No. 24-271

IN THE UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT

X. CORP.,

Plaintiff-Appellant

v.

ROBERT BONTA,
IN HIS OFFICIAL CAPACITY
AS ATTORNEY GENERAL OF CALIFORNIA,

Defendant-Appellee.

On Appeal from the United States
District Court for the Eastern District of California

BRIEF OF THE ELECTRONIC PRIVACY INFORMATION CENTER AS *AMICUS CURIAE* IN SUPPORT OF DEFENDANT-APPELLEE AND AFFIRMANCE

Megan Iorio
Schuyler Standley
ELECTRONIC PRIVACY
INFORMATION CENTER
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140
iorio@epic.org

March 20, 2024

Attorneys for Amicus Curiae

CORPORATE DISCLOSURE STATEMENT

Pursuant to Fed. R. App. P. 26.1, *amicus curiae* the Electronic Privacy Information Center states that they have no parent corporation and that no publicly held corporation owns 10% or more of their stock.

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Other Authorities

INTEREST OF THE *AMICUS CURIAE*

The Electronic Privacy Information Center ("EPIC") is a public interest research center in Washington, D.C., established in 1994 to focus public attention on emerging privacy and civil liberties issues.[1] EPIC regularly participates as *amicus* in cases concerning platform accountability, the First Amendment, and Section 230.[2]

---

[1] All parties consent to the filing of this brief. In accordance with Rule 29, the undersigned states that no party or party's counsel authored this brief in whole or in part nor contributed money intended to fund the preparation of this brief. No outside person contributed money intended to fund the preparation of this brief.

[2] https://epic.org/?s=&_content-type=amicus-brief.

SUMMARY OF THE ARGUMENT

Social media platforms have a significant impact on society. Platforms use opaque policies, practices, and algorithms to control users' online experiences. The companies' choices in policies, practices, and design can cause harm to users' health and wellbeing as well as to broader societal concerns like democracy, public health, and human rights. Platform transparency is of critical importance because the public cannot understand the harms platforms pose or have a fulsome discussion of how to avoid these harms without understanding how platforms work.

Despite public pressure for transparency, social media companies have resisted public scrutiny of their practices. While platforms provide some disclosures in the form of content moderation policies, reports, and data for independent research, the information provided is limited in scope and not standardized across platforms. Recently, some platforms—most notably X Corp. (hereinafter, "X")—have rolled back their transparency programs and are cracking down on independent research critical of their practices, even bringing lawsuits against researchers whose work they hope to shut down.

Given the importance of platform transparency and platforms' unwillingness to provide it voluntarily, the government's power to compel disclosures must be preserved. A finding that all transparency mandates concerning content moderation are presumptively unconstitutional and can be struck down solely based on hypothetical future harm, as X and its *amici* urge, ignores the public's compelling interest in transparency and would be inconsistent with established First Amendment jurisprudence.

Content moderation transparency requirements do not compel any changes to a platform's message and ensure that the public has access to important information. As the Supreme Court recognized in *Zauderer v. Office of Disc. Counsel*, 471 U.S. 626 (1985), laws that compel a commercial service to disclose truthful, non-controversial information to the public do not involve the same First Amendment interests as laws that prevent companies from engaging in protected speech. To outweigh the government's and public's substantial interest in the free flow of information, particularly regarding content moderation practices, the challenging party must show an actual intrusion on First Amendment

rights. A successful challenge cannot be supported by hypothetical future harm alone.

X has not made the required showing of harm. X makes no showing of administrative burden or harm from public retaliation, and its concern that the Attorney General will misuse his authority to force a change to X's content moderation practices is purely hypothetical. The specter of these consequences, without more, is not enough to support a facial challenge.

ARGUMENT

I. THE GOVERNMENT MUST COMPEL TRANSPARENCY ABOUT PLATFORMS' CONTENT MODERATION PRACTICES.

The public has a robust interest in platform transparency. But social media companies resist public scrutiny. Courts must preserve the government's power to require transparency because the public cannot understand the potential harmful effects of social media without access to information currently locked inside platforms.

A. It is important for individuals to understand how platforms' content moderation practices affect them both as users and as members of society.

Billions of people use social media platforms to connect, share, and talk about topics, from the most mundane to the most urgent. These platforms' content moderation policies and practices impact peoples' experiences as users and as members of society, and the public has an interest in understanding these effects. Platforms' increasing reliance on automation for content moderation heightens concerns for consumer protection, democracy, and human rights.

i. Platforms' content moderation practices are opaque and heavily governed by complex, inscrutable algorithms.

Social media companies' content moderation practices can be broken down into two major categories: (1) removing content and users from the service and (2) selecting and ranking content tailored to

individual users in features such as feeds and search results.[3] Each presents its own unique transparency challenges.

Content removal and user bans are dictated by a combination of a company's content moderation policies and their enforcement. While companies disclose many of their policies in public documents such as terms of service and community standards, some policies and customary practices are kept out of public view. Companies enforce their content moderation policies through a combination of human and automated review. There can be a disconnect between companies' stated policies and their enforcement. This disconnect can be the result of inconsistent, discriminatory, arbitrary, or erroneous enforcement, and can either be targeted at or have an outsized impact on certain groups. Use of automated systems to flag or remove content can increase the potential for biased and erroneous enforcement.

---

[3] There are a variety of other content moderation activities, such as adding labels to content, but removing and ranking content and users are the primary ways in which major platforms moderate content.

Companies' selection and ranking of content is more complicated than content removal. Selection and ranking of content are automated through a platform's recommendation algorithm, which may have several components. Arvind Narayanan, *Understanding Social Media Recommendation Algorithms*, Knight First Amendment Institute 9 (Mar. 9, 2023).[4] On most platforms, the recommendation algorithm is primarily designed to optimize engagement— the probability that a user will interact with the content they are shown. *Id.* at 18–19. The output of the engagement algorithm reflects a company's prediction of user behavior, not the company's viewpoint on content. *Id.* at 22–24. Recommendation algorithms also typically include a component that increases or decreases the ranking or "amplification" of content, based on an assessment of the quality of the content or whether it is the type of content the company wants to promote. *Id.* at 24. Companies often use downranking to enforce vague or opaque policies, as opposed to concrete, public content guidelines or community standards. *See*

---

[4] https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.

Tarleton Gillespie, *Do Not Recommend? Reduction as a Form of Content Moderation*, Social Media & Society 8–9 (2022). For example, downranking is the main tool used for hard-to-define-but-harmful content and what social media companies call "borderline" content that comes close to violating the company's policies but doesn't qualify for removal. *Id.* at 3, 4–6. Unlike in the content removal context, users may find it difficult to know whether they have been shadowbanned or their content downranked.

Understanding how social media platform design affects what users see is difficult because social media platforms are hyper-individualized, complex, and built using tools that imperfectly effectuate the companies' content policies. *See* Daphne Keller, *Platform Transparency and the First Amendment*, 4 J. of Free Speech L. 1, 44 (2023) (comparing the difficulty of discerning platform decision making with the relative ease of inferring newspaper decision making). Meanwhile, the feeds themselves, and the policies that govern them, change frequently. *See id.* Access to internal data is thus necessary for users and researchers to infer the actual content moderation practices of a platform.

ii. Transparency is needed to illuminate the harms platform content moderation practices pose to users and society.

The scale, automated complexity, and hyper-personalization of social media platforms can harm users and society. Platforms can sometimes be toxic and unsafe environments, particularly for children and marginalized groups. *See, e.g.*, Georgia Wells et al., *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, Wall St. J. (Sep. 14, 2021).[5] They can spread misinformation and hate in a more rapid, targeted, and unexpected ways than other disseminators of information. *See, e.g.*, Sam Schechner et al., *How Facebook Hobbled Mark Zuckerberg's Bid to Get America Vaccinated*, Wall St. J. (Sep. 17, 2021).[6] Because these harms are directly tied to a company's content moderation practices, the public needs to understand these practices to determine how we, as individuals and as a society, can adapt to and prevent these harms.

---

[5] https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739.
[6] https://www.wsj.com/articles/facebook-mark-zuckerberg-vaccinated-11631880296.

Transparency is needed to understand how and whether platforms' policies and algorithmic design choices are amplifying the spread of harmful content. Some of the most significant harms to the public derive from social media's potential to amplify misinformation and hate speech. As described above, when companies primarily rank what to show users based on "engagement"—a calculation of how likely a user is to interact with a given piece of content—users tend to interact disproportionately with divisive, sensationalist, and provocative content. Narayanan, *supra*, at 18–19, 36; *see also* Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, Facebook (May 5, 2021).[7] The unfortunate result is that platform algorithms tend to spread misinformation, hate speech, and other harmful content. *See* Jeff Allen, *Misinformation Amplification Analysis and Tracking Dashboard*, Integrity Institute (Oct. 13, 2022);[8] Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier*

---

[7] https://www.facebook.com/notes/751449002072082/.
[8] https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard.

Instead, Wall St. J. (Sep. 15, 2021).[9] Transparency requirements help the public understand how algorithm design choices are impacting the spread of such content.

One way platforms' content moderation practices may cause harm is by subjecting users to a torrent of hate speech and abuse. Different platforms take different approaches to defining and setting thresholds for abusive behavior. The result can significantly impact the experiences of people with marginalized identities and even have carry-on effects into the real world. For example, targeted deadnaming and misgendering is a rising threat to prominent transgender people online because the content attracts engagement, which helps it go viral, and not all platforms have or fully enforce policies against such hate. Leanna Garfield & Jenni Olson, *All Social Media Platform Policies Should Recognize Targeted Misgendering and Deadnaming as Hate Speech*, GLAAD (Dec. 11, 2023).[10] Women and people of color who run

---

[9] https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215.

[10] https://glaad.org/social-media-platform-policies-targeted-misgendering-deadnaming-hate-speech/.

for office are more likely to experience hate on social media based on their gender and race, and this harassment makes them less likely to run for reelection or higher office. Gowri Ramachandran et al., *Intimidation of State and Local Officeholders*, Brennan Center for Justice (Jan. 25, 2024).[11] The emerging threat of deepfake nonconsensual intimate images could push more women out of the political arena and off social media entirely unless platforms develop robust policies and practices for removing this content. Vandinika Shukla, *Deepfakes and Elections: The Risk to Women's Political Participation*, Tech Policy Press (Feb. 29, 2024).[12]

Transparency can also unearth instances in which platforms provide specific users or groups special privileges or disadvantages by moderating content in a biased way. Facebook, for example, was found to have a secret policy exempting high-profile users from its content moderation policies. Jeff Horwitz, *Facebook Says Its Rules Apply to All.*

---

[11] https://www.brennancenter.org/our-work/research-reports/intimidation-state-and-local-officeholders.
[12] https://www.techpolicy.press/deepfakes-and-elections-the-risk-to-womens-political-participation/.

Company Documents Reveal a Secret Elite That's Exempt, Wall St. J. (Sep. 13, 2021).[13] Many groups have discovered that their speech was secretly being throttled on platforms, one notable example being the systematic silencing of Palestinians and advocates for Palestinian human rights on platforms like Facebook. Marwa Fatafta, *It's Not a Glitch: How Meta Systematically Censors Palestinian Voices*, Access Now (Feb. 19, 2024).[14] Informing users of these practices can help them decide where to speak and push for change.

Transparency can also inform the public about the trustworthiness of the information they see on social media. If a user mistakenly believes the content in their feed is trustworthy, they can be fooled into believing misinformation, which might lead them into taking actions that negatively impact their—and others'—health, wellbeing, and rights. Changes in a company's policies or practices can dramatically alter the trustworthiness of a content feed, as has occurred

---

[13] https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353.
[14] https://www.accessnow.org/publication/how-meta-censors-palestinian-voices/#flawed-content-moderation-policy.

on X since Elon Musk's takeover. Jim Rutenberg & Kate Conger, *Elon Musk Is Spreading Election Misinformation, but X's Fact Checkers Are Long Gone*, N.Y. Times (Jan. 25, 2024);[15] Vittoria Elliott, *Elon Musk's Main Tool for Fighting Disinformation on X Is Making the Problem Worse, Insiders Claim*, Wired (Oct. 17, 2023).[16]

Transparency about misinformation is especially important because such content, if spread widely, can have disastrous impacts on democracy, public health, and human rights. Misinformation about COVID vaccines flourished on platforms like Facebook, contributing to vaccine hesitancy. Schechner et al., *supra*. Amplification of misinformation and hate speech has also stoked violence, like the January 6th riots and the massacre of Rohingya in Myanmar. Craig Timberg et al., *Inside Facebook, Jan. 6 Violence Fueled Anger, Regret Over Missed Warning Signs*, Wash. Post (Oct. 22, 2021);[17] Amnesty

---

[15] https://www.nytimes.com/2024/01/25/us/politics/elon-musk-election-misinformation-x-twitter.html.
[16] https://www.wired.com/story/x-community-notes-disinformation/.
[17] https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/.

International, *The Social Atrocity: Meta and the Right to Remedy for the Rohingya* 45–46 (Sep. 29, 2022);[18] *see also* Newley Purnell & Jeff Horwitz, *Facebook Services Are Used to Spread Religious Hatred in India, Internal Documents Show*, Wall. St. J. (Oct. 23, 2021) (linking spread of misinformation with deadly religious riots in India);[19] Jamal Abukhater, *The Impact of Platform's Content Moderation Policies on Palestinian Digital Rights*, 7amleh Center, 6–7 (Mar. 14, 2024) (linking spread of misinformation and hate with incitement to violence in the West Bank and Gaza).[20]

Misinformation about elections is especially pernicious, undermining confidence in the election system, and can spread like wildfire on platforms. Paul M. Barrett, *Spreading the Big Lie: How Social Media Sites Have Amplified False Claims of U.S. Election Fraud*,

---

[18] https://www.amnesty.org/en/documents/ASA16/5933/2022/en/.
[19] https://www.wsj.com/articles/facebook-services-are-used-to-spread-religious-hatred-in-india-internal-documents-show-11635016354.
[20] https://7amleh.org/2024/03/14/new-position-paper-highlights-impacts-of-x-platform-s-content-moderation-policies-on-palestinian-digital-rights.

NYU Center for Business and Human Rights (Sept. 2022);[21] *see also* Cecilia Kang, *Five Unfounded Claims About Voting in the Midterm Election*, N.Y. Times (Nov. 2, 2022). Despite the threat, no platform provides access to information about the best-performing and most-engaged-with election-related content, accounts, and groups, which makes it difficult for users, researchers, and election officials to understand how such misinformation spreads on social media. Accountable Tech, *Democracy by Design: Social Media's Policy Scores* (Feb. 16, 2024).[22] Only two major platforms—TikTok and Snapchat—have published policies prohibiting deepfakes of public figures, *id.*, creating a significant risk that this content will spread on platforms and manipulate voters.

Leaks of internal documents from Facebook show that the platform knows that its automated tools have trouble detecting and controlling misinformation, hate speech, and other harmful content that

---

[21] https://bhr.stern.nyu.edu/tech-big-lie.
[22] https://accountabletech.org/research/democracy-by-design-social-medias-policy-scores/.

violates its policies. *See, e.g.*, Deepa Seetharaman et al., *Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts*, Wall St. J. (Oct. 17, 2021);[23] Schechner et al., *supra*. Part of the reason why platforms have trouble controlling the spread of harmful content is that automated detection of such content is hard. Misinformation, for instance, is designed to look like the truth, so it is difficult to detect. Esma Aïmeur et al., *Fake News, Disinformation and Misinformation in Social Media: A Review*, 13 Soc. Network Analysis and Mining 30 (2023).[24] But part of the explanation involves company priorities that are kept secret from the public. A recurring theme in the Facebook leaks is that leadership prefers permissive filters that err on the side of keeping content up, and they are also reluctant to sacrifice engagement to achieve content moderation goals. *See, e.g.*, Seethamaran et al., *supra*; Hagey & Horwitz, *supra*; Narayanan, *supra*, at 34–35.

---

[23] https://www.wsj.com/articles/facebook-ai-enforce-rules-engineers-doubtful-artificial-intelligence-11634338184.
[24] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9910783/.

Independent research is sorely needed on how harm proliferates on social media—and how to mitigate the harm—but it is difficult for researchers to conduct these studies without greater transparency from platforms into their algorithms and data. *See, e.g.*, Benjamin Kaiser & Jonathan Mayer, *It's the Algorithm: A Large-Scale Comparative Field Study of Misinformation Interventions*, Knight First Amendment Institute (Oct. 23, 2023) (calling for platforms to publish policies for content deamplification and the weighted contribution to the algorithm);[25] *see generally*, Barrett, *supra*; Narayanan, *supra*, at 22. Consistent, robust transparency is needed to ensure that the public is properly apprised of potential harm. But platforms are unwilling to provide it.

B. Adequate transparency cannot be accomplished through voluntary measures because platforms resist public scrutiny.

The public should be informed of the harms social media can cause to individuals and society. But much of what we currently know about

---

[25] https://knightcolumbia.org/content/its-the-algorithm-a-large-scale-comparative-field-study-of-misinformation-interventions.

platform harms comes from the Facebook leaks. *See, e.g.*, *The Facebook Files*, Wall St. J. (2021).[26] The public should not have to rely on occasional leaks for their understanding of platform-moderated speech.

Locked within social media companies is a wealth of information about how their platforms impact users and society. Companies are reluctant to allow the public a glimpse into their fortresses, and when they do, it is a highly curated view governed by overbearing terms and subject to the whims of company executives. Some companies—notably, X—have tried to silence research critical of their practices by shutting down researchers' access to their platforms and bringing burdensome lawsuits against them. Because platforms will not provide adequate transparency voluntarily, the government must demand it.

Platforms have provided limited transparency in response to public pressure. The first platform transparency reports were a response to civil society pressure to reveal the extent of government demands for user information. Trust & Safety Prof. Assoc., *History of*

---

[26] https://www.wsj.com/articles/the-facebook-files-11631713039.

Transparency Reports, Trust & Safety Curriculum.[27] Google published the first formal transparency report in 2010. *Id*. Soon, civil society broadened its attention to companies' own content policies and removal actions, and Etsy released the first policy enforcement report in 2014. *Id*.

Hoping to pressure companies to standardize and expand their enforcement reports, civil society groups published the Santa Clara Principles in 2018. *See* The Santa Clara Principles.[28] But the effort has been largely unsuccessful. Analysis shows that there continue to be large discrepancies in the information disclosed by platforms and that no companies' reports are in accordance with the Santa Clara Principles. Aleksandra Urman & Mykola Makhortykh, *How Transparent are Transparency Reports? Comparative Analysis of Transparency Reporting Across Online Platforms*, 47 Telecommunications Policy 3 (April 2023).[29] In fact, there has been a

---

[27] https://www.tspa.org/curriculum/ts-fundamentals/transparency-report/history-transparency-reports/ (last visited Mar. 20, 2024).
[28] https://santaclaraprinciples.org (last visited Mar. 20, 2024).
[29] https://doi.org/10.1016/j.telpol.2022.102477.

decline in platform transparency reporting in the last few years, notably X's abandonment of transparency reporting. The last content moderation transparency report that X voluntarily published was in April 2023. X, *An Update on Twitter Transparency Reporting* (Apr. 25, 2023).[30] The report was based on data from the first half of 2022, *id.*, and was much more abbreviated than the reports it published in 2021 and earlier. *Compare id.*, *with* X, *Rules Enforcement* (July 28, 2022).[31]

AB 587 pushes transparency reporting forward by imposing standards for format and types of information disclosed, something public pressure and voluntary measures have so far failed to achieve. The law also sets a baseline for transparency and prevents companies like X from abandoning disclosures altogether. AB 587 does not solve all transparency problems, and it need not do so. More transparency laws will be necessary in the future, but that does not mean that AB 587 does not serve an important government interest.

---

[30] https://blog.x.com/en_us/topics/company/2023/an-update-on-twitter-transparency-reporting.
[31] https://transparency.x.com/en/reports/rules-enforcement.html#2021-jul-dec.

Transparency reports contain useful information about companies' content moderation practices, but a more complete view into platform practices requires independent research. Such research requires access to granular data about the content on a service and users' interactions with that content. In response to public pressure, some platforms have provided researchers, journalists, and others with application programming interfaces ("APIs") through which they can access more granular data. *See* Emma Lurie, *Comparing Platform Research API Requirements*, Tech Policy Press (Mar. 22, 2023) (discussing differences among the APIs that TikTok, Meta, X, and YouTube provided at the time of writing).[32] But tech companies have a reputation of overpromising and underdelivering for research programs. For example, when civil society and government officials raised concerns about the lack of transparency into paid political ads on platforms, many major companies, such as Facebook, X, and Google, promised to provide research tools that would allow access to ad data. But when the tools

---

[32] https://www.techpolicy.press/comparing-platform-research-api-requirements/.

were finally rolled out, researchers found that Twitter and Google's databases were not comprehensive. For instance, Google's excluded all issue ads (as opposed to candidate ads), while Facebook's was so riddled with technical errors that it was practically unusable. Matthew Rosenberg, *Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised*, N.Y. Times (July 25, 2019).

In general, platform's data access programs are strictly limited and subject to the whims of executives. *See* Nathaniel Persily & Joshua A. Tucker, *How to Fix Social Media? Start with Independent Research*, Brookings (Dec. 1, 2021).[33] Platforms provide incomplete data sets through their APIs, either excluding key metrics like engagement or entire categories of content. *See* Rosenberg, *supra*; Craig Timberg, *Facebook Made Big Mistake in Data It Provided to Researchers, Undermining Academic Work*, Wash. Post (Sep. 10, 2021).[34] Most platform APIs also have rate limits and other restrictions that prevent

---

[33] https://www.brookings.edu/articles/how-to-fix-social-media-start-with-independent-research/.
[34] https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/.

researchers from collecting enough data for large-scale research. *See* Brittany I. Davidson et al., *Platform-Controlled Social Media APIs Threaten Open Science*, 7 Nature Human Behaviour 2054 (2023). API terms of service also limit the kind of research that can be done and impose procedural hurdles on publishing that are antithetical to independent research. *Id.*

Companies have ended data access programs and attacked independent researchers when they have realized the projects could reveal damaging information about the platform. For instance, when researchers deployed tools to help them collect information about political ads that Facebook promised—and failed—to deliver through its Ad Library, Facebook revoked the researchers' access to the platform and threatened legal action if they did not stop collecting data. Knight First Amendment Institute, *Researchers, NYU, Knight Institute Condemn Facebook's Effort to Squelch Independent Research about*

Misinformation, (Aug. 4, 2021);[35] Gilad Edelman, *Facebook's Reason for Banning Researchers Doesn't Hold Up*, Wired (Aug. 4, 2021);[36] Algorithm Watch, *AlgorithmWatch Forced to Shut Down Instagram Monitoring Project After Threats from Facebook* (Aug. 13, 2021).[37] Meta also moved to shut down an entire transparency tool, CrowdTangle, after researchers and journalists used it to show that right-wing commentators were spreading misinformation with great success. Kevin Roose, *Inside Facebook's Data Wars*, N.Y. Times (Oct. 4, 2021).[38] After winding down the tool for the last few years, Meta will take CrowdTangle completely offline in August 2024—without an adequate replacement and right before the U.S. (and other worldwide) elections, which will disrupt research into political activity on Facebook and Instagram at a critical time. Jeff Horwitz, *Meta to Replace Widely Used*

---

[35] https://knightcolumbia.org/content/researchers-nyu-knight-institute-condemn-facebooks-effort-to-squelch-independent-research-about-misinformation.
[36] https://www.wired.com/story/facebooks-reason-banning-researchers-doesnt-hold-up/.
[37] https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/.
[38] https://www.nytimes.com/2021/07/14/technology/facebook-data.html.

Tool—And Largely Cut Off Reporter Access, Wall St. J. (Mar. 14, 2024); *see also* Brandon Silverman, *CrowdTangle is Dead, Long Live CrowdTangle!*, Some Good Trouble Substack (Mar. 14, 2024).[39]

Recently, X has been particularly aggressive in silencing research into its platform. Before Elon Musk took over, Twitter had the most functional API of all major platforms, which also made it the go-to platform for social media research. Justine Calma, *Twitter Just Closed the Book on Academic Research*, The Verge (May 31, 2023). After Musk's takeover, newly-renamed X announced that it would end free access to its API and began charging up to $42,000 a year for API access. *Id*. The change essentially cut off researchers' access to X's API, resulting in more than 100 studies about X being canceled, suspended, or changed to focus on another platform. Sheila Dang, *Exclusive: Elon Musk's X Restructuring Curtails Disinformation Research, Spurs Legal Fears*, Reuters (Nov 6, 2023).[40]

---

[39] https://brandonsilverman.substack.com/p/crowdtangle-is-dead-long-live-crowdtangle.
[40] https://www.reuters.com/technology/elon-musks-x-restructuring-curtails-disinformation-research-spurs-legal-fears-2023-11-06/.

X is also using the legal system to silence researchers that publish findings critical of the platform. X is currently suing the Center for Countering Digital Hate for revealing that the company failed to remove numerous instances of disinformation that violated its content guidelines. *See X Corp. v. Center for Countering Digital Hate*, No. 23-cv-03836 (N.D. Cal. July 3, 2023); Vittoria Elliott, *Elon Musk's Lawsuit Against a Group That Found Hate Speech on X Isn't Going Well*, Wired (Mar. 1, 2024).[41] X also sued Media Matters after it reported that big name advertisers were being shown alongside antisemitic content. *See X Corp. v. Media Matters for America*, No. 23-cv-01175 (N.D. Cal. Nov. 20, 2023); Vittoria Elliott, *Elon Musk's Media Matters Lawsuit Will Have a 'Chilling Effect'*, Wired (Nov. 21, 2023).[42]

X's legal attack on AB 587 is part of the company's broader strategy to silence public discussion of the harm its platform might pose to users and society. Transparency laws are necessary precisely because companies like X are only transparent where it benefits them. The

---

[41] https://www.wired.com/story/elon-musk-lawsuit-hate-speech-x/.
[42] https://www.wired.com/story/x-elon-musk-media-matters-lawsuit/.

public needs to know more than just the platforms' positive spin to understand how content moderation impacts them and society.

There are positive changes on the horizon for platform transparency—not due to voluntary actions by platforms but because of new regulations like the European Union's Digital Services Act ("DSA"). To comply with Article 40.12 of the DSA, the largest platforms are developing tools to allow researchers access to real-time data, many for the first time ever. Silverman, *supra*. U.S. lawmakers should enact similar laws. But courts may foreclose this possibility by adopting the view of X and its *amici* on the constitutionality of platform transparency laws.

II. THE GOVERNMENT'S STRONG INTEREST IN CONTENT MODERATION TRANSPARENCY LAWS CAN ONLY BE OUTWEIGHED BY A SUBSTANTIAL, ACTUAL CHILLING EFFECT OR BURDEN ON PROTECTED SPEECH.

As described in Section I, the government and the public have a strong interest in access to information about social media companys' content moderation practices. Like other disclosure laws, content moderation transparency laws promote the core tenets of the First Amendment—by increasing access to information, they cultivate

informed decision-making and robust public debate. "Facts, after all, are the beginning point for much of the speech that is most essential to advance human knowledge and to conduct human affairs." *Bates v. State B. of Arizona*, 433 U.S. 350, 375 (1977).

Accordingly, and in line with precedent, constitutional scrutiny of content moderation transparency laws must weigh the pre-existing government interest in transparency against the degree of intrusion on protected First Amendment interests. Because of the strength of the government's interest in content moderation transparency, only an actual showing that the statute burdens or chills protected speech interests suffices, and X has not met that burden.

A. Challenges to transparency laws are subject to lesser scrutiny than laws that directly burden speech and generally require a showing that disclosure will actually burden protected speech.

Courts have long distinguished transparency laws from other types of speech regulations—and subjected them to more lenient scrutiny—because transparency laws only minimally infringe on First Amendment rights. Courts also require challengers to demonstrate a real burden on their protected interests.

Transparency laws "impose no ceiling on speech, and do not prevent anyone from speaking." *Citizens United v. Fed. Election Commn.*, 558 U.S. 310, 366 (2010) (cleaned up) (internal citations omitted). When compared to laws suppressing speech, "the First Amendment interests implicated by disclosure requirements are substantially weaker." *Zauderer*, 471 U.S. at 653 n.14. Unlike laws that mandate warning labels, disclaimers, or inclusivity in messaging, transparency laws do not compel or dictate a specific message. *See, e.g.*, *Natl. Assn. of Wheat Growers v. Bonta*, 85 F.4th 1263, 1267 (9th Cir. 2023); *Wooley v. Maynard*, 430 U.S. 705, 713 (1977); *Envtl. Def. Ctr., Inc. v. E.P.A.*, 344 F.3d 832, 849 (9th Cir. 2003); *Hurley v. Irish–American Gay, Lesbian and Bisexual Group of Boston, Inc.*, 515 U.S. 557, 573 (1995). There is a material difference between laws that mandate disclosures of facts, and compelled speech that "prescribe[s] what shall be orthodox in politics, nationalism, religion, or other matters of opinion or force citizens to confess by word or act their faith therein." *Zauderer*, 471 U.S. at 651 (quoting *W. Virginia State Bd. of Educ. v. Barnette*, 319 U.S. 624, 642 (1943)).

Because transparency laws embody the government's interest in promoting access to information, and the inherent intrusion into protected speech is comparatively minimal, successful challenges to transparency regulations require a showing of an actual burden on protected First Amendment rights resulting from disclosure.

Fifty years of precedent in this Court and the Supreme Court supports this contention. *Smith v. Helzer*, No. 22-35612, 2024 WL 1125095, at *7 (9th Cir. Mar. 15, 2024) (declining to consider "hypothetical" or "imaginary" cases to sustain a facial challenge to an elections disclosure law); *Chula Vista Citizens for Jobs and Fair Competition v. Norris*, 782 F.3d 520, 538 (9th Cir. 2015) (upholding election disclosure law under exacting scrutiny where "the actual burden imposed on First Amendment rights by the petition disclosure requirement is quite small."); *Fam. PAC v. McKenna*, 685 F.3d 800, 807 (9th Cir. 2012) (upholding law where challenger "made no showing that Washington's disclosure requirements expose contributors to a significant or systemic risk of harassment or retaliation"); *Doe v. Reed*, 561 U.S. 186 (2010) (finding "no reason to assume that any burdens imposed by disclosure of typical referendum petitions would be remotely

like the burdens plaintiffs fear in this case."); *Citizens United*, 558 U.S. at 367 (rejecting facial challenge to state election law where it was not shown that the law would impose an actual burden on association or political speech); *Buckley v. Valeo*, 424 U.S. 1, 68 (1976) (counseling courts to "look to the extent of the burden that they place on individual rights" to determine whether the government's interest in transparency outweighed its intrusiveness).

Successful challenges to disclosure laws are therefore generally supported by a real, not hypothetical, burden on speech. In *Americans for Prosperity Found. v. Bonta*, 141 S. Ct. 2373, 2388 (2021), for example, the court concluded that the challenged law was substantially overbroad where it "indiscriminately" included "the information of *every* major donor with reason to remain anonymous" and the plaintiffs "introduced evidence that they and their supporters have been subjected to bomb threats, protests, stalking, and physical violence." *Id.* Even further, the Court highlighted that "[t]he gravity of the privacy concerns in this context is further underscored by the filings of hundreds of organizations as *amici curiae* in support of the petitioners," which "[f]ar from representing uniquely sensitive causes . . . span the

ideological spectrum." *Id.* The Court ultimately concluded that the "disclosure requirement imposes a widespread burden on donors' associational rights" which, under exacting scrutiny, "cannot be justified on the ground that the regime is narrowly tailored to investigating charitable wrongdoing, or that the State's interest in administrative convenience is sufficiently important." *Id.* at 2389. A strong showing of burden on First Amendment interests was therefore necessary to outweigh even a baseline government interest in transparency.

B. Challenges to content moderation transparency laws must show an actual chilling effect resulting from disclosure.

The commercial context of content moderation transparency laws counsels for a permissive standard that balances any actual intrusion on First Amendment rights with the government's interests in content moderation transparency.

The standard used to evaluate content moderation transparency laws must reflect the fact that the government has a pre-established, substantial interest in "determining appropriate disclosure requirements for business corporations." *Pacific Gas & Electric Co. v.*

Public Utilities Commission, 475 U.S. 1, 16 n.12 (1986). In contrast, "[t]he right of a commercial speaker not to divulge accurate information regarding his services is not . . . a fundamental right." *Zauderer*, 471 U.S. at 651 n.14; *accord. Natl. Elec. Mfrs. Ass'n v. Sorrell*, 272 F.3d 104, 114–15 (2d Cir. 2001) ("[M]andating that commercial actors disclose commercial information ordinarily does not offend the important utilitarian and individual liberty interests that lie at the heart of the First Amendment.").

The *Zauderer* framework for evaluating commercial disclosure laws is accordingly the best fit. Under *Zauderer*, "the government may compel truthful disclosure in commercial speech as long as the compelled disclosure is reasonably related to a substantial governmental interest and involves purely factual and uncontroversial information that relates to the service or product provided," and does not impose an undue burden on speech. *CTIA - The Wireless Assn. v. Cty of Berkeley*, 928 F.3d 832, 842 (9th Cir. 2019).

The *Zauderer* Court's logic shows exactly why the standard rationally applies to content moderation transparency laws. In *Zauderer*, the court declined to strike down a statute requiring

attorneys to disclose in their advertisements that their clients may be required to pay their fees and costs. *Zauderer*, 471 U.S. at 651. The Court found that "[b]ecause the extension of First Amendment protection to commercial speech is justified principally by the value to consumers of the information such speech provides . . . appellant's constitutionally protected interest in not providing any particular factual information in his advertising is minimal." *Id.* (internal citations omitted).

This reasoning applies to transparency laws, like AB 587, that seek disclosure of true information about content moderation practices. Like advertisements, content moderation policies and practices are primarily commercial. They are the internet's "no shirt, no shoes, no service" signs, establishing rules and boundaries of social media companies' services, and the circumstances in which access to those services can be removed. In addition to the established government interest in the free flow of commercial information, the government and the public's specific interests in content moderation transparency are robust, as discussed in Section I. And because they shed light on activities that are already occurring, rather than limiting speech or

imposing a message, content moderation transparency statutes pose a minimal, if any, intrusion upon First Amendment interests. The same sliding-scale test enumerated in *Zauderer* therefore applies.

C. X's First Amendment interests in non-disclosure of its data are minimal.

X must make a showing of *some* actual burden on its First Amendment rights. X's burden to demonstrate some harm derives not just from the level of scrutiny applied by this Court, but also from the posture of this case: a request for a preliminary injunction on a facial First Amendment challenge requires a *prima facie* showing that First Amendment rights will be infringed in a substantial number of the statutes' applications when judged against the law's plainly legitimate sweep. *See Smith*, 2024 WL 1125095, at *4. Assuming that content moderation is protected speech, X might have demonstrated a burden to its speech interests if it had showed that AB 587 chilled its content moderation activities or forced it to change its message. X has not made this showing.

One way a disclosure law might burden speech is by imposing such an enormous compliance cost that companies are compelled to

change their content moderation goals to reflect a specific message. *See, e.g.*, *NetChoice, LLC v. Atty. Gen., Fla.*, 34 F.4th 1196, 1230 (11th Cir. 2022), *cert. granted in part sub nom. Moody v. NetChoice, LLC*, 144 S. Ct. 478 (2023) (finding a chilling effect where compliance with a platform transparency law imposes such substantial costs on content moderation that it compels companies to do less of it, with the effect that the platforms spread more of the government's preferred speech).

X has not alleged that AB 587's mandated disclosures are an impossible administrative burden. X is already complying with AB 587 and there is no evidence that it has changed its content moderation practices in the process. In its Q3 2023 report, X only disclosed definitions, policies, and figures for the categories of speech that it already included and tracked in its public-facing policies and disclosures to the European Union.[43] For four out of six of the requested categories

---

[43] Compare X, Q3 2023 TOS Report, available at: https://oag.ca.gov/sites/default/files/X%20Corp.%20-%20Q3%202023%20California%20TOS%20Report.pdf/X%20Corp.%20-%20Q3%202023%20California%20TOS%20Report.pdf, with The X Rules, available at https://help.twitter.com/en/rules-and-policies/x-rules,

of definitions, X simply stated that it did not have a policy or definition regarding these types of content. For example, for "extremism or radicalization," X stated that "[t]he current version of the X Terms of Service does not define 'extremism' or 'radicalization.' X has policies that address 'Violent & Hateful Entities' and 'Hateful Conduct,' among other policies." *See* X, Q3 2023 TOS Report, at 1-2. It made similar statements for hate speech or racism, foreign political interference, and controlled substances distributions, demonstrating that X has not made any modifications to its practices as a result of this law. *See id.* A perusal through other social media companies' disclosures show similar results: the disclosures reflect the current state of their practices, which have not substantially changed as a result of AB 587.[44]

Although X interprets AB 587 to express government preferences for specific categories of content moderation, and to threaten undue enforcement actions or burdens should it not moderate this content, this

---

and X, DSA Transparency Report, available at https://transparency.twitter.com/dsa-transparency-report.html (last visited March 20, 2024).

[44] https://oag.ca.gov/ab587/submissions.

subjective interpretation is not clear on the face of the law, nor in any evidence of future application. *See Smith*, 2024 WL 1125095 at *4 (noting that "the Supreme Court has cautioned courts 'not to go beyond the [regulations'] facial requirements and speculate about "hypothetical" or "imaginary" cases.'"). Where X cannot even show that, as applied to its own disclosures, its right to editorial discretion has been infringed upon, it is unlikely that it can show that a substantial number of applications will do so.

Another burden disclosure laws may impose is a deterrent effect on association and other First Amendment rights due to public reactions to the disclosures. Making this showing requires an actual likelihood of threats, harms, or harassment resulting from disclosure. *Compare, e.g.*, *Reed*, 561 U.S. at 200 (upholding regulation that requiring disclosure of petition signatories due to low likelihood of reprisal); *with NAACP v. Alabama*, 357 U.S. 449, 462 (1958) (striking down disclosure law where the organization had "made an uncontroverted showing that on past occasions revelation of the identity of its rank-and-file members [had] exposed these members to economic

reprisal, loss of employment, threat of physical coercion, and other manifestations of public hostility.").

X has not shown evidence of threats that might cause a chilling effect on association or other protected speech. For instance, X has not shown that, after releasing its transparency report, it has been subjected to any threats, harassment, or even negative public reactions. *See Reed*, 561 U.S. at 201 (finding relevant that post-disclosure, none of the alleged harms came to fruition); *accord. Citizens United*, 558 U.S. at 370 ("Citizens United has been disclosing its donors for years and has identified no instance of harassment or retaliation"). Its assertion that its content moderators have been "harassed, abused, and attacked" in the past for their involvement in content moderation decisions, *see* Appellant's Br., at 12., does not demonstrate that its moderators will be subject to *more* of this conduct as a result of complying with AB 587, nor does it show that the company will *change its practices* as a result of any such conduct, and is therefore an insufficient showing.

Public scrutiny of business practices is a *benefit*, not a harm. As a business, X runs the risk that its commercial decisions will impact its revenue and market share. The First Amendment cannot shield it from

market forces. X does not have a protected First Amendment right to prevent consumers from making informed decisions about the platforms that they use.

The final harm that X alleges boils down to a generalized fear that the California attorney general will abuse the enforcement provisions of this statute to impermissibly infringe on X's editorial judgment. But the hypothetical risk of abuse of a regulation has never been considered a cognizable burden on speech. Any regulation can be misused. As the Second Circuit aptly put it,

> [T]otalitarian tendencies do not lurk behind every instance of a state's collection of information about those within its jurisdiction. Any form of disclosure-based regulation—indeed, any regulation at all—comes with some risk of abuse. This background risk does not alone present constitutional problems.

Citizens United v. Schneiderman, 882 F.3d 374, 383 (2d Cir. 2018).

The proper forum for addressing potential future misuse of a statute is in an as-applied challenge when the misuse becomes a real threat. For example, X itself has previously sued Texas Attorney General Ken Paxton for issuing civil investigative demands under a

consumer protection statute in what X perceived as retaliation for the platforms' content moderation practices. *See Twitter, Inc. v. Paxton*, 56 F.4th 1170 (9th Cir. 2022). In that case, X properly made a pre-enforcement challenge to the order, rather than contending that the statute was unconstitutional because of the possibility that it could be used to chill speech. *See id.*; *see also*, Ramya Krishnan, *How the Supreme Court Could Encourage Platform Transparency*, Slate (Jan. 9, 2023) ("[P]rivacy laws, tax laws, employment discrimination laws, and SEC disclosure laws … [can] be abused to retaliate against companies for editorial decisions government officials don't like, and they sometimes are. That doesn't render them categorically unconstitutional.").[3]

The only evidence that X offers to support the claim that the AG will misuse AB 587 is a November 2022 letter from the California Attorney General sent to five large social media platforms, including X. *See* 6-ER-1067–1070. AB 587 is mentioned only once, among four other laws that more directly require disclosures of elections-related spending and prohibit interference with elections. *See id.* And commensurate with the text of AB 587, the letter merely urges the companies to continue to

take action pursuant to *their own pre-existing policies*, not to adopt the government's preferred policies. *Id.*

Because there is no evidence that the Attorney General has used its investigative powers under AB 587 to bully X—or any other company—into moderating the government's preferred content, nor is there evidence of any threats that could lead this Court to assume that such abuse would occur in the future, this hypothetical intrusion is more of a fringe case than evidence of a substantial number of likely unconstitutional applications. Should X's fears be realized, and there arises a "reasonable probability" the Attorney General will leverage AB 587 to require the company to moderate content in the government's preferred way, an as-applied challenge remains available.

CONCLUSION

For the foregoing reasons, *amicus* respectfully urges the Court to affirm the District Court's order denying Plaintiff's motion for a preliminary injunction.

Date: March 20, 2024

/s/ Megan Iorio
Megan Iorio
Schuyler Standley
ELECTRONIC PRIVACY INFORMATION CENTER
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140

Attorneys for Amicus Curiae Electronic Privacy Information Center

CERTIFICATE OF COMPLIANCE

I am the attorney or self-represented party.

This brief contains 6,967 words, excluding the items exempted by Fed. R. App. P. 32(f). The brief's type size and typeface comply with Fed. R. App. P. 32(a)(5) and (6).

I certify that this brief *(select only one)*:

[ ] complies with the word limit of Cir. R. 32-1.

[ ] is a **cross-appeal** brief and complies with the word limit of Cir. R. 28.1-1.

[X] is an **amicus** brief and complies with the word limit of Fed. R. App. P. 29(a)(5), Cir. R. 29-2(c)(2), or Cir. R. 29-2(c)(3).

[ ] is for a **death penalty** case and complies with the word limit of Cir. R. 32-4.

[ ] complies with the longer length limit permitted by Cir. R. 32-2(b) because *(select only one)*:

[ ] it is a joint brief submitted by separately represented parties;

[ ] a party or parties are filing a single brief in response to multiple briefs; or

[ ] a party or parties are filing a single brief in response to a longer joint brief.

[ ] complies with the length limit designated by court order dated _____.

[ ] is accompanied by a motion to file a longer brief pursuant to Cir. R. 32-2(a).

Signature: */s/ Megan Iorio* **Date:** March 20, 2024

CERTIFICATE OF SERVICE

I certify that on March 20, 2024, this brief was e-filed through the ACMS System of the U.S. Court of Appeals for the Ninth Circuit. I certify that all participants in the case are registered ACMS users and that service will be accomplished by the ACMS system.

Date: March 20, 2024

/s/ Megan Iorio
Megan Iorio
Schuyler Standley
ELECTRONIC PRIVACY
INFORMATION CENTER
1519 New Hampshire Ave. NW
Washington, DC 20036
(202) 483-1140

Attorneys for Amicus Curiae
Electronic Privacy Information Center