrict the speech rights of its users, the Settlement Agreement violates this rule.

The recent emphatic reaffirmation of the rule against jawboning came in *National Rifle Association v. Vullo*,[116] which should control the constitutional analysis of the Settlement Agreement. That case concerned allegations[117] that the superintendent of the New York Department of Financial Services (DFS), Maria Vullo, pressured certain insurance companies to stop doing business with the NRA in order to suppress the NRA's pro-Second Amendment advocacy.[118] The alleged coercion took a number of forms. Following a tip, Vullo had investigated the insurance companies and found that some of the services they provided to the NRA violated New York law.[119]

In a series of private meetings and public statements, Vullo then allegedly threatened to conduct broader investigations into the insurance companies' businesses unless they stopped doing business with the NRA.[120] Vullo's pressure campaign culminated in the DFS entering into consent decrees with the insurance com-

---

[114] Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175, 190 (2024).

[115] Lakier, *supra* note 82, at 784.

[116] 602 U.S. 175 (2024).

[117] The case reached the Court on appeal from a motion to dismiss, and so the Court assumed the truth of the factual allegations in the complaint. *Id.* at 181.

[118] *Id.* at 187.

[119] *Id.* at 175.

[120] *Id.* at 183–84.

panies in which the companies admitted liability, paid fines, and agreed not to provide any further NRA-endorsed insurance programs, even if lawful.[121] Justice Sotomayor, for a unanimous Court, held that while "Vullo was free to criticize the NRA and pursue the conceded violations of New York insurance law,"[122] the alleged pressure campaign, if proved, would have crossed a constitutional line because Vullo "wield[ed] her power . . . to threaten enforcement actions against DFS-regulated entities in order to punish or suppress the NRA's gun-promotion advocacy."[123]

Note some important, and remarkable, features of *Vullo*. Vullo's investigations were within her remit as superintendent—there was no suggestion that they exceeded her authority. Nor did it matter that the insurance companies had, in fact, violated New York law. "[T]he conceded illegality of the NRA-endorsed insurance programs does not insulate Vullo from First Amendment scrutiny," the Court made clear.[124] Vullo's alleged actions were unconstitutional notwithstanding this because they were done to suppress disfavored expression.[125]

Vullo and its consent decrees are, in material respects, on all fours with the Meta Settlement Agreement. The Attorneys General have power to investigate unfair and deceptive practices, and they can bring legal actions for violations of privacy laws. What they cannot do is use these powers to achieve the suppression of protected expression, including of third parties: here, the rights of Meta's users.

It is true that in this case, the Attorneys General are seeking to protect the third parties (children) rather than punish them (as Vullo was allegedly trying to do to the NRA), but this is a distinction without a difference under the First Amendment: Paternalistic censorship is as much a constitutional violation as politically motivated censorship.[126] It is also true that the Court in *Vullo* emphasized the allegation

---

[121] *Id.* at 176.

[122] *Id.* at 187.

[123] *Id.*

[124] *Id.* at 196.

[125] *Id.* at 188.

[126] Brown v. Ent. Merchants Ass'n, 564 U.S. 786, 794 (2011) ("No doubt a State possesses legitimate power to protect children from harm, but that does not include a free-floating power to restrict the ideas to which children may be exposed.") (citation omitted).

that Vullo targeted the NRA because it expressed particular views,[127] while many of the speech restrictions in the Settlement Agreement might be described as content-neutral. But, as Part I showed, not all the provisions can be so described: The age-appropriate content provisions are explicitly not only content- but also viewpoint-based. Furthermore, there is no reason why the rule from *Vullo* should not extend to overly burdensome content-neutral restrictions too: At its core, the rule is about preventing the government from circumventing the First Amendment's protections.[128]

Indeed, in some ways the First Amendment arguments against the Meta Settlement Agreement are stronger than those in *Vullo*. In *Vullo*, the insurance companies admitted liability, whereas Meta did not.[129] And in *Vullo*, the consent decrees did not include speech-restrictive terms themselves but instead targeted non-expressive business activities and insurance services.

Vullo emphasized both points in arguing that in entering into the consent decrees and otherwise pressuring the insurance companies to cease business with the NRA, she was simply carrying out her regulatory responsibilities.[130] The Court dismissed these arguments, however, holding that even though the consent decrees included admitted infringements and did not themselves directly concern speech, they nonetheless worked to suppress protected expression by depriving the NRA of resources. "The analogy," the Court said, quoting a Seventh Circuit jawboning case, "is to killing a person by cutting off his oxygen supply rather than by shooting him."[131] By contrast, the speech restrictions are included in the Meta Settlement Agreement itself—it is a direct shot.

---

[127] *Vullo*, 602 U.S. at 187 ("At the heart of the First Amendment's Free Speech Clause is the recognition that viewpoint discrimination is uniquely harmful to a free and democratic society.").

[128] Indeed, if it were otherwise, government actors would be incentivized to engage in broader content-neutral jawboning to achieve their speech suppressive ends but avoid the rule against viewpoint-based jawboning. *See also* Lakier, *supra* note 82, at 784 (describing *Vullo* as making clear "that what the First Amendment rule against informal censorship [or what today tends to be referred to simply as 'jawboning'] prohibits is the practice of constitutional evasion").

[129] Consent Judgment, *supra* note 106, at § X.C ("This Consent Judgment and the Agreement were entered into for settlement purposes only and do not constitute an admission by Defendant of any liability, wrongdoing, or violation of any local, state, federal or international law.").

[130] *Vullo*, 602 U.S. at 194–97.

[131] *Id.*

Ironically, this may be in part why the First Amendment concerns with the Agreement have not attracted as much outcry. The standard definition of jawboning emphasizes the use of *informal* government power to suppress speech.[132] Typical jawboning cases include government actors who write informal letters that threaten or imply potential legal consequences but themselves have no formal legal force,[133] or officials who have private meetings or contact with private actors in which they pressure them to take certain actions.[134]

Indeed, if instead of filing a complaint the Attorneys General had held non-public meetings or had sent private letters to Meta proposing that the company take the very actions outlined in the Settlement Agreement, or else they would bring a lawsuit seeking $1.4 trillion in penalties,[135] it would be easy to conceptualize this as a jawboning case, and the constitutional problem would be immediately apparent. The Settlement Agreement, by contrast, is not informal at all. It is a formal legal agreement, out in the open, approved by a court and having the force of law.

But the fact that jawboning typically involves the use of informal government power does not mean that this is the only way it can occur. The pressure campaign found to violate the First Amendment in *Vullo* itself involved consent decrees, much like the Meta Settlement Agreement. True, the case did not involve a direct challenge to the terms of those decrees and instead focused on Vullo's broader campaign. But it is hard to see why this should matter.

The jawboning cases emphasize that courts should not focus on the precise form of the government actions being challenged, but should instead engage in a

---

[132] *See, e.g.*, Lakier, *supra* note 82, at 784 (defining jawboning as "the purposeful use of informal governmental power to evade the constraints that the First Amendment imposes on the government's formal powers"); Derek E. Bambauer, *Against Jawboning*, 100 MINN. L. REV. 51, 57 (2015) (similarly emphasizing the use of "informal" government power); Blake E. Reid, *Jawbreaking and Counterboning*, 58 CONN. L. REV. 905, 907 (2026); Daniel Browning, *Jawboning as Collaborative Governance*, 6 J. FREE SPEECH L. 1245, 1253 (2026).

[133] *See, e.g.*, Bantam Books, Inc. v. Sullivan, 372 U.S. 58 (1963); Kennedy v. Warren, 66 F.4th 1199 (9th Cir. 2023); Backpage.com, LLC v. Dart, 807 F.3d 229 (7th Cir. 2015); Okwedy v. Molinari, 333 F.3d 339 (2d Cir. 2003); Penthouse Int'l, Ltd. v. Meese, 939 F.2d 1011 (D.C. Cir. 1991).

[134] *See, e.g.*, Murthy v. Missouri, 603 U.S. 43 (2024).

[135] Diana Novak Jones, *Meta Says US States Are Seeking $1.4 Trillion in Penalties in August Youth Safety Trial*, REUTERS (Jul. 7, 2026), https://www.reuters.com/business/meta-says-us-states-are-seeking-14-trillion-penalties-august-youth-safety-trial-2026-07-07/.

big-picture, realistic analysis of how governmental power is being wielded. *Bantam Books v. Sullivan*,[136] the foundational jawboning case which the *Vullo* Court made clear continues to guide jawboning analysis today,[137] was explicit on this point. Courts must "look through forms to the substance" in assessing when and how governmental power is operating to suppress speech.[138] Justice Gorsuch wrote a separate concurrence in *Vullo* just to underline the holistic and non-formalist nature of the jawboning inquiry.[139] Indeed, the important doctrinal innovation of the jawboning cases is to say that *informal* governmental power can *also* violate the Constitution, not that formal governmental power should be subject to *less* scrutiny.

Any other analysis would produce an absurd result. If the state passes a law that induces or requires a private party to restrict the rights of others, the law will be subject to constitutional constraints.[140] If a government actor forms a relationship with a private party with respect to certain conduct such that the party's actions can be said to be fairly attributable to the state, then those actions can be deemed state action and subject to constitutional scrutiny.[141] And if the government informally coerces or induces a private party to take action against third parties in a way that would violate those third parties' rights if the state were the actor, that would violate the rule against jawboning.[142]

---

[136] 372 U.S. 58 (1963).

[137] *Vullo*, 602 U.S. at 197 ("The Court does not break new ground in deciding this case. It only reaffirms the general principle from *Bantam Books*.").

[138] Bantam Books, Inc. v. Sullivan, 372 U.S. 58, 67 (1963).

[139] Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175, 199 (2024) (Gorsuch, J., concurring).

[140] *See, e.g.*, Skinner v. Ry. Lab. Executives' Ass'n, 489 U.S. 602, 615–16 (1989) (finding that a regulatory scheme designed to induce private railroads to test their employees for drugs and alcohol, including by providing them certain immunities, showed significant state "encouragement, endorsement, and participation" to constitute state action); Smith v. California, 361 U.S. 147, 154 (1959) (finding that a statute that incentivized booksellers not to stock certain books implicated the First Amendment because "[t]he bookseller's self-censorship, compelled by the State, would be a censorship affecting the whole public, hardly less virulent for being privately administered").

[141] *See, e.g.*, Lugar v. Edmondson Oil Co., 457 U.S. 922, 937 (1982); West v. Atkins, 487 U.S. 42, 54 (1988) (finding that the actions of a doctor contracted to provide medical services to state prison inmates were "conduct . . . fairly attributable to the State").

[142] *See supra* text accompanying notes 122–125.

To find that the Settlement Agreement is not also subject to constitutional scrutiny would be to allow the rights violations contained within it to slip between the cracks of these many rules designed to prevent state actors from finding loopholes in the Constitution—exactly the kind of formalistic outcome *Bantam Books* and *Vullo* warn against. It would mean that the First Amendment would prevent a government actor from coercively threatening to bring litigation or signing a contract intended to have Meta restrict the speech rights of its users, but if the government actor brought litigation and *then* signed such a contract, the First Amendment would be nowhere to be seen. This is untenable.

The strongest argument to the contrary is that unlike the foregoing examples, Meta's decision to settle its case is a voluntary act, and not one taken at the state's behest. There is no doubt that if Meta were to independently make the changes in the Settlement Agreement, the First Amendment would not be implicated.[143] Similarly, the argument might run, Meta could voluntarily agree to make these changes as a way of managing its litigation risk without the Constitution being implicated. Indeed, the framing of the content-based provisions in the Agreement seems aimed at setting up this argument: They suggest that Meta is not being forced by the Agreement to restrict any speech but instead merely required to continue to enforce and improve its own Community Standards. This argument seems designed to invoke *Murthy v. Missouri*, which involved a First Amendment challenge by plaintiffs whose posts were removed from social media after government actors pressured platforms to remove certain kinds of content.[144] The Court denied plaintiffs standing in part because the platforms had independent incentives to engage in content moderation, and so the plaintiffs had not proven that the suppression of their content was traceable in any particular instance to the governmental pressure.[145]

Likewise, state action cases have been clear that a mere contractual relationship between a private party and the state does not make the private party's exercise of

---

[143] *See* Moody v. NetChoice, LLC, 603 U.S. 707, 740 (2024) ("When the platforms use their Standards and Guidelines to decide which third-party content those feeds will display, or how the display will be ordered and organized, they are making expressive choices. And because that is true, they receive First Amendment protection."). *See also* Manhattan Cmty. Access Corp. v. Halleck, 587 U.S. 802, 808 (2019) ("By enforcing that constitutional boundary between the governmental and the private, the state-action doctrine protects a robust sphere of individual liberty.").

[144] 603 U.S. 43 (2024).

[145] *Id.* at 62, 68.

its own discretion state action for constitutional purposes.[146] Instead, "a State normally can be held responsible for a private decision only when it has exercised coercive power or has provided such significant encouragement, either overt or covert, that the choice must in law be deemed to be that of the State."[147]

But these cases limiting the circumstances in which government actors can be held constitutionally responsible for the actions of private parties do not apply to the Settlement Agreement for a number of reasons. First, the link between the government pressure and the change in platform rules in the case of the Meta Settlement Agreement is far clearer and more direct than in *Murthy*. In *Murthy*, the Court did not suggest that such a link *could* not establish state action, but simply that the evidence, which the Court described as "murky,"[148] did not do so in that case. By contrast, the Settlement Agreement is the explicit source of the specific speech restrictions Meta has agreed to impose. Many of the obligations are new obligations that Meta has aggressively fought in court through its trade association and would not be implementing but for the Agreement. The potential trillion-dollar liability in the state suit would certainly be coercive, just as the Court found it could be in *Vullo*.

The fact that the most explicitly viewpoint- and content-based commitments (those concerning age appropriate content, discussed in Part I.C above) are described as Meta agreeing to "maintain" its own existing practices does not make those commitments akin to *Murthy* either. First, that Meta's policies were voluntary before the Settlement Agreement does not render truly voluntary an ongoing legal obligation—extracted through government pressure—to maintain those policies on pain of contempt of court. This is enough to distinguish the case from *Murthy*, where the Court found that after the governmental pressure campaign at issue, "the platforms remain free to enforce, or not to enforce, [their] policies."[149] This is precisely the state of affairs the Settlement Agreement seeks to change—Meta is no

---

[146] *Manhattan Cmty. Access Corp.*, 587 U.S. at 814 ("[A]s the Court has long held, the fact that the government licenses, contracts with, or grants a monopoly to a private entity does not convert the private entity into a state actor.").

[147] Blum v. Yaretsky, 457 U.S. 991, 1004 (1982).

[148] Murthy v. Missouri, 603 U.S. 43, 68 n.8 (2024).

[149] *Id.* at 73.

longer "free . . . not to enforce [its] policies," but instead could be brought to court for violating the Settlement Agreement if it changes its rules.

And this fact implicates users' rights, not just Meta's rights. That Meta's users have no First Amendment right to complain about Meta's content moderation actions taken against them does not mean that they have no First Amendment right to complain when those actions are undertaken at the behest of the government. As the Court said in the context of a state law that required employers to fire their at-will employees in violation of the Fourteenth Amendment, "[t]he fact that the employment is at the will of the parties, respectively, does not make it one at the will of others. The employee has manifest interest in the freedom of the employer to exercise his judgment without illegal interference or compulsion."[150]

Second, the ongoing oversight by the Attorneys General of Meta's enforcement of its Community Standards creates additional First Amendment problems, as it gives the Attorneys General leverage to insist on their own interpretations of the Settlement Agreement's vague terms.[151] This is especially so given that Meta has agreed not only to maintain its existing policies but also to "continuously improve" those policies and change them where they are "not capturing content or experiences that are causing harm to Teen Users."[152] As a result, the Attorneys General enjoy considerable influence over the future direction of Meta's speech rules that they would not have in the absence of the Agreement. This suggests ongoing coercive power.

None of this analysis is undermined by the Consent Judgment's statement that Meta is entering into the agreement "freely and voluntarily . . . without any degree of duress or compulsion."[153] That the Agreement contains an assertion to this effect is hardly surprising. As Judge Posner observed in a Seventh Circuit jawboning case cited approvingly by the Supreme Court in *Vullo*, "what would one expect [a target

---

[150] Truax v. Raich, 239 U.S. 33, 38 (1915).

[151] *See* Daphne Keller, *Platform Transparency and the First Amendment*, 4 J. FREE SPEECH L. 1, 22–29 (2023); Eric Goldman, *The Constitutionality of Mandating Editorial Transparency*, 73 HASTINGS L.J. 1203 (2022).

[152] *See* discussion *supra* Part I.C.

[153] Consent Judgment, *supra* note 106, at § V.

of coercion] to say?"[154] Targets "will be reluctant to acknowledge that [they are] submitting to threats but will instead ascribe [their] abandonment of the activity to [their] having discovered that it offends [their] moral principles."[155] But this does not make the protestations true.

For the same reason, Meta's newfound enthusiasm for changes to its platform that it has long fought cannot be plausibly viewed as something arising independently of government pressure, even if Meta attests that it is making the changes "voluntarily." California Attorney General Rob Bonta put it plainly: "They didn't agree voluntarily out of the goodness of their heart. We had a trial where we were hammering them. We had a litigation and a trial verdict hanging over their head, which was not going well for them."[156]

Thus, "look[ing] through forms to the substance,"[157] the First Amendment must apply to the Settlement Agreement. The capacious and anti-formalist rule against jawboning is intended precisely to capture these kinds of new and innovative ways in which government actors might evade constitutional constraints.

B. The Standard of Review

The conclusion that the Settlement Agreement implicates the First Amendment does not automatically mean it is unconstitutional. This section turns to the question of what standard of review should apply to cases such as the Settlement Agreement.

As Genevieve Lakier has written, *Bantam Books* and *Vullo* suggest a categorical rule applies once the government is found to have engaged in jawboning—that is, that "all efforts . . . to coerce the private suppression of speech" are per se unconstitutional, regardless of whether the speech is protected, regardless of whether the efforts are actually so strong as to be coercive, and regardless of whether the efforts actually succeed.[158] This does not mean that all informal contacts between state actors and persons suspected of violating the law are unconstitutional—the Court in

---

[154] Backpage.com, LLC v. Dart, 807 F.3d 229, 233 (7th Cir. 2015); Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175, 190 (2024).

[155] *Backpage.com*, 807 F.3d at 238.

[156] Myrow, *supra* note 18.

[157] Bantam Books, Inc. v. Sullivan, 372 U.S. 58, 67 (1963).

[158] Lakier, *supra* note 82, at 832.

Bantam Books distinguished between contact "genuinely undertaken with the purpose of aiding the [private party] to comply with [the] laws and avoid prosecution under them," which was constitutionally permissible, and contact made with the intent "to suppress," which was not.[159] This is a boundary question about whether the government action is jawboning at all. But once a government actor *had* been found to have engaged in jawboning, the Court in both *Bantam Books* and subsequently *Vullo* found that it was categorically impermissible. That is, it "did not engage in any further balancing of means and ends"[160] in the way First Amendment doctrine often requires after a court finds a governmental action burdens protected speech. Applying this rule to the Settlement Agreement would mean that the Agreement cannot stand once it is shown to implicate the rights of third parties.

The reason for this extremely broad rule is prophylactic.[161] Given the intermediaries' incentives to simply suppress the speech of others that they are not invested in, as well as the informal (and thus often private) nature of jawboning, much jawboning will not be subject to constitutional challenge; those pressured will often simply choose to comply, and no one else will necessarily know about the government's involvement. Thus, the Court has explained that this kind of governmental pressure can be *more dangerous* than the exercise of the formal power to censor.

As the Court put it in *Bantam Books*, "the vice of" jawboning is that by operating outside the formal legal system it "eliminate[s] the safeguards" involved in legal processes.[162] And as the Court further explained in *Vullo*, because jawboning often happens "behind closed doors, the 'ballot box' is an especially poor check on that official's authority."[163] Thus, because both procedural safeguards and public accountability are often lacking in jawboning cases, the First Amendment erects strong prophylactic and categorical protections against it.

Perhaps, though, these dangers are less acute in the context of a settlement agreement which is, after all, a formal legal process overseen by a court. Indeed, it may often be in the public interest for regulators to seek settlements, rather than go through the full process of a trial. Furthermore, there is nothing secret about such

---

[159] 372 U.S. 58, 71–72 (1963).

[160] Lakier, *supra* note 82, at 802.

[161] *Id.* at 832.

[162] 372 U.S. at 69–70.

[163] 602 U.S. at 198.

settlements. In these cases, the third parties whose rights are implicated are fully aware of the governmental role in the outcome. For all these reasons, perhaps the categorical rule can be relaxed in this context, and instead of finding the agreement per se unconstitutional, courts could apply whatever First Amendment standard of review would ordinarily attach to that kind of speech restriction in a given case. Justice Jackson, writing only for herself in a concurrence in *Vullo*, suggested this to be her preferred approach.[164]

The question is a novel one not strictly answered by existing precedent, but this rule seems attractive here. Yet abandoning the categorical rule only makes sense if the reason is because *First Amendment safeguards will be in place*. This requires a court to review the First Amendment implications of any settlement agreement. An analogy here might be to cases where a court hears a challenge from a third party who intervenes in a proceeding to have court records unsealed, asserting its own First Amendment right of access to court records.[165] In such cases, courts regularly consider the First Amendment interests of third parties implicated by the agreement (in those cases, an agreement to seal certain documents).[166]

To ensure that settlement agreements respect First Amendment constraints, third parties whose speech is affected by them must be allowed to challenge the enforcement of those agreements to the extent the agreements restrict their rights. The government would then have to discharge the burden of establishing that such speech restrictions are constitutionally permissible, for instance because they target constitutionally unprotected speech or pass the scrutiny necessary for restrictions on generally protected speech.

Perhaps some of the provisions in the Settlement Agreement might meet this bar. But some might not, for reasons outlined in Part I. Recall that even many of the Attorneys General party to the Agreement acknowledged that its provisions went beyond what they would likely otherwise get from a court. They should know—almost every Attorney General who lost in the cases described in Part I was a party

---

[164] Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175, 200 (2024) (Jackson, J., concurring) ("[T]he fact of coercion, without more, does not state a First Amendment claim. Rather, in addition to finding that the government has crossed a line from persuasion to coercion, courts must assess how that coercion actually violates a speaker's First Amendment rights.").

[165] *See* Eugene Volokh, *Settlements and Third Parties' First Amendment Interests*, 8 J. FREE SPEECH L. 45 (2026).

[166] *Id.*

to the Settlement. And the very court that approved the Settlement Agreement had held that certain features of Meta's platforms that are restricted under the Agreement were protected by the First Amendment.[167]

Affirming that third parties have a right to challenge settlement agreements that affect their speech rights is an important safeguard beyond this particular settlement. It would, for example, create a mechanism through which the hypothetical settlement agreement from the Introduction requiring the suppression of all pro-Democrat content could be challenged and struck down. Courts must make clear that settlement agreements do not create a loophole in the First Amendment.

C. Government Will Continue to Function

A common objection to a broad rule against jawboning is that it will impede the effective operation of the regulatory state.[168] Vullo herself made this argument, insisting that if government officials investigating violations of the laws whose enforcement they oversee and reaching settlement agreements with the entities they regulate were found to violate the First Amendment, it would "wreak havoc at all levels of government."[169] Similarly here, the Attorneys General might argue that it would be untenable to make the enforcement of every settlement agreement they enter into potentially challengeable by third parties. Settlement agreements, after all, can be an important regulatory tool.[170] But nothing I have argued here would unduly handicap the regulatory state.

First, it is worth noting that the Court considered and rejected similar arguments about regulatory capacity in both *Bantam Books* and *Vullo*. The argument

---

[167] *See, e.g.*, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., 702 F. Supp. 3d 809, 837 (N.D. Cal. 2023) (finding, for example, that the timing and clustering of defendants' own notifications is First Amendment-protected).

[168] Lakier, *supra* note 82, at 836 (noting that this belief "was what led many courts to embrace a very narrow interpretation of the First Amendment anti-jawboning rule in the decades prior to *Vullo*"). *See also* Reid, *supra* note 132, at 925. Justice Jackson also raised this concern in her separate concurrence in *Vullo*. 602 U.S. at 201 (Jackson, J., concurring) ("[O]ur democracy can function only if the government can effectively enforce the rules embodied in legislation; by its nature, such enforcement often involves coercion in the form of legal sanctions.").

[169] Brief for Respondent at 50, Nat'l Rifle Ass'n of Am. v. Vullo, 602 U.S. 175 (2024) (No. 22-842), 2024 WL 791438, at *50.

[170] *See, e.g.*, Solove & Hartzog, *supra* note 99.

that finding actions like Vullo's to be jawboning would "interfere with the government's ability to function properly . . . falls flat," said the *Vullo* Court.[171] Government actors are free to carry out their regulatory responsibilities—they just cannot do so in a way that coerces the suppression of protected speech.

Similarly, in *Bantam Books*, Justice Harlan complained in dissent that the Court's opinion would impede the state's ability to deal with "a most baffling social problem" (juvenile delinquency due to reading obscene material).[172] But, as in *Vullo*, the Court had no sympathy for this argument. Government actors are free to engage in legitimate law enforcement activity—they simply cannot do so if their objective is to suppress protected expression.[173] This was surely the aim of the Attorneys General in entering into an agreement that included speech-restrictive terms, with mechanisms for continuing oversight of their enforcement. Indeed, as Part I recounted, many Attorneys General remarked that obtaining such changes was the entire goal of their case.

The Court rejected arguments about regulatory capacity twice, sixty years apart, for good reason. The government wields significant power in the modern state. To conclude that the existence of any formal source of legal authority is enough to evade constitutional review would mean giving the government enormous leverage over private actors to shape the marketplace of ideas as it likes.

And far from spelling the end of the regulatory state, the First Amendment constraint imposed on the government's power to enforce its laws is actually quite limited. Attorneys General can still pursue violations of their states' unfair and deceptive practices laws. They can still enter into settlement agreements with those they investigate and bring action against. But those settlement agreements cannot burden the speech rights of third parties without being subjected to First Amendment scrutiny. The Constitution demands no less, but also no more.

III. The Thwarted Ambitions of Platform Regulation

It is difficult to think of another social problem that has consumed as much public and regulatory attention in recent years with as little to show for it as social media. In the years following the 2016 election, tech executives were hauled before Congress multiple times a year to answer for some problem or other caused by their

---

[171] 602 U.S. 175, 197 (2024).

[172] 372 U.S. 58, 76 (1963) (Harlan, J., dissenting).

[173] *Id.* at 72.

products. Mark Zuckerberg alone has appeared on the Hill at least eight times since 2018, answering questions spanning privacy, child safety, disinformation, political bias, and more.[174] Ire at social media companies has spanned the political spectrum—but often for very different reasons.[175] In part because of the polarized nature of the complaints about social media, all this Congressional attention has resulted in little legislative action.[176]

States have attempted to fill the void. In the last decade alone, there have been dozens of laws passed at the state level regulating platforms. These laws have run into First Amendment problems, however. Texas and Florida's attempts to reimagine the doctrinal status quo to allow greater regulatory leeway in the internet era were largely batted away by the Supreme Court in *Moody v. NetChoice*.[177] And as Part I recounted, state laws have largely continued to flail in the lower courts in *Moody*'s wake.

In such circumstances, it is unsurprising that regulators might start looking for solutions outside the legislative process to achieve their regulatory objectives. Indeed, as Lakier recounts, this is a common historical pattern when regulators are seized by a moral panic and frustrated with the barriers imposed by existing law.[178] Regulators have resorted to jawboning to deal with concerns about smutty magazines, comic books, movies and more over the years.[179] The broad prophylactic rule against jawboning was crafted precisely as a response to this predictable pattern of suppression efforts aimed at finding ways around existing doctrinal hurdles that constrain formal lawmaking.

The Settlement Agreement also reflects some of the risks of using more informal processes. The outcomes reached in a settlement will necessarily be shaped by

---

[174] Cecilia Kang & Mike Isaac, *Mark Zuckerberg, Serial Witness, Takes the Hot Seat Again*, N.Y. TIMES (Apr. 14, 2025).

[175] Evelyn Douek, *The Politics and Perverse Effects of the Fight Against Online Medical Misinformation*, 134 YALE L.J.F. 249 (2025) (describing how "platforms' content-moderation practices became a central battleground in the emerging culture wars").

[176] A notable exception is the law aimed at forcing TikTok to divest from foreign ownership. I have critiqued this legislation elsewhere: Evelyn Douek, *Tiktok v. Garland and the First Amendment Anticanon*, 2025 SUP. CT. REV. 125 (2026).

[177] 603 U.S. 707 (2024). *See also* Douek & Lakier, *supra* note 10, at 133.

[178] Lakier, *supra* note 82, at 794–95.

[179] *Id.*

different considerations than those that attend the lawmaking process.[180] There is no reason to think that an agreement, the terms of which were proposed by Meta, calculated to respond to the reputational and political incentives of the corporation and political actors involved, and settled over bagels in a private meeting with Attorneys General,[181] will adequately account for the interests of all those it affects. This process sidesteps the transparency and deliberative nature of the legislative process. And rather than thinking about whether, as the First Amendment requires, the measures in the agreement are appropriately tailored to the harms they are intended to address, the Attorneys General will be thinking about exactly how much leverage they have.[182]

The problem of potentially misaligned incentives in the settlement process is particularly pronounced when, as in this case, those primarily affected—kids—are barely represented in the settlement conversations or political process. Evidence about whether measures similar to those in the Settlement Agreement that have been imposed overseas have actually benefitted minors is decidedly mixed.[183] And

---

[180] *See, e.g.*, Engstrom & Rabin, *supra* note 108, at 353 n.346 ("[T]ort law is inferior to regulatory activity in certain respects. It is both backward-looking and blunt. It can also be slow, unpredictable, nondemocratic, unduly incremental, 'powerfully influenced . . . by everchanging normative judgments about the scruples of the contestants,' and heavily dependent on 'extraordinary investments [in] lawyering.'") (citation omitted); Steven Davidoff Solomon & David Zaring, *Transactional Administration*, 106 GEO. L.J. 1097, 1104 (2018) ("[U]sing deals to routinely evade government processes affects core values and often lowers the quality of governance."); Donald G. Gifford, *Impersonating the Legislature: State Attorneys General and Parens Patriae Product Litigation*, 49 B.C. L. REV. 913, 961–62 (2008) ("[T]he deliberative legislative process, with its assortment of checks and balances, either constitutionally mandated or arising from legislative tradition, offers comparative institutional advantages over the attorney general's process of deciding whether to regulate a given industry . . . through state-sponsored litigation."); NOLETTE, *supra* note 98, at 13 ("Unlike policymaking in legislatures or administrative agencies in which the public has the opportunity to follow the process via C-Span or notice-and-comment rule-making hearings, the AGs' national policymaking is conducted largely behind closed doors. Out-of-court settlements involving billions of dollars and establishing new regulations require only the approval of a single judge, if at all, to go into effect.").

[181] Kang & Tan, *supra* note 23.

[182] *Id.* (quoting Tennessee Attorney General Skrmetti as encouraging other states to take the deal, saying "Pigs get fat, and hogs get slaughtered").

[183] *See, e.g.*, Courtney Barnes et al., *Assessing Early Effects of Australia's Social Media Minimum Age Act on Adolescents' Social Media Use: Observational Study*, 393 BMJ e363695 (2026); Michael Dezuanni et al., *Australian Teens Impacted by the Social Media Ban Are Getting Less News: New*

while the Settlement Agreement includes monitoring obligations to ensure that Meta is adhering to the terms of the Agreement itself,[184] the auditor's reports are confidential, and there is no obligation to independently study whether or to what extent the changes actually benefit teens.

It is hard, then, not to look at the Settlement Agreement as a sorry indicator of the state of platform regulation and institutional capacity to respond to complex regulatory challenges. Conversations about structural reforms, transparency mandates, independent oversight, or rigorous independent research into exactly what happens on social media platforms and its effects have all stalled in the United States. Instead, the most significant reckoning Meta has faced is to sign a deal that it itself crafted to make potentially ineffective changes aimed at one small segment of the concerns that people have about social media overall. That the Agreement will do *something* is not a reason to overlook its infirmities.

Conclusion

The Settlement Agreement is a product of real, and bipartisan, concern and frustration with social media platforms and the often irresponsible and exploitative ways they have wielded their power over the public sphere in general and their most vulnerable users in particular. Perhaps to some the First Amendment appears merely an annoying obstruction to dealing with these very real problems.

But one only needs to zoom out slightly to see why there has never been a more important time to reaffirm that government actors should not be permitted to evade the constitutional limits on their power by using the threat or reality of a lawsuit to obtain commitments from a private party to suppress the speech of others. The First Amendment must apply to this form of speech regulation as to any other.

Research, The Conversation (May 17, 2026), https://perma.cc/X2SP-7H8N. *See also* Azeen Ghorayshi, *Mental Health Among Teenagers Is Improving, C.D.C. Finds*, N.Y. Times (Sep. 21, 2026) (noting improvements in U.S. teen mental health, even as social media use stayed constant).

[184] Settlement Agreement, *supra* note 1, at § III.