ROB BONTA
Attorney General of California
TODD GRABARSKY
LARA HADDAD
Supervising Deputy Attorneys General
ASHA ALBUQUERQUE (SBN 332901)
WILLIAM BELLAMY (SBN 347029)
SHIWON CHOE (SBN 320041)
CHRISTOPHER J. KISSEL (SBN 333937)
SAMUEL SOKOLSKY (SBN 348173)
MEGHAN H. STRONG (SBN 324503)
ZELDA VASSAR (SBN 313789)
Deputy Attorneys General
455 Golden Gate Avenue, Suite 11000
San Francisco, CA 94102-7004
Telephone: (415) 510-4400
Fax: (415) 703-5480
E-mail: Asha.Albuquerque@doj.ca.gov
William.Bellamy@doj.ca.gov
Shiwon.Choe@doj.ca.gov
Christopher.Kissel@doj.ca.gov
Samuel.Sokolsky@doj.ca.gov
Meghan.Strong@doj.ca.gov
Zelda.Vassar@doj.ca.gov
Attorneys for Defendant Rob Bonta, in his official capacity as Attorney General of California

IN THE UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

TIKTOK INC,

Plaintiff,

v.

ROB BONTA, in his official capacity as Attorney General of California,

Defendant.

Case No. 5:25-cv-09789-EJD

DEFENDANT'S OMNIBUS OPPOSITION TO PLAINTIFFS' MOTIONS FOR PRELIMINARY INJUNCTION

Date/Time: May 7, 2026 at 9:00 a.m.
Courtroom: 4 (Fifth Floor)

META PLATFORMS, INC., Plaintiff, **v.** **ROB BONTA, in his official capacity as Attorney General of California,** Defendant.	Case No. 5:25-cv-09792-EJD **DEFENDANT'S OMNIBUS OPPOSITION TO PLAINTIFFS' MOTIONS FOR PRELIMINARY INJUNCTION** Date/Time: May 7, 2026 at 9:00 a.m. Courtroom: 4 (Fifth Floor)
GOOGLE LLC and YOUTUBE, LLC, Plaintiffs, **v.** **ROB BONTA, in his official capacity as Attorney General of California,** Defendant.	Case No. 5:25-cv-09795-EJD **DEFENDANT'S OMNIBUS OPPOSITION TO PLAINTIFFS' MOTIONS FOR PRELIMINARY INJUNCTION** Date/Time: May 7, 2026 at 9:00 a.m. Courtroom: 4 (Fifth Floor)

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Statutes

Other Authorities

INTRODUCTION

California enacted the Protecting Our Kids from Social Media Addiction Act (Senate Bill No. 976, or SB976) with the goal of protecting the mental and physical health of children. SB976 was inspired by a body of research showing that the features internet companies employ to keep people maximally "engaged" on their platforms are fueling an epidemic of compulsive and excessive internet use among children. This epidemic is causing real harm. Children who spend excessive time online experience negative self-image, poor sleep, and anxiety and depression, among other health and behavioral issues. Children report feeling addicted to their screens—they want to turn their attention to other things but are unable to do so. An abundance of scientific, reliable, peer-reviewed research confirms that the mechanisms employed by internet platforms have habit-forming effects. There is good news, however: these same reliable studies have found that limiting exposure to these addictive mechanisms improves users' well-being and self-image, leads to better sleep and better self-regulation, and decreases depression and anxiety.

But Plaintiffs Meta Platforms, Inc., which owns Facebook, Instagram, and Threads ("Meta"); TikTok Inc. ("TikTok"); and Google LLC and its wholly-owned subsidiary YouTube, LLC (collectively, "YouTube") have each filed lawsuits, seeking to block certain provisions of SB976 from being applied to them: namely, §§ 27000.5, 27001, and 27002(b)(2) and (4), which regulate "addictive feeds" (or, as Plaintiffs dub them, "personalized feeds") with respect to children.[1] Addictive feeds are endless scrolls of user-posted material ("posts") that internet platforms like Plaintiffs compile and personalize into "feeds" for each individual user, based on the information the platform has tracked and collected about that user (often measuring into the millions of data points), so as to keep that user maximally engaged and spending as much time

[1] Section 27000.5 defines "addictive feeds." Section 27001 requires platforms to obtain parental consent to provide addictive feeds to children. Section 27002(b)(2) requires platforms to default to limiting children's access to addictive feeds to one hour per day unless modified by their parents. Section 27002(b)(4) requires platforms to offer parents a mechanism to be able to set that the default feed provided to their children will be a non-addictive feed. All statutory citations are to California's Health and Safety Code, except where otherwise indicated.

on the platform as possible. The reason Plaintiffs try to maximize user engagement is simple: Plaintiffs make money from advertisers paying them to target their users with ads. The more "engaged" that users—including children—are, the more ad space Plaintiffs can sell, and the more money they make. Essentially, Plaintiffs monetize children's time and attention for profit.

To that end, Plaintiffs have designed their platforms to be addictive. Public statements and internal documents confirm Plaintiffs' knowledge of and discussions about the addictive nature of their platforms—including, e.g., how they were designed to "consume as much of your time and conscious attention as possible" (Facebook), how they are "a drug... We're basically pushers" (Instagram), how youth addiction is viewed internally as "a good thing for the platform" (TikTok), and how engineers "aspire[d] to create an app that is...Addictive[:] Our app experience should compel users to come back more and more often" (YouTube). Ex. 1 at 21–22, 86, 162.[2]

In order to compile their addictive feeds, each Plaintiff employs machine-learning computer algorithms that, for each individual user, select and display those posts that the algorithms "think" will most hook that user and maximize that user's "engagement" on Plaintiffs' platforms. Critically, no actual human being reviews, selects, or exercises editorial judgment over the compilation in any user's personalized feed. Instead, these feed-personalization algorithms take the millions or billions of data points they have tracked and harvested about each user and use them to calculate, as numeric scores, the likelihood that new posts will engage the user, and then compile their feeds for the user based on those scores. These algorithms have no understanding of the actual content or substance of the posts they are compiling or the feeds they are generating.

[2] Exhibit 1 is a copy of the partially redacted Omnibus Opposition to Defendants' Motions for Summary Judgment that was filed in *In re Social Media Adolescent Addiction/Personal Injury Products Liability Litigation*, No. 4:22-md-03047-YGR (PHK) (N.D. Cal.) ("MDL MSJ Opposition"). Defendant cites the Opposition here because it has been publicly filed and thus is not subject to any confidentiality restrictions. Where possible, Defendant has also attached the underlying exhibits (filed provisionally under seal) for the Court and, where pertinent, cites to those exhibits in this brief. (During the course of discovery, Plaintiffs refused to produce many of the Opposition exhibits, so in some cases Defendant does not have access to exhibits here.)

And no actual human being understands how the algorithms operate or why the algorithms select the posts they select for any given user's feed (the so-called "black box" phenomenon). In many instances, the algorithms—in pursuit of their overarching goal to maximize user engagement—send children down "rabbit holes" of feeds of increasingly extreme posts that the algorithms "think" those children will engage with more, such as posts encouraging suicide and self-harm. They do this not because any human being at Plaintiffs' companies necessarily has made an editorial choice to promote suicide or self-harm. They do this precisely because human beings are *not* the ones deciding how feeds are personalized. Instead, the algorithms are the ones personalizing the feeds on their own, in pursuit of the companies' goal—maximizing user "engagement" and the time users spend on Plaintiffs' platforms—without having any understanding of the actual content of the feeds that they are presenting to users and children.

Plaintiffs have each filed motions for preliminary injunction ("MPIs") to enjoin §§ 27000.5, 27001, and 27002(b)(2) and (4) from being applied to them. Plaintiffs' MPIs should be denied.

First, Plaintiffs have not established a likelihood of success on the merits of their pre-enforcement as-applied First Amendment claims. As this Court has held, "[t]he touchstone of First Amendment speech rights is [] the protection of *expression*." *NetChoice v. Bonta*, 761 F. Supp. 3d 1202, 1219 (N.D. Cal. 2024) (*NetChoice I*) (emphasis in original), *aff'd in part and rev'd in part*, 152 F.4th 1002 (9th Cir. 2025) (*NetChoice II*). Plaintiffs' algorithmically-generated feeds are not expressive. They do not reflect the expressive choices of any actual human being—in fact, the humans at Plaintiffs' companies cannot predict what the algorithms will "say," and the algorithms often promote user posts that their human creators desire *not* to promote. As this Court has observed, the idea that "an algorithm to maximize engagement, i.e., the time spent on a social media platform" or one that "amplifies messages that its creator expressly disagrees with" nonetheless "implements some expressive choice and conveys its creator's message should be met with great skepticism." *NetChoice I*, 761 F. Supp. 3d at 1222; *see id.* at 1221–23. Nor do these algorithmic feeds convey an expressive message of Plaintiffs' own—they just reflect back users what the algorithms "think" users will be interested in. As the Ninth Circuit affirmed, "an algorithm that 'respond[s] solely to how users act online,' merely 'giving them the content they

appear to want,' probably is not expressive." *NetChoice II*, 152 F.4th at 1014 (quoting *Moody v. NetChoice, LLC*, 603 U.S. 707, 736 n.5 (2024); *id.* at 746 (Barrett, J., concurring)).

Additionally, even if Plaintiffs could establish that SB976 as applied to them burdens expressive conduct protected by the First Amendment, at most intermediate scrutiny would apply, because SB976 is content-neutral: it regulates only digital mechanisms, not topics, ideas, or viewpoints. And SB976 satisfies intermediate scrutiny. Indeed, SB976 is constitutional even under strict scrutiny, since it pursues an undeniably compelling interest—the health of children—while narrowly tailoring its restrictions to the mechanisms that cause children harm.

Second, Plaintiffs have not established irreparable harm. Plaintiffs do not proffer any idea, thought, or opinion that SB976 prevents them from expressing. At most, SB976 here restricts the output of computerized algorithms. But algorithms are not irreparably harmed by restrictions on their output, and Plaintiffs have not established irreparable harm to any actual human being.

Third, Plaintiffs have not established that equities and public interest weigh in their favor. Plaintiffs' addictive-feed mechanisms are inflicting real harm on children every day. Allowing SB976 to remain in effect will alleviate these, while leaving unrestricted (1) Plaintiffs' ability to express their own viewpoints and (2) users' ability to access any online content of their choice.

The Court should deny Plaintiffs' MPIs.

BACKGROUND

I. PLAINTIFFS EMPLOY COMPUTERIZED MACHINE-LEARNING ALGORITHMS TO COMPILE ADDICTIVE, PERSONALIZED FEEDS TO KEEP USERS—INCLUDING CHILDREN—ON PLAINTIFFS' PLATFORMS FOR AS LONG AS POSSIBLE

A. Plaintiffs Designed Their Feeds to Be Addictive in Order to Maximize Profits

Platforms like Facebook, Instagram, Threads, TikTok, and YouTube monetize users' time online through data collection and targeted advertising. Radesky Decl. ¶ 75; Hilbert Decl. ¶¶ 14–15. The more "engaged" a user is on Plaintiffs' platforms and the longer the user spends there, the more ads can be sold for that user, and thus the more money Plaintiffs will make. Egelman Decl. ¶¶ 17–20. *See generally* Ex. 1 at 12–21, 77–86, 150–62; *accord, e.g.*, Ex. M11 at 194:9–195:12 (Meta CEO Mark Zuckerberg agreeing that "[redacted]" and thus "[redacted]

[redacted]"); Ex. M8 at 23:10–24:7, 93:20–94:4, 246:14–249:12, 305:15–306:22; Ex. M10 at 31:16–32:14; Ex. M12 at 31:1–32:23; Ex. M13 at 104:4–13; Ex. T427 at 123:11–22, 125:3–7 134:23–136:16 (TikTok's algorithm team head Zhu Wenjia agreeing that "[redacted]" and "[redacted]"); Ex. T541 at 0547, 9176[3]; Ex. T632 at 45:7–15; Ex. Y715 at 337:13–339:6, 351:8–352:11 (YouTube Vice President of Engineering Christos Goodrow agreeing that "[redacted]").

In pursuit of their goal to have users maximally engaged, Plaintiffs have designed their platforms to be addictive. *See generally* Ex. 1 at 21–33, 36–43, 86–93, 162–65, 168–69. For example, Sean Parker, the second president of Facebook (now Meta), discussed how:

> The thought process that went into building these applications, Facebook being the first of them, ... was all about: "How do we consume as much of your time and conscious attention as possible?" And that means that we need to sort of give you a little dopamine hit every once in a while, because someone liked or commented on a photo or a post or whatever. And that's going to get you to contribute more content, and that's going to get you ... more likes and comments. It's a social-validation feedback loop ... exactly the kind of thing that a hacker like myself would come up with, because you're exploiting a vulnerability in human psychology. The inventors, creators—it's me, it's [Facebook CEO] Mark [Zuckerberg], it's [Instagram co-founder] Kevin Systrom on Instagram, it's all of these people—understood this consciously. And we did it anyway.

Mike Allen, *Sean Parker unloads on Facebook: "God only knows what it's doing to our children's brains"*, AXIOS, Nov. 9, 2017 (attached as Ex. 3). As Meta researchers discussed:

> Oh my gosh yall IG [Instagram] is a drug... We're basically pushers... We are causing Reward Deficit Disorder bc people are binging on IG so much they can't feel reward anymore...like their reward tolerance is so high...I know [Head of Instagram] Adam [Mosseri] doesn't want to hear it—he freaked out when I talked about dopamine in my teen fundamentals leads review but its undeniable! It[']s biological and psychological....the top down directives drive it all towards making sure people keep coming back for more. That would be fine if its productive but most of the time it isn't...the majority is just mindless scrolling and ads.

Ex. 1 at 21; Ex. M74 at 2152. As former Meta "Responsible AI" researcher Joshua Simons, Ph.D., testified, "given that the whole News Feed system is aimed at engagement, that's a set of repeated patterns of behavior—clicking, liking, sharing, and so on—the best way to increase

[3] Non-deposition bates-stamped documents are pincited by the last four digits of those stamps.

engagement is addiction in some way[.]" Ex. 1 at 37; Ex. M13 at 67:11–17.

Internal TikTok documents confirm that TikTok was designed to be addictive as well. As an internal "Product Experience Report" put it, while parent concerns about youth addiction were "concerning for young minds," this phenomenon was viewed internally as "a good thing for the platform." Ex. 1 at 86; Ex. T532 at 7593. Strategy documents confirm that "the advertising-based business model encourages optimization for time spent in the app," Ex. 1 at 78, Ex. T541 at 9176, that "[redacted]" and that "[redacted]" Ex. T537 at 3044, Ex. T556 at 1514.

Internal YouTube documents confirm that YouTube was designed to be addictive as well. For example, an internal slide deck titled "[redacted]" discussed how YouTube "aspire[d] to create an app that is...Addictive[:] Our app experience should compel users to come back more and more often[.]" Ex. 1 at 162; Ex. Y765 at 1526; *see* Ex. Y715 at 378:9–379:9 (identifying participants). Other documents were equally candid, discussing how YouTube's "[redacted]" to "[redacted]" and to "[i]ncrease habitual users" and "focus on making YouTube a daily habit [redacted]" with the "aim of increasing frequency and long-term YouTube Time." Ex. 1 at 162; Ex. Y701 at 1235–38, 1255 (emphasis in original).

Plaintiffs' platforms are not just addictive from a lay perspective—they satisfy medical and clinical standards for addictiveness. *See* Christakis Decl. ¶¶ 13, 116–26; Radesky Decl. ¶¶ 48, 50; Telzer Decl. ¶¶ 68, 157–60. MRIs have indicated altered brain function among individuals addicted to social media, similar to that of individuals addicted to drugs or alcohol. Telzer Decl.

¶ 173; Twenge Decl. ¶ 152; Christakis Decl. ¶ 129. The addictive nature of Plaintiffs' platforms, and the harm they cause, are discussed further below. *See infra* pp. 20–25.

B. Plaintiffs Employ Computerized Machine-Learning "Black Box" Algorithms to Track and Harvest Millions of Data Points on Each Individual User and Use Them to Compile Personalized Feeds That Are Designed to Be Addictive

One of the key mechanisms by which Plaintiffs achieve their platforms' addictive effects is through the use of addictive personalized feeds, compiled by computerized machine-learning algorithms. These algorithms track and harvest millions of data points on each of their users, e.g., which posts the user has most engaged with in the past, and then use that harvested data to try to calculate which other posts will most engage the user and keep the user on Plaintiffs' platform longer—so that Plaintiffs can sell more ads and monetize that user's time and attention. These computerized algorithms have no understanding of the actual content of the feeds they are compiling or the posts that go into them. And they are so immensely complex that humans (even their own programmers) cannot understand or explain how they make their decisions or predict with any certainty what they might do or what posts they might display in any given user's feed.

1. Plaintiffs Choose to Make Their Feeds Addictive to Increase Their Profits

Platforms can compile feeds in non-addictive ways, such as sorting posts from friends and family chronologically. Golbeck Decl. ¶ 22; Hilbert Decl. ¶ 38; Egelman Decl. ¶ 30. Users want non-addictive feeds. For example, a Meta study revealed how "[r]epeatedly over the past year . . . we've heard teens complain bitterly about Instagram's nonchronological feed.... Many say that if they could change one thing about IG, this is what they would change. Although the non-chronological feed fuels increased time spent on the site, teens say it makes it more difficult for them to find relevant info about their closer friends." Ex. 1 at 42; Ex. M280 at 0439.

Plaintiffs realized, however, that tracking, harvesting, and utilizing users' personal data to create personalized feeds would keep users on their platforms for longer, which would be far more profitable. *See* Radesky Decl. ¶¶ 75–83; Egelman Decl. ¶¶ 17–20; Hilbert Decl. ¶¶ 53–54. For example, internal Meta documents discussed how Instagram's switch to a personalized feed led users "to spend 40% more time in the app each day." Ex. 1 at 35; Ex. M212 at 8565; *see also* Ex. 1 at 35 (discussing how feed personalization "show[s] significant increases in consumption

and time spent," "increas[es] the number of sessions and time spent when compared to chronological feed," and "drives revenue! I've never seen anything like it!"); Ex. M213 at slide 19; Ex. M217 at 9529; Ex. M219 at slide 5. Internal YouTube documents likewise discussed how [redacted]. *See* Ex. Y702 at 185:8–17; *accord, e.g.*, Ex. Y715 at 160:5–161:13 ([redacted]). TikTok at one point considered a "non personalized feed" option to reduce harm to users—but implemented it only in Europe, because "ad revenues [would] decrease" if a non-personalized feed option were offered in the United States. Ex. 1 at 107; Ex. T519 at 1580; *see also* Telzer Decl. ¶ 62.

2. How Plaintiffs' Machine-Learning Feed-Personalization Algorithms Work

No Plaintiff compiles its addictive personalized feeds through human beings reviewing the pool of user posts and material available on Plaintiffs' platforms or human beings selecting which posts should be displayed in which user's feeds. Instead, as Plaintiffs freely acknowledge, they employ computer algorithms to compile their addictive personalized feeds and select the posts that appear in each user's individual feeds. Meta MPI 11; TikTok MPI 2; YouTube MPI 4–5.

Plaintiffs' descriptions of their algorithms in their MPIs are cursory. *Cf. NetChoice I*, 761 F. Supp. 3d at 1223 ("[M]uch of the First Amendment analysis depends on a close inspection of how regulated feeds actually function."); *NetChoice II*, 152 F.4th at 1022 (when it comes to "recommendation algorithms or how they work, . . . record development matters"). And Plaintiffs have not disclosed the complete details of how their algorithms work.[4] Nevertheless, the record to date reveals at least five aspects that explain how Plaintiffs' feed-personalization algorithms operate and how they compile and present addictive feeds to users:

a. humans do not write the rules that govern how feeds are personalized; instead, the algorithms write, and constantly rewrite, their own rules;

b. the algorithms have no understanding of the content of user posts or the feeds they are compiling; instead, all they can measure, and try to maximize, are numeric metrics like the amount of time users spend on Plaintiffs' platforms;

[4] For example, Plaintiffs specifically refused to produce to Defendant expert reports filed with the MDL MSJ Opposition (as Exhibits 985 and 986) discussing how their algorithms work.

c. not only do humans not write the rules that govern how feeds are personalized, humans do not even *know* the rules, and humans cannot explain how the algorithms operate or why the algorithms select the posts they do;

d. humans do not know and cannot predict the algorithms' output, i.e., what the algorithms will "say" or what the contents of any user's feed will be; indeed, the algorithms often amplify messages that their human creators want to *not* convey, e.g., suicide or self-harm posts; and

e. feed personalization is separate and distinct from content moderation.

a. Humans Do Not Write the Rules That Govern How Plaintiffs' Feeds Are Personalized; Instead, the Algorithms Write Their Own Rules

In the historical "traditional paradigm," human programmers wrote precise instructions for computers to follow and carry out, based on predefined logic and human-defined steps (e.g., "if a post contains 'X', delete it"). Hilbert Decl. ¶ 27. The human programmers knew what the rules were and had certainty over how the algorithms would process data to achieve a desired output. *Id.* This traditional paradigm is often referred to as "Good Old-Fashioned AI" or "GOFAI." *Id.*

Machine learning turns this paradigm on its head. *Id.* ¶¶ 28–29. With machine-learning algorithms, humans do *not* write or define the rules that algorithms follow or carry out. *Id.* ¶ 29; Golbeck Decl. ¶¶ 27, 30; *see also, e.g.*, Mackenzie Austin & Max Levy, *Speech Certainty*, 77 STAN. L. REV. 1, 48 (2025) (quoting former Google head of search and AI describing how "[b]y building [machine] learning systems, we don't have to write these rules anymore"). Instead, human developers design a model to optimize for a particular goal—e.g., maximizing the time users stay on a platform—and the algorithms themselves calculate how to achieve that goal. Golbeck Decl. ¶¶ 30, 39; Hilbert Decl. ¶ 29. In effect, a machine-learning algorithm is its own "master": "**the algorithm to compute the best algorithm**." Hilbert Decl. ¶ 29. "[T]he recommendations made by the algorithmic production worker are expressions of this socially embedded algorithm in this very moment, not of any person or anything else, including the programmers who coded the underlying machine learning logic in the first place or the executives who deploy them in the wild." *Id.* ¶¶ 10, 82–91; *see* Golbeck Decl. ¶¶ 47–49; Austin & Levy 48.

Plaintiffs' feed-personalization algorithms are provided large datasets of user behavior and, from this data, learn to identify statistical relationships between users, posts, and outcomes of interest. Golbeck Decl. ¶¶ 27, 30. These datasets can contain millions or billions of individual

data points. *Id.* ¶ 32; Arvind Narayanan, *Understanding Social Media Recommendation Algorithms* 23 (2023) (attached as Ex. 4) ("If a user spends an hour a day on TikTok for four years, the average video length is 20 seconds, and they skip half the videos, the platform has interaction records on over *half a million* videos for that single user.") (emphasis in original); *accord, e.g.*, YouTube MPI Goodrow Decl. ¶ 11 (YouTube Recommender System "is based on billions of signals"). The datasets can range from extremely narrow measurements ("How many users have clicked on the post . . . to view it in full screen in the past hour") to extremely broad ones ("How much time the user has spent [on the platform] in the past 84 days"). Ex. 1 at 35; *accord, e.g.*, Ex. 5, ¶ 12; TikTok MPI Zhou Decl. ¶ 12; YouTube MPI Goodrow Decl. ¶ 11.

The algorithms reduce these millions or billions of data points into numeric values. Golbeck Decl. ¶ 37. The algorithms then use those numeric values to try to identify relationships between them and use those relationships to calculate, as outputs, numeric scores predicting the user's future behavior, e.g., for new posts, estimating the probability that the user will spend time on and engage with that new post. *Id.* ¶¶ 27, 32, 35–37; Narayanan 20, 22. The algorithms' processes for analyzing these millions or billions of data points themselves number in the millions or billions of internal parameters. Golbeck Decl. ¶¶ 32, 35.

The algorithms then construct a personalized feed for that user by ranking posts according to these numeric scores, placing posts with higher predicted likelihoods of the desired behavior—such as the likelihood that the user will "engage" with those posts—higher up in a user's feed, while placing posts with lower predicted likelihoods down lower in the feed, or omitting them entirely. *Id.* ¶¶ 38, 41, 47–52; Austin & Levy 65–66; Narayanan 18. As YouTube's own engineers publicly discussed, "[t]he ranking network . . . assign[s] a score to each video according to a desired objective function using a rich set of features describing the video and user. The highest scoring videos are presented to the user, ranked by their score." Paul Covington et al., *Deep Neural Networks for YouTube Recommendations* 192 (2016) (attached as Ex. 6).

The algorithms also continually update themselves as new data is collected (e.g., data on how users respond to each new feed and new post they see). Golbeck Decl. ¶¶ 36, 39–40, 50. In other words, there are no set rules on how the algorithms compile personalized feeds or select

which posts go into the feeds; the algorithms continually adjust and rewrite their own rules to better achieve their overarching goal. *Id.* ¶¶ 39–40. For example, "[i]f a user was recently recommended a video but did not watch it[,] then the model will naturally demote this impression on the next page load." Covington 196; *see* Austin & Levy 51–57 (describing the process of "gradient descent—the critical step in the process by which the [algorithmic] model writes its own rules"); Narayanan 31 ("If some types of content are more engaging than others, that will be automatically learned from the data—in fact, it will be learned on a per-user level, so that the algorithm may prefer photos for one user and events for another."). Plaintiffs themselves tout this fact. Meta Vice President, Global Head of Safety, Antigone Davis, attests how:

> Facebook and Instagram use artificial intelligence systems to decide what content appears and in what order, informed by a variety of factors including users' own choices, at any given time. On Facebook and Instagram, Meta suggests certain content in users' Feeds. These Feeds employ prediction models that use many input signals, such as the user's location, who created a post, how a user previously interacted with posts from the same user or category, and the number of people mentioned in a post's comments, among others, to help select content users are most likely to find interesting and engage with. The models and their input signals are dynamic and change frequently over time.

Ex. 5, ¶ 12. Likewise, YouTube Vice President of Engineering Cristos Goodrow attests how:

> The YouTube Recommender System is not a static system, but instead was built by Google's engineers to dynamically evolve and adapt. It is based on billions of signals, and there is no set weight assigned to one specific data point over another. The weighting changes constantly, and varies depending on how to accomplish the primary goal of the YouTube Recommender System, which is to predict which content will be relevant, enjoyable, and valuable to a particular user at a particular moment and deliver that content to that user.

YouTube MPI Goodrow Decl. ¶ 11. An internal slide deck titled "[redacted]" discussed how the "[redacted]" and has "[redacted]" Ex. Y1173 at 6936. TikTok documents likewise discuss how TikTok uses "[redacted]" Ex. T504 at 8324; Ex. T427 at 492:2–495:12. TikTok's algorithm head Zhu Wenjia explained, TikTok's algorithms take in "[redacted]" and output [redacted]: "[redacted]" Ex. T427 at 493:6–22.

b. The Algorithms Have No Understanding of the Content of the Feeds They Are Compiling; Instead, All They Can Measure and Attempt to Maximize Are Metrics Like the Time Users Spend on Platforms

Plaintiffs' feed-personalization algorithms have no understanding of the actual content or substance of the feeds they are compiling or the underlying user posts that they are selecting for display in each user's feed. "The model does not 'understand' the content in a human sense, nor does it form beliefs or intentions about what users should see. It produces rankings based on statistical predictions about behavior, continuously updated as new data is collected." Golbeck Decl. ¶¶ 39, 47–49. In other words, the personalized feeds that Plaintiffs' algorithms compile do not reflect any content or idea—rather, the feeds are just the outputs of mathematical functions. *Id.* ¶¶ 47–54. As former Meta Responsible AI researcher Dr. Simons explained, "[redacted]" Ex. M13 at 116:16–18. Likewise, as TikTok's algorithm team head Zhu testified or agreed, "[redacted]" Ex. T427 at 520:4–7. "[redacted]" *Id.* at 520:13–19. Likewise, as YouTube Senior Director of Product Management James Beser testified, "[redacted]" Ex. Y702 at 252:16–253:1; *see also* Ex. Y1003 at 119:19–20 (YouTube Vice President of Engineering John Harding explaining, "[redacted]").

The only things that Plaintiffs' feed-personalization algorithms can perceive, and try to predict, are numeric data metrics, like the amount of time a user spends on a platform and how much the user engages with content (e.g., watches videos to completion, comments on them, re-shares them, etc.). *See* Golbeck Decl. ¶¶ 37–41; Hilbert Decl. ¶¶ 50–53; Narayanan 28. Maximizing the amount of engagement and the time the user spends on the platform thus serves as the basis for how the algorithms select what goes into each user's personalized feeds. Golbeck Decl. ¶¶ 38–39, 47–49, 53; Hilbert Decl. ¶¶ 50–53; *see also, e.g.*, Ex. M229 at slide 5; Ex. Y715 at 561:3–9.

c. Not Only Do Humans Not Write the Rules That Govern How Plaintiffs' Feeds Are Compiled, Humans Do Not Even *Know* the Rules and Cannot Explain How the Algorithms Operate or Why They Do What They Do

Not only do humans not write the rules that govern how Plaintiffs' feed-personalization algorithms operate, humans do not even *know* what the rules are, and cannot explain how or why the algorithms select, or do not select, the posts they do for inclusion in any user's feed.

Machine-learning algorithms are inherently unpredictable. Hilbert Decl. ¶¶ 35–38; *see also id.* ¶ 106 ("**[M]achine learning comes with inherent and irreducible uncertainty in its conduct** …, and that nobody, not the engineers who designed, trained and fine-tuned it, nor the executives who execute it, can predict or completely control its machine learned conduct."). This is the case for multiple reasons. First, the algorithms are simply too large and complex, with too many variables, for any human to understand them or be able to explain them:

> Modern machine learning models derive their internal parameters—often numbering in the millions or billions—through training on large datasets. These parameters are not individually interpretable and interact in complex, non-linear ways, making it practically impossible to trace how specific inputs lead to specific outputs. As a result, even the engineers who design and train these systems cannot predict with certainty how the model will rank or score a particular piece of content for a given user.

Golbeck Decl. ¶¶ 32–33; *see id.* ¶¶ 34–35, 48, 50–52; Hilbert Decl. ¶ 36; Austin & Levy 62–63. Second, the environment in which the algorithms are operating is also constantly changing: "these systems operate within non-stationary environments characterized by feedback loops, dynamically changing the stationarity used for prediction," thus creating an "uncontrollable and unpredictable result." Hilbert Decl. ¶¶ 35–36; *see* Golbeck Decl. ¶ 34. The millions or billions of data points (including data points on individual users) that the algorithms track are constantly changing, as are the millions or billions of internal parameters that the algorithms use to assess that data and compile each user's personalized feed. Golbeck Decl. ¶¶ 34, 50; Hilbert Decl. ¶ 37; *accord* Ex. 5, ¶ 12; YouTube MPI Goodrow Decl. ¶ 11; Ex. T427 at 493:20–22. Even assuming (counterfactually) that Plaintiffs were using "Good Old-Fashioned AI" algorithms that operated on a set of predefined steps instead of machine-learning algorithms that rewrite their own rules, their output would still be unpredictable: "even if the code might be deterministic, [the] social

algorithms adjust their chosen behavior in real time to human dynamics, which makes their behavior as unpredictable [as] the social phenomena it draws from." Hilbert Decl. ¶ 37 (quotation marks omitted). Simply put, "[t]oday's social algorithms are so complex that no single person can fully understand them." *Id.*; *see* Golbeck Decl. ¶¶ 32–34; Austin & Levy 62–63.

Plaintiffs do not deny this—rather, they concede that humans do not know how their machine-learning algorithms operate or why the algorithms do what they do. For example, former Meta Vice President Brian Boland testified before Congress that "[c]ontrary to popular belief, technology companies cannot accurately predict what their algorithms will do. . . . These platforms function as a black box by design." Testimony of Brian Thomas Boland, Former Vice President, Facebook, before the S. Comm. on Homeland Sec. & Governmental Affairs 4–5 (Sept. 12, 2022) (attached as Ex. 7). As Boland further testified in deposition:

[redacted]

Ex. M8 at 274:1–15. Similarly, TikTok documents discuss how [redacted] Ex. T504 at 8324; Ex. T427 at 492:17–497:8. Likewise, YouTube's "[redacted]" slide deck prepared by YouTube engineers explains how [redacted] and how they "[redacted]" Ex. Y1173 at 6931.

d. Humans Do Not Know and Cannot Predict the Algorithms' Output, i.e., What the Algorithms Will "Say" or What the Contents of Any User's Feed Will Be; Indeed, the Algorithms Often Amplify Messages That Their Human Creators Want to *Not* Convey

"Eight-year-old Lalani Erika Walton wanted to become 'TikTok famous.' Instead, she wound up dead." Brian Contreras, *TikTok is sued over death of two young girls in viral 'blackout challenge'*, L.A. TIMES, July 1, 2022 (attached as Ex. 8). According to lawsuits filed by parents

of the deceased children, TikTok's algorithms promoted videos of the so-called "blackout challenge"—in which participants attempt to choke themselves into unconsciousness—in the personalized feeds of children as young as eight and nine. *Id.* These young children, who did not understand the dangerous nature of the videos they saw in their feeds, tried to copy what they saw, and died from it. *Id.* The mother of one eight-year-old found her "hanging from her bed with a rope around her neck." *Id.* The police told her that her daughter had been watching TikTok blackout challenge videos "on repeat" before she died. *Id.*

These are not isolated incidents, nor are they limited to TikTok. *Id.* Rather, it is the natural result of all three Plaintiffs' use of machine-learning algorithms to compile their addictive feeds. As discussed above, the algorithms have no understanding of the feeds they are generating. And humans cannot predict the algorithms' output—i.e., humans do not know in advance what the algorithms will "say" or what the actual content of any user's personalized feed will be. *See* Golbeck Decl. ¶¶ 32–34; Hilbert Decl. ¶¶ 41, 82, 91; Austin & Levy 60–64.

All three Plaintiffs' algorithms promote harmful content or content that Plaintiffs claim to desire *not* to promote. Hilbert Decl. ¶¶ 43–46, 111. Meta, for example, states in its official policies that it does not allow posts that promote or encourage suicide or self-injury—yet a study showed that where teenage user accounts had previously accessed posts on Instagram Reels with suicide, self-harm, or highly depressive content, Instagram's algorithms would continue to recommend more such posts: between 14–20% of the recommended posts referenced suicide methods, 44–55% referenced suicide and/or self-harm ideation, and 97–99% referenced feelings of intense depression, misery, and despair. *Id.* ¶ 45. TikTok similarly states that it does not allow posts that promote suicide or self-harm—yet an audit with user accounts set at age 13 that "paused briefly on videos about body image and mental health, and liked them" found that "[w]ithin 2.6 minutes, TikTok recommended suicide content. Within 8 minutes, TikTok served content related to eating disorders. Every 39 seconds, TikTok recommended videos about body image and mental health." *Id.* YouTube states that it does not allow content that endangers the emotional and physical well-being of minors—yet a study found that in 3.5% of cases where a toddler starts from a video that appears in a top results of a toddler-appropriate search (e.g.,

"Peppa Pig"), YouTube's recommendations will send the user to a disturbing video within ten hops. *Id.* As YouTube Senior Director of Product Management James Beser testified, [redacted]. Ex. Y702 at 373:4–374:22.

Plaintiffs are well aware that their machine-learning feed-personalization algorithms can promote harmful material that Plaintiffs state they oppose, and lead users down "rabbit holes" of increasingly extreme types of content—because, as described above, the algorithms have no actual understanding of the content they are feeding to users. The Wall Street Journal conducted an investigation finding how users were pushed into extreme rabbit holes: in one example, after one user account watched videos about heartbreak, emotional pain, and mental health for just 36 minutes, TikTok's algorithm began filling 93% of the user's feed with depression-related content. *Investigation: How TikTok's Algorithm Figures Out Your Deepest Desires*, WALL ST. J., July 21, 2021, at 7:30–9:00, https://tinyurl.com/ydrb9wa7. The remaining 7% were mostly ads—i.e., money for TikTok. *Id.* at 9:00–9:30; *see also* Rob Barry et al., *How TikTok Serves Up Sex and Drug Videos to Minors*, WALL ST. J., Sept. 8, 2021 (attached as Ex. 9); Tawnell D. Hobbs et al., *'The Corpse Bride Diet': How TikTok Inundates Teens With Eating-Disorder Videos*, WALL ST. J., Dec. 17, 2021 (attached as Ex. 10). TikTok's own Investigations team determined that the "[e]ase of entering a filter bubble of [suicide/self harm] content in U.S." remains "HIGH." Ex. 1 at 107; Ex. T517 at 8485. A 2020 document drafted by TikTok's [redacted] described how "[redacted]" Ex. T509 at 4763. TikTok's own internal review of [redacted] Ex. T518 at 5925. TikTok's own data confirmed that [redacted] of eating-disorder videos viewed by minors were

algorithmically recommended. Ex. 1 at 107; Ex. T515 at 1706. *See generally* Ex. 1 at 106–08.

Similarly, an internal 2021 Meta slide deck titled "[redacted]" discussed how content containing "Bullying and Harassment (B&H), Suicide and Self Injury (SSI), and Violence and Incitement (V&I)...are all >2x higher prevalence for teens than the average user" and how "[redacted]" Ex. 1 at 39; Ex. M231 at slides 43, 45. After the Wall Street Journal's first two articles in its TikTok investigation came out, then-Instagram Senior Director [redacted] discussed internally in response to "[redacted]" how in Meta's own platforms, "[redacted]" and "if you start following borderline accounts or interacting with such content, our recommendation algorithms will start pushing you down a rabbit hole of more egregious content." Ex. 1 at 39; Ex. M243 at 7939. As Lotte Rubaek, a former member of Meta's Suicide and Self Injury Global Expert Group, testified, "once the young people start to self-injure, they tell us that their feed...on Instagram suddenly gets full of this content. [redacted]" Ex. 1 at 39–40; Ex. M256 at 24:22–25:8. Meta's former Head of Youth Policy Vaishnavi Jayakumar likewise testified, Instagram's algorithm "does sometimes push content that, in the aggregate, could be seen as encouraging suicide" and "might promote content that, in aggregate, could be seen as promoting self-injury," including pushing that content out to children. Ex. 1 at 40; Ex. M2 at 221:4–222:2. *See generally* Ex. 1 at 37–42.

Similarly, internal YouTube documents confirm how "[redacted]" how "[redacted]" and identified how "[redacted]" citing [redacted]

[redacted]” Ex. Y1026 at 2617–20. They discuss how “[redacted]” how [redacted] [redacted]” how “[redacted] [redacted]”—and how “[redacted]” Ex. Y1025 at 5399. An internal study found that [redacted]. Ex. Y702 at 368:7–377:5. *See generally* Ex. 1 at 169–71.

The reason this content appears in Plaintiffs’ addictive personalized feeds is not because Plaintiffs necessarily want to promote this sort of content. Plaintiffs do not assert that any human being at their companies has made an editorial choice to promote posts that advocate suicide, self-harm, “blackout challenges,” and so on, in children’s feeds. Rather, these posts end up in children’s personalized feeds precisely because humans are *not* the ones compiling the feeds—and because it instead is the algorithms that are doing so, in pursuit of their goal to maximize user “engagement,” without any understanding of the actual content of the posts they are pushing out to children in the personalized feeds they are compiling. *See* Hilbert Decl. ¶ 91.

e. Feed Personalization Is Distinct from Content Moderation

Each Plaintiff makes a point of mentioning how it engages in content moderation, including by compiling a set of “Community Standards” or “Community Guidelines” (collectively, “Community Standards” or “Standards”) that bar certain kinds of user posts (e.g., violent or graphic content, nudity, bullying, hate speech, etc.) from appearing on their platforms, and then by using algorithms to remove content that violates their Standards from their platforms. Meta

MPI 5–6; TikTok MPI 5–6; YouTube MPI 5. This content moderation is beside the point in these cases, however, as it is not regulated or restricted in any way by SB976. *See infra* p. 26.

Plaintiffs' content-moderation activities and algorithms are separate from Plaintiffs' feed-personalization activities and algorithms. Golbeck Decl. ¶¶ 55–65. As all three Plaintiffs acknowledge, their content-moderation algorithms work first to remove (or prioritize or deprioritize) material on their platforms based on Plaintiffs' user-agnostic guidelines. Only after that do Plaintiffs' feed-personalization algorithms operate on the pool of material that remains and, from that pool, select posts to display in each individual user's feed. Meta MPI Blackstrom Decl. ¶ 7 (discussing Meta's "multi-stage process" where "Meta excludes content, including content flagged by humans and by automated review for violating Meta's Community Standards," and "then" uses "signals to make personalized predictions about which content the user may find relevant and valuable"); TikTok MPI Schiller Decl. ¶¶ 3–6 (discussing how all user posts on TikTok "first" goes through content moderation before they appear on platform); TikTok MPI Zhou Decl. ¶ 17 ([redacted]); YouTube MPI Goodrow Decl. ¶¶ 24–25 (discussing how YouTube applies its Community Guidelines to user posts first and how its personalization algorithm only recommends videos that have not been removed during content moderation); *see* Golbeck Decl. ¶¶ 55–65.

Conversely, if material survives content moderation, Plaintiffs' feed-personalization algorithms do not take the content of the material into account in determining whether to place that material into user feeds. Golbeck ¶ 63. Content moderation is not perfect, and as discussed above, material that violates Plaintiffs' Community Standards can make it through content-moderation processes. Ironically, such material is often then amplified at the feed-personalization stage, as this violative material is often "interesting" and drives user engagement—and user engagement, not the actual content of any material, is what drives Plaintiffs' feed-personalization algorithms. *See, e.g.*, Ex. M208 at slide 11 ("[redacted]"); Ex. T402 at 407:10–408:23 ("[redacted]

[redacted]"); Ex. Y702 at 372:21–375:16 (discussing how, [redacted]).

II. PLAINTIFFS' ADDICTIVE PERSONALIZED FEEDS CAUSE HARMS TO CHILDREN AND TEENS

There is ample evidence, including Plaintiffs' own internal documents, demonstrating that Plaintiffs' addictive feeds cause an array of harms to children and teens. And as explained in detail below, the link between addictive feeds and harms is *causal*, not a mere correlation.

A significant portion of American teens report being constantly connected to internet platforms. Radesky Decl. ¶¶ 25–26. In the most recent Pew Research Center survey, over 76% of minor teens (13–17) reported using YouTube daily (43% several times a day and 17% almost constantly), 61% reported using TikTok daily (34% several times a day, 21% almost constantly), and 61% reported using Instagram daily (31% several times a day, 12% almost constantly). *Id.*; *accord* Christakis Decl. ¶¶ 90–93. Teens do not perceive this time to be well spent, however—rather, they find it hard to stop using these platforms, spend more time than they want there, and believe the platforms manipulate them into staying online. Radesky Decl. ¶ 26.

Children and teens are more susceptible than adults to social-media designs. *Id.* ¶¶ 37–41; Telzer Decl. ¶¶ 98–105; Christakis Decl. ¶¶ 65–69. Children and teens have underdeveloped executive functions and thus less impulse control, perspective taking, and critical thinking about digital media. Radesky Decl. ¶¶ 37, 40; Telzer Decl. ¶¶ 101–05, 113–18; Christakis Decl. ¶¶ 65–69. Their behavior is more reinforced by immediate gratification than thinking about the long-term consequences of their actions, which can lead to unhealthy habit formation. Radesky Decl. ¶ 39; Telzer Decl. ¶¶ 99–102, 113–14. During adolescence in particular, the brain undergoes significant changes, including a hypersensitivity to rewarding stimuli that makes teens more likely than adults to seek out and repeat behaviors that provide immediate gratification. Telzer Decl. ¶¶ 81–90, 113. Conversely, the portions of the brain that support decision-making, cognitive control, impulse control, and emotional regulation are still developing, leading to a "gap" in adolescence where youth are highly sensitive to rewards while not yet being able to effectively engage in cognitive control. *Id.* ¶¶ 92–94, 114. This is coupled with reduced self-

monitoring, meaning that teens can easily lose track of time and go into a dissociative mental state while scrolling social media feeds—facts of which Plaintiffs are well aware. Radesky Decl. ¶¶ 39, 41. Teens in particular have heightened sensitivity to social cues and increased emotional responses as a product of more mature emotion-processing regions of the brain and underdeveloped judgment regions, which are receptive to social media's ability to activate reward-processing regions in the brain to motivate continued engagement. Telzer Decl. ¶ 104. Teens and young adults who biologically lack higher cortical functions, including impulse control, are more vulnerable to potentially addictive substances and behaviors. Christakis Decl. ¶ 114. And a nationally representative sample of 11- to 15-year-old girls found that over a third said they felt "addicted" to social media. Radesky Decl. ¶ 51. Another study of 11- to 17-year-olds found that 14% found it difficult to stop using technology once they start, 33% were short on sleep due to being online late at night, and 39% often neglected daily obligations. *Id.* ¶ 52.

Plaintiffs are well aware of how teens are uniquely vulnerable to their platforms' addictive designs. For example, an internal Meta document titled "[redacted]" explains:

[redacted]

Ex. M224 at 0672–81 (emphasis in original, internal hyperlinks omitted). Likewise, an internal TikTok document titled "[redacted]" discusses how:

[redacted]

[redacted] Ex. T556 at 1514. Similarly, an internal YouTube document titled "[redacted]" discusses how "[redacted]" Ex. 735 at 1252 (emphasis in original).

Problematic and addictive social-media use has been linked with sleep disruption; diminished attentional capacity; lower working memory; increased impulsivity; ADHD; poor school performance; depression; anxiety; body-dysmorphic disorder and eating disorders; and suicide and self-harm. Radesky Decl. ¶¶ 49, 58–74; Christakis Decl. ¶¶ 195–322; Telzer Decl. ¶¶ 116–228, 302; Twenge Decl. ¶¶ 50–54, 81–105, 116–21. As noted, MRIs have indicated altered brain function among individuals addicted to social media, similar to that of individuals addicted to drugs or alcohol. Telzer Decl. ¶ 173; Twenge Decl. ¶ 152; Christakis Decl. ¶ 129.

These problems are specifically exacerbated by addictive personalized feeds. Addictive personalized feeds increase the time that children remain on internet platforms. Radesky Decl. ¶¶ 77–83; Telzer Decl. ¶¶ 62–65. They do this by "automatically serv[ing] users with content, removing the need for them to make any decisions or actively seek content to engage with. . . . These feeds enable children to keep scrolling (sometimes for hours) without ever needing to actively search for something, and without the platform prompting or encouraging them to stop." Radesky Decl. ¶ 85. Viewing personalized videos activates the dopamine system of the brain, in a feedback loop that leads to craving for yet more personalized videos and to compulsive use. Telzer Decl. ¶ 64. This addiction is content-agnostic: even if the underlying content that a user is viewing does not hold intrinsic addictive qualities, users may become addicted because platforms continually amplify more and more of that content to the user, whatever it might be. *Id.* ¶ 69. Additionally, feed personalization removes friction and uncertainty from the user experience, increasing vulnerability to compulsive usage patterns, particularly among youth with weaker

impulse regulation. *Id.* ¶ 103. As teens become accustomed to highly personalized feeds and rapid feedback, their ability to delay gratification may weaken, making it more challenging for them to develop long-term self-discipline and impulse control. *Id.* ¶ 114. Excessive exposure may alter dopamine reward pathways in the brain, restructuring these pathways over time to become increasingly tuned to these cues and making it even harder to disengage. *Id.* ¶¶ 173, 179.

Addictive feeds create a targeted stream of material that feels endless, keeping users hooked as they anticipate the next video, which might be even more entertaining or rewarding. *Id.* ¶ 144. One study involving brain scans of young adults showed significantly greater neural activation when viewing videos that had been personalized as compared to generic, non-targeted videos, even when both sets of videos came from the same platform and could feature similar content—the brain responds not merely to the intrinsic properties of a video itself, but specifically to the algorithmic selection and sequencing of that video, as in an addictive feed. *Id.* ¶¶ 47–49, 144–48.

Additionally, during adolescence, the "social brain" undergoes significant development in the systems processing information from peers, leading to greater sensitivity to cues about social acceptance, standing and belonging (e.g., asking "What do others think of me?" and "Where do I fit in?"). *Id.* ¶ 98. Algorithmic recommendations (which in actuality are just computer-generated) are often interpreted by the adolescent brain as if they were meaningful social signals, providing a constant stream of what the brain perceives as socially relevant cues at a stage when it is especially tuned to them, shaping beliefs about norms, popularity, and personal value. *Id.*

In addition to increasing the amount of time spent on platforms, personalized feeds amplify extreme views and lead users down rabbit holes of increasingly extreme content, through a process some companies call "preference amplification." Radesky Decl. ¶¶ 94–96. For example, when teens with a tendency toward poor body image engage with posts about dieting or fitness, their feeds then amplify this and serve them more and more of this content. *Id.* ¶¶ 96–97. Teens report that they have a hard time getting eating-disorder content out of their feeds once they have engaged with it. *Id.* ¶ 96. One study found that individuals with an eating disorder are *over 4000% more likely* to be shown videos that are appearance-oriented or that showcase toxic eating behaviors, as compared to those without an eating disorder. Telzer Decl. ¶¶ 50, 192. Similarly,

personalized feeds amplify out to teens posts with sexual content; racist, homophobic, and misogynistic content; violence; body-image and disordered-earing content; alcohol-, steroid-, and supplement-related content; and suicide and self-harm content, once teens have engaged with such posts. Radesky Decl. ¶¶ 97–101; Telzer Decl. ¶¶ 52–57, 61; Twenge Decl. ¶¶ 47–48.

Addictive feeds perpetuate themselves in a feedback loop—the longer children stay on their feeds, the more user actions the feed-personalization algorithms have to measure, and the more the algorithms can then further tailor the user's feed to be even more addictive:

> This feedback loop can accelerate the transition from casual use to compulsive engagement, as the platform increasingly optimizes the feed to match the user's reward sensitivities. In this way, platform algorithms may actively amplify existing vulnerabilities, transforming individual differences in social reward processing into sustained behavioral risk.

Telzer Decl. ¶ 166; *see id.* ¶ 69; *accord* Radesky Decl. ¶ 90. One study involving 2000 users directly measured the addictive features of platforms including Facebook, Instagram, and YouTube, and concluded that self-control problems magnified by habit formation might be responsible for 31 percent of all social-media use. Christakis Decl. ¶¶ 175–176.

In summary, Plaintiffs' platforms and personalized feeds cause an array of harms to children and teens—and the link is *causal*, not merely correlative. As Dr. Telzer states:

> Social media leads to, contributes, and causes risks including depression, poor mental health, loss of sleep, poor body image and negative self-esteem, and functional and structural changes in areas of the brain associated with addiction, executive control, social belonging, and social comparison. . . . Recommendation algorithms contribute to, exacerbate, and increase the addictive nature of social media's harms on adolescents. . . . These findings raise critical concerns about how algorithmic design features—not just the content itself—may amplify compulsive use and alter neurodevelopmental pathways, particularly in adolescents whose brains are highly sensitive to reward and social information.

Telzer Decl. ¶¶ 44, 48. Dr. Radesky opines similarly: "problematic, compulsive, and addictive-like use of social media is recognized by global mental health experts and researchers and causes significant decreases in functioning in teens." Radesky Decl. ¶ 53. Dr. Christakis likewise opines that, to a reasonable degree of medical and scientific certainty, there is a causal relationship between social-media use and adverse outcomes in youth, including sleep problems, ADHD, anxiety, depressive symptoms, body dysmorphia, and eating disorders. Christakis Decl. ¶¶ 215, 223, 237, 264, 269, 301, 349–50. Dr. Twenge likewise opines, "[t]he totality of the

evidence demonstrates a clear causal path from social media use to low psychological well-being in adolescents[.]" Twenge Decl. ¶ 8.

The situation is not hopeless, however. Studies show that when children and teens reduce time on internet platforms, their mental-health outcomes improve, with less anxiety and depression, less sleep deprivation, lower emotional and behavioral difficulties, and higher happiness and life satisfaction. *Id.* ¶¶ 7, 99–105; Radesky Decl. ¶¶ 66–67, 71; Christakis Decl. ¶¶ 145, 261; Hilbert Decl. ¶¶ 64–65. Studies further show that when addictive personalized feeds are replaced with non-addictive feeds, the time users spend on social-media platforms decreased, while self-regulation increased. Radesky Decl. ¶¶ 78–81; Telzer Decl. ¶ 62; Hilbert Decl. ¶ 65.

III. SB976 WAS ENACTED TO PROTECT CHILDREN FROM THE HARM OF ADDICTIVE FEEDS

Motivated to address the harms caused to children by addictive online engagement features, the Legislature passed SB976. The Legislature found that "the algorithmic delivery of content and other design features . . . pose a significant risk of harm to the mental health and well-being of children and adolescents." SB976 § 1. It noted that "the risk of poor mental health outcomes doubles for children and adolescents who use social media at least three hours a day." *Id.* It also noted that heavy social media use "leads to less healthy sleep patterns and sleep quality, which can in turn exacerbate both physical and mental health problems." *Id.* And it noted a U.S. Surgeon General's report concluding that children and adolescents who use social media at least three hours a day have double the risk of poor mental health outcomes. *Id.* The Legislature emphasized, however, that "[s]ocial media provides an important tool for communication and information sharing." SB976 § 1(a). Given these findings, the Legislature narrowly tailored SB976 to address the harms caused by excessive time online while preserving minors' access to content and platforms across the internet and avoiding intrusion on companies' speech rights.

Plaintiffs challenge three provisions: §§ 27000.5, 27001, and 27002(b)(2) and (4). Section 27000.5 defines "addictive feeds" as feeds of user posts that are "recommended, selected, or prioritized for display to a user based, in whole or in part, on information provided by the user, or otherwise associated with the user or the user's device," with certain exceptions. Section 27001 requires platforms to obtain parental consent to provide addictive feeds to children. Section

27002(b)(2) requires platforms to limit children's access to addictive feeds to one hour per day unless modified by their parents. Section 27002(b)(4) requires platforms to offer parents a mechanism to be able to set that their children's default feed will be a non-addictive feed.

SB976 is also clear about what it does *not* do. It expressly does not limit the content any user can search for, access, and view. §§ 27000.5(a)(1), (2), (4). It does not require companies to give parents any special access or control over their children's accounts. § 27003(a). It does not interfere with companies' ability to moderate content or to prioritize or deprioritize content "independently—i.e., unrelated to any individual." § 27000.5(a); *cf.* TikTok MPI 14.

LEGAL STANDARD

"[A] preliminary injunction is 'an extraordinary remedy never awarded as of right.'" *Benisek v. Lamone*, 585 U.S. 155, 158 (2018) (quoting *Winter v. Nat. Res. Def. Council, Inc.*, 555 U.S. 7, 24 (2008)). An MPI plaintiff "must show that: (1) she is likely to succeed on the merits, (2) she is likely to suffer irreparable harm in the absence of preliminary relief, (3) the balance of equities tips in her favor, and (4) an injunction is in the public interest." *Garcia v. Google, Inc.*, 786 F.3d 733, 740 (9th Cir. 2015) (en banc) (citing *Winter*, 555 U.S. at 20). Where the government is a party, the last two factors merge. *Drakes Bay Oyster Co. v Jewell*, 747 F.3d 1073, 1092 (9th Cir. 2014).

"[T]he Federal Rules of Evidence do not strictly apply in the preliminary injunction context," and courts "may give even inadmissible evidence some weight[.]" *Flathead-Lolo-Bitterroot Citizen Task Force v. Montana*, 98 F.4th 1180, 1189 (9th Cir. 2024). Among other things, courts may consider hearsay. *Johnson v. Couturier*, 572 F.3d 1067, 1083 (9th Cir. 2009).[5]

ARGUMENT

I. PLAINTIFFS HAVE NOT ESTABLISHED A LIKELIHOOD OF SUCCESS ON THE MERITS

A. Plaintiffs Have Failed to Demonstrate That SB976 Restricts Conduct Protected by the First Amendment Because Their Addictive Feeds Are Not Expressive

A plaintiff may not bring a pre-enforcement as-applied First Amendment challenge "by

[5] For example, the Court may consider the MDL MSJ Opposition, notwithstanding that Plaintiffs have refused to produce to Defendant all of the underlying Opposition exhibits.

nakedly asserting that his or her speech was chilled[.]" *Stockton v. Brown*, 152 F.4th 1124, 1144 (9th Cir. 2025). Rather, at a minimum, Plaintiffs must "articulate[] a concrete plan to violate the law in question" (here, SB976) and "give details about their future speech such as 'when, to whom, where, or under what circumstances' they intend to violate the law in question." *Id.* (brackets omitted). Plaintiffs further must establish "a *genuine* threat of *imminent* prosecution" based on that speech. *Project Veritas v. Schmidt*, 125 F.4th 929, 941 (9th Cir. 2025) (en banc) (emphasis in original). It thus is critical to establish (1) what conduct, specifically, Plaintiffs have a concrete plan to engage in, and (2) whether SB976 actually regulates that conduct.

Plaintiffs do not allege any concrete plan for actual human beings to review the user posts on Plaintiffs' platforms or to compile those posts into any user's feeds. Nor do Plaintiffs allege a concrete plan to engage in content moderation (e.g., applying Community Standards) or any other user-agnostic activities where they face a genuine threat of imminent prosecution, as SB976 does not cover such activities. Rather, the conduct at issue in these as-applied challenges is the computerized processes undertaken by Plaintiffs' feed-personalization algorithms. When Plaintiffs' conduct is properly framed, their as-applied claims must fail, because the First Amendment does not protect the addictive-feed output of machine-learning algorithms.

"First Amendment protection extends 'only to conduct that is inherently expressive[.]'" *Project Veritas*, 125 F.4th at 946 (citing *Rumsfeld v. Forum for Acad. & Inst. Rts., Inc.*, 547 U.S. 47, 66 (2006) (*FAIR*); *Clark v. Cmty. for Creative Non-Violence*, 468 U.S. 288, 294 (1984)); *accord NetChoice I*, 761 F. Supp. 3d at 1219. Plaintiffs bear the burden of establishing that their activity "is expressive" in the first instance and "that the First Amendment even applies." *Clark*, 468 U.S. at 293 n.5; *B&L Prods., Inc. v. Newsom*, 104 F.4th 108, 112 (9th Cir. 2024). And as the Ninth Circuit held specifically in the context of SB976, "an algorithm that 'respond[s] solely to how users act online,' merely 'giving them the content they appear to want,' probably is not expressive." *NetChoice II*, 152 F.4th at 1014 (quoting *Moody*, 603 U.S. at 736 n.5; *id.* at 746 (Barrett, J., concurring)). The addictive-feed output of Plaintiffs' algorithms are not expressive for at least two separate and independent reasons, either one of which is sufficient to confirm that the conduct at issue is not protected by the First Amendment. First, Plaintiffs' feeds are outputs

generated by computerized algorithms, and not human beings, and are thus not protected speech. And second, Plaintiffs' feeds are merely compilations of third-party speech arranged to reflect users' interests back to them that do not convey an expressive message of Plaintiffs' own.

1. Addictive Feeds Compiled by Non-Human Machine-Learning "Black Box" Algorithms Are Not Protected Speech

As described above, *see supra* pp. 8–18, no human being reviews the pool of user posts and material available on Plaintiffs' platforms. No human being selects which posts should be displayed in which user's feeds. Indeed, humans do not even know and cannot predict what the contents of Plaintiffs' feeds will be, because they are generated by machine-learning "black box" algorithms that humans do not understand. These types of feeds—feeds generated by machine-learning algorithms without the exercise of human editorial judgment, and where humans do not even know what the algorithms are "saying"—are not protected speech or expression.

First Amendment protection extends only to human beings. In those instances where First Amendment rights have been extended to non-human entities like corporations, this has been only because those rights derive from human beings. *Burwell v. Hobby Lobby Stores, Inc.*, 573 U.S. 682, 706–07 (2024) ("An established body of law specifies the rights and obligations of the *people* . . . who are associated with a corporation in one way or another. When rights, whether constitutional or statutory, are extended to corporations, the purpose is to protect the rights of these people.") (emphasis in original); *accord Moody*, 603 U.S. at 746–47 (Barrett, J., concurring) (discussing corporations' First Amendment rights in context of how corporations "are composed of human beings with First Amendment rights"); *see also* David Atkinson et al., *Intentionally Unintentional: GenAI Exceptionalism and the First Amendment*, 23 FIRST. AMEND. L. REV. 173, 177, 192 (2024). Where, as here, humans are so far removed from the computerized processes by which Plaintiffs' personalized feeds are compiled, and no human understands those processes or knows what content is actually being delivered, First Amendment protection does not apply.

It is of no moment that Plaintiffs' algorithms were originally created by humans. The fact that a mechanism might be created by humans does not derivatively mean that any output created by that mechanism is the product of a human, much less expressive speech by a human. Atkinson

186–87. For example, the fact that humans can create mechanisms that can create video footage (namely, video cameras) does not mean that any footage that the mechanism creates is necessarily the expressive speech of a human. *See Project Veritas*, 125 F.4th at 946. To be expressive, the creation must be the result of expressive choices by a human being: video footage created by a camera as part of a human film director's creation of a cinematic work is expressive, but video footage created by a wall-mounted security camera is not. *See id.* Likewise, a wall-mounted security camera programmed to film when it detects movement has not created expressive speech, because even though the camera's footage might be the product of a choice (whether by the camera itself, or its human creator, to film when motion is detected), that is not an *expressive* choice. *See id.* ("First Amendment protection extends 'only to conduct that is inherently expressive'"). Plaintiffs' feed-personalization algorithms might be infinitely more complex than a motion-detecting camera, but at their core, they are non-human mechanisms simply responding to stimuli whose output does not reflect an expressive choice by an actual human being. *See id.*; *Garcia v. Character Techs., Inc.*, 785 F. Supp. 3d 1157, 1177–79 (M.D. Fla. 2025) (declining to hold that "words strung together by an [AI large language model] are speech").

Citing *Moody*, Plaintiffs claim that their feeds are expressive because they reflect Plaintiffs' editorial judgment. But the algorithmic feed personalization at issue here does not reflect the kinds of editorial judgments described in the Supreme Court's caselaw. For example, in the context of newspaper editors, the Court explained that "the function of editors" is "first and foremost, to make decisions about 'the content of the paper' and 'the choice of material to go into' it." *Moody*, 603 U.S. at 728 (brackets omitted) (quoting *Mia. Herald Publ'g Co. v. Tornillo*, 418 U.S. 241, 258 (1974)). "'The choice of material,' the 'decisions made as to content,' the 'treatment of public issues'—'whether fair or unfair'—all these 'constitute the exercise of editorial control and judgment.'" *Id.* at 738 (brackets omitted) (quoting *Tornillo*, 418 U.S. at 258). "Those choices rest on a set of beliefs about which messages are appropriate and which are not (or which are more appropriate and which less so)." *Id.* In *Moody*, the Court applied these editorial-judgment concepts to internet platforms' content moderation based on their user-agnostic Community Standards. The platforms' Standards—which no one disputed were written

entirely by human beings—set forth rules about what content would not be allowed (for example, they disallowed pornography, hate speech, and misinformation on select topics) and reflected a "wealth of user-agnostic judgments about what kinds of speech, including what viewpoints, are not worthy of promotion." *Id.* at 719–20, 736 n.5. The platforms' content-moderation algorithms just "implement[ed] those standards," which, again, were decided upon and written by human beings. *Id.* at 735. In that scenario, editorial judgment was exercised by actual human beings.

But in stark contrast to the content moderation in *Moody* (which, again, is not restricted by SB976 in any way), Plaintiffs' algorithmically-driven *feed personalization*—an entirely separate process from content moderation, *see supra* pp. 18–20—does not reflect the editorial judgment of any human being in any meaningful way. *See Moody*, 603 U.S. at 732 n.4 ("Of course, an entity engaged in expressive activity when performing one function may not be when carrying out another."). No human being reviews the pool of user posts and material available on Plaintiffs' platforms.[6] *See supra* pp. 8–18. No human being is making the personalization decisions about which posts should be displayed in which user's feeds based on a set of beliefs about which messages are appropriate and which are not—e.g., a belief that children and teens who might have watched some suicide or self-harm videos should be fed more of them. *See id.* Instead, the decisionmakers about which posts should be displayed in which user's feeds are Plaintiffs' machine-learning algorithms, which write and rewrite their own rules for how they make these decisions. *See supra* pp. 9–11. This occurs to such an extent that humans do not and cannot fully understand the algorithms or predict what content they will feed users, *see supra* pp. 13–14; in fact, the algorithms often amplify content that Plaintiffs assert they want to *not* convey, *see supra* pp. 14–18. As this Court has explained, "[r]ather than reflecting human judgments about the messages that should be disfavored," Plaintiffs' algorithms "reflect more and more of [the algorithms'] own 'judgment'" and therefore "it would become harder to say that the algorithm

[6] Humans may make some user-*agnostic* decisions, such as decisions to not promote flat-earth videos or to designate posts from certain sources as "high-quality" and promote them. *See* YouTube MPI 5. But user-agnostic decisions, like content moderation, are separate from addictive feed personalization, and, as such, not regulated by SB976 and thus not at issue here.

implements human expressive choices[.]" *NetChoice I*, 761 F. Supp. 3d at 1223.

Plaintiffs elide the fact that no actual human being is exercising editorial judgment over their personalized feeds by referring to themselves as undifferentiated corporate entities ("Meta," "TikTok," and "YouTube" as a whole), and framing these corporate entities as the actors that are exercising editorial judgment. But as discussed above, a corporation has First Amendment rights only as derived from "the rights and obligations of the *people*"—the actual "human beings"—"who are associated with a corporation in one way or another." *Hobby Lobby*, 573 U.S. at 706 (emphasis in original). As Justice Barrett explained in her concurrence in *Moody*,[7] the First Amendment protects "human beings' constitutionally protected right to *decide for themselves* the ideas and beliefs deserving of expression, consideration, and adherence." *Moody*, 603 U.S. at 746 (emphasis in original, quotation marks and brackets omitted). As Justice Barrett asked, when an AI algorithm is the one making the decisions, "has a human being with First Amendment rights made an inherently expressive choice not to propound a particular point of view?" *Id.* (quotation marks and ellipsis omitted). Likewise, as Justice Alito discussed in his concurrence, there is a difference between "editors [who] are real human beings" versus how "when AI algorithms make a decision, 'even the researchers and programmers creating them don't really understand why the models they have built make the decisions they make.'" *Id.* at 795. Here, no actual human being is exercising editorial judgment over Plaintiffs' feed personalization.

Humans have, of course, made some choices with respect to Plaintiffs' personalized feeds—for example, the choice to have Plaintiffs' machine-learning algorithms try to maximize user "engagement" and keep users on their feeds for as long as possible. But these choices are not exercises of *editorial* judgment—they make no decisions about "'the content of [any user's feed]' and 'the choice of material to go into' it," *cf. id.* at 728 (majority op.)—as evidenced by the fact that these choices tell one nothing about what the actual content of the feed will be.

Were there any doubt about whether Plaintiffs could be said to be exercising editorial judgment over their personalized feeds, it is put to rest by the fact that Plaintiffs do not know

[7] Because Justice Barrett's vote was necessary to form a majority, her concurrence carries "particular weight." *Schmitz v. Zilveti*, 20 F.3d 1043, 1045 (9th Cir. 1994).

what their feeds contain. No human at Plaintiffs' companies knows or can predict what Plaintiffs' algorithms will "say." Even Plaintiffs as corporate entities as a whole do not know what their algorithms will "say"—as evidenced by the fact that their algorithms will often "say" things that the corporate entities profess to want to *not* "say," such as feeds with posts promoting suicide, self-harm, or other content that violates Plaintiffs' guidelines. *See supra* pp. 13–18. And of course, the algorithms themselves do not know, and have no concept of, the actual content of the feeds they are generating or what they are "saying." *See supra* p. 12. The First Amendment has never extended to protect speech where the speaker does not even know what he or she is saying. *See* Austin & Levy 11–26. It is nonsensical to speak of a party exercising "editorial judgment" where the "editor" does not even know what he or she is editing. As no actual human being knows what Plaintiffs' machine-learning algorithms are "saying" or why they "say" what they do, the First Amendment does not protect their output. *See id.* at 67–83.

Plaintiffs compare personalized feeds with librarians recommending books to patrons. But they analogize themselves to the wrong party: the actual human beings that make up Plaintiffs' corporate entities are not the librarians in this scenario—they are the *patrons* (seeking books on behalf of third parties, their users), who are not exercising any editorial judgment of their own and instead are outsourcing the decision-making to the librarians. And the "librarians" in this scenario are not Plaintiffs or any of the actual human beings that make up Plaintiffs' corporate entities, but rather are non-human algorithms, which do not even understand the concept of *what a book is* in any meaningful, human way, let alone the actual content of the metaphorical books (their feeds) that they are recommending. *See supra* p. 12 (algorithms have no understanding of content). Whereas actual human librarians know with certainty what they are saying and what recommendations they are making, no human actor can predict with meaningful certainty what Plaintiffs' algorithms will "say." *See supra* pp. 13–18. And more broadly, unlike Plaintiffs' feed-personalization algorithms, whose programming directs them to feed content (regardless of its quality) that will keep users engaged on Plaintiffs' platforms for as long as possible so as to maximize profits, librarians do not merely "feed" content to children in a deadpan manner. Rather, they draw on their own education, training, and life experiences to express subjective

editorial judgments about the quality of books' content, and can explain their recommendations or otherwise assist children in utilizing those books—something Plaintiffs' algorithms cannot do.

Plaintiffs raise two fallback arguments, neither of which persuades.

First, Plaintiffs argue that their feed algorithms do not respond *solely* to how their users act online, because they also have algorithms that implement their Community Standards and other user-agnostic content-moderation decisions, in addition to their user-specific feed-personalization decisions. Meta MPI 11 n.2; TikTok MPI 13–14; YouTube MPI 13. But as the Supreme Court has explained after *Moody*, "[w]hen speech has both protected and unprotected features, . . . the unprotected features of the speech are, despite their communicative character, essentially a 'nonspeech' element for purposes of the First Amendment." *Free Speech Coal. v. Paxton*, 606 U.S. 461, 492 (2025) (quotation marks and brackets omitted). Even if Plaintiffs may be engaged in expressive activity when implementing their Community Standards or other user-agnostic decisions (activity that SB976 does *not* regulate), this does not mean that their *feed-personalization* algorithms are engaged in expressive activity, or that SB976's regulation of those feed-personalization algorithms implicates the First Amendment. *Id.*; *see Moody*, 603 U.S. at 732 n.4 ("Of course, an entity engaged in expressive activity when performing one function may not be when carrying out another."); *Green v. U.S. Dep't of Justice*, 392 F. Supp. 3d 68, 91 (D.D.C. 2019) (*Green I*) ("Even when conduct or activity is sufficiently expressive to be appropriately designated as 'speech' that implicates First Amendment rights, that does not mean that that expressive activity cannot be bifurcated into separate speech and non-speech components."), *aff'd*, 111 F.4th 81 (D.C. Cir. 2024) (*Green II*). SB976 targets only Plaintiffs' non-expressive feed-personalization activities, while leaving unregulated Plaintiffs' ability to implement their Community Standards or any other "user-agnostic judgments about what kinds of speech, including what viewpoints, are not worthy of promotion." *Moody*, 603 U.S. at 736 n.5. Indeed, as this Court has already correctly suggested, SB976's targeted approach in regulating only the feed-personalization algorithms does not restrict expressive activity:

> Regulating feeds that use algorithms mostly relying on non-expressive factors may not trigger First Amendment scrutiny at all because doing so only incidentally burdens any expressive component of those algorithms. Or, if it is very easy to

> separate an algorithm's expressive content moderation functions from non-expressive user-activity-based functions, a law prohibiting personalized feeds from relying on user activity information may also only incidentally burden speech. A covered company could just remove the user activity factors from the recommendation algorithms driving its media feeds. This latter possibility is especially significant here because SB 976 targets only recommendations based on user information. It does not prohibit covered entities from incorporating their content moderation guidelines into their recommendation algorithms.

NetChoice I, 761 F. Supp. 3d at 1223.

Second, TikTok and YouTube attempt to reframe the First Amendment question in terms of their users' rights to receive addictive feeds, as opposed to the companies' right to disseminate addictive feeds. TikTok MPI 17; YouTube MPI 14–16. But "no precedent exists for the proposition that the listener's rights are greater than those of the speaker"; rather, "the listener's rights [] derive from the existence of the speaker's rights." *NAACP v. Jones*, 131 F.3d 1317, 1322 & n.5 (9th Cir. 1997). "The First Amendment protects a right to read, but it does not grant unimpeded access to every reading material a reader might wish for." *Green II*, 111 F.4th at 98 (rejecting argument that readers have a First Amendment "right to access and learn from digital works" such that statute that places access restrictions on digital works is unconstitutional). For example, the Supreme Court has upheld restrictions on the speech of students at school based on the theory that students do not have full First Amendment rights, regardless of whether listeners might have wanted to receive that speech. *See, e.g.*, *Morse v. Frederick*, 551 U.S. 393, 404–05 (2007). Listeners do not have any greater right vis-à-vis the computer-generated output of Plaintiffs' feed-personalization algorithms than they do vis-à-vis actual human student speakers.

In short, addictive personalized feeds generated by machine-learning algorithms lack any meaningful human editorial judgment, to the point that humans do not even know what the algorithms are "saying." Accordingly, those feeds are not protected speech, and SB976's regulation of Plaintiffs' addictive, personalized feeds do not implicate the First Amendment.

2. Additionally, Because the Addictive Feeds Consist of Third-Party User Posts Compiled to Reflect Users' Perceived Interests Back to Them, They Do Not Convey a Protected Expressive Message of Plaintiffs' Own

The fact that Plaintiffs' addictive feeds are machine-generated and lack meaningful human editorial judgment is alone sufficient to render Plaintiffs' feeds non-protected under the First

Amendment. But additionally, Plaintiffs' feeds are non-protected for another reason: they are comprised exclusively of compilations of third-party posts that reflect users' preferences and do not convey any expressive message of Plaintiffs' own.

The First Amendment "bears less heavily when speakers assert the right to make other people's speeches." *Project Veritas*, 125 F.4th at 955. Not all compilations of third-party speech are expressive or protected by the First Amendment. "Instead, the First Amendment protects only those compilations that are 'inherently expressive' in their own right, meaning that they select and present speech created by other persons in order to spread the compiler's *own* message." *Moody*, 603 U.S. at 781 (Alito, J., concurring in the judgment) (emphasis added, quotation marks and brackets omitted) (citing *FAIR*, 547 U.S. at 66); *NetChoice I*, 761 F. Supp. 3d at 1219. "[T]he message that a compiler seeks to send need not be particularly articulate, but to trigger the First Amendment, there must clearly be a message of some sort." *NetChoice I*, 761 F. Supp. 3d at 1221 (citing *Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., Inc.*, 515 U.S. 557, 574 (1995)). "[R]estrictions on a private speaker's ability to compile and organize third-party speech implicate speech rights only if those restrictions impair the speaker's *own* expression." *Id.* at 1219 (emphasis added) (citing *Moody*, 603 U.S. at 730–31, 740). By contrast, algorithms that just "reflect users' revealed preferences to them" and "merely 'giv[e] them the content they appear to want,' probably [are] not expressive." *NetChoice II*, 152 F.4th at 1014 (quoting *Moody*, 603 U.S. at 736 n.5; *id.* at 746 (Barrett, J., concurring)). As this Court explained, "[i]f a third party's interpretations triggered the First Amendment, essentially everything would become expressive and receive speech protections," in violation of the Supreme Court's dictates that "there is not 'limitless variety of conduct [that] can be labeled "speech" even when the person engaging in the conduct thereby intends to express an idea.'" *NetChoice I*, 761 F. Supp. 3d at 1222 (quoting *United States v. O'Brien*, 391 U.S. 367, 391 (1968)).

The Supreme Court's decision in *FAIR* is instructive. There, the Court rejected free-speech claims brought by an association of law schools and faculty against a law requiring them to include military recruiters in on-campus recruiting events (or lose federal funding). *FAIR*, 547 U.S. at 52–54. In holding that law schools' decisions to allow recruiters on its campus "is not

inherently expressive," the Court explained that "[l]aw schools facilitate recruiting to assist their students in obtaining jobs" and a requirement to include military recruiters "does not sufficiently interfere with any message of the school." *Id.* at 64. The Court recognized that providing recipients simply what the *recipients* might be interested in (whether recruiters that might offer recipients a job, or internet posts that recipients might find interesting) does not convey a message of the *speaker* and "is not inherently expressive." *Id.*; *see Moody*, 603 U.S. at 740 (discussing how in *FAIR*, "the host of the third-party speech was not itself engaged in expression").

Plaintiffs here do not claim that their feeds convey Plaintiffs' *own* expression. *Cf. Stockton*, 152 F.4th at 1144–45 (First Amendment plaintiff bears the burden of identifying "what [it] wishes to say on [] topics"). Instead, Plaintiffs assert only that the message that their feeds convey to users is, *we think you will be interested in this feed*. Meta MPI 6; TikTok MPI 15; YouTube MPI 9. In other words, Plaintiffs' feeds merely reflect users' perceived interests back to them.

That Plaintiffs do not claim their feeds convey any expression of Plaintiffs' own likely is not accidental. By not having any message of their own, Plaintiffs can appear neutral and target the widest possible audience and try to be all things to all users. *See generally* Narayanan 33–35. Additionally, and perhaps far more significantly, Plaintiffs avoid conveying a message of their own due to Section 230 of the Communications Decency Act, which (broadly speaking) grants a defense to internet companies with respect to claims to hold them responsible as a publisher or speaker of information on their platforms provided by third parties (e.g., user posts)—but only insofar as they are not themselves "responsible, in whole or in part, for the creation or development of [the] information" being provided. 47 U.S.C. § 230(c), (f)(3); *see Fair Hous. Council v. Roommates.com*, 521 F.3d 1157, 1162–63 (9th Cir. 2008) (en banc). If Plaintiffs are compiling user posts into feeds that convey an expressive message of Plaintiffs' own, Plaintiffs are creating content and cannot avail themselves of Section 230. *See Roommates.com*, 521 F.3d at 1169–70 (holding that internet platform did not benefit from Section 230 for its involvement in filtering process that showed some user posts to some users but not to others based on discriminatory criteria). So Plaintiffs conspicuously avoid claiming that their feeds convey any message of Plaintiffs' own, and instead say only that their feeds reflect their *users*' interests.

But Plaintiffs cannot have it both ways. *See Doe ex rel. Roe v. Snap, Inc.*, 144 S. Ct. 2493, 2494 (2024) (Thomas, J., dissenting from cert. denial) ("In the platforms' world, they are fully responsible for their websites when it results in constitutional protections, but the moment that responsibility could lead to liability, they can disclaim any obligations and enjoy greater protections from suit than nearly any other industry."). If Plaintiffs claim only that their feeds just reflect users' interests back to them and do not assert any message of Plaintiffs' own, then their feeds are not inherently expressive, and thus are not protected under the First Amendment. *NetChoice I*, 761 F. Supp. 3d at 1222; *NetChoice II*, 152 F.4th at 1014; *accord, e.g.*, *State ex rel. Jackson v. TikTok Inc.*, No. 24CV032063-910, 2025 WL 2399525, at *10 (N.C. Super. Ct. Aug. 19, 2025) (where "[t]he algorithm does not convey a message by its programmer" and "simply bows to user preferences and propensities," "it's hard to discern any expressive activity.").

Plaintiffs' feed personalization does not give user feeds a "particular expressive quality." *Cf. Moody*, 603 U.S. at 738. Once user posts make it through Plaintiffs' content-moderation processes, Plaintiffs' feed-personalization algorithms can place any posts in any user's feed if they "think" those posts will engage that user. Feeds that can potentially encompass *anything* do not have a "particular" expressive quality. *See id.* at 736 n.5 (declining to hold that feeds "without any regard to independent content standards" are expressive).[8] For example, Plaintiffs' algorithms will amplify posts that Plaintiffs claim to disagree with, such as posts encouraging suicide or self-harm. *See supra* pp. 14–18. Algorithms that "would recommend and amplify both favored and disfavored messages alike" do not have a "particular expressive quality" or convey an expressive choice. *NetChoice I*, 761 F. Supp. 3d at 1222; *cf. Hurley*, 515 U.S. at 574 (a parade is expressive because "the [organizer] selects the expressive units of the parade from potential participants" and "each contingent's expression *in the [organizer]'s eyes* comports with what

[8] Plaintiffs' *content moderation*—which, again, is not affected by SB976—may contain a "particular expressive quality." For example, applying Community Standards prohibiting pornography will mean feeds will not be XXX-rated, which may be a particular quality. But Plaintiffs' *feed personalization*, by contrast, does not do this—it allows for any feed of content, regardless of the quality of that content, to potentially be presented to a user.

merits celebration on that day") (emphasis added).

Plaintiffs argue that "entities engaged in protected First Amendment activity do not lose this protection simply because one of the goals of that activity may be to please an audience or increase future engagement with their expressive products." Meta MPI 15–16; *accord* TikTok MPI 13; YouTube MPI 13–14. Meta, for example, argues that the First Amendment would not allow a "mandate that authors not end chapters on cliffhangers." Meta MPI 16. But whether entities engaged in expressive activity (e.g., writing a book) can take their audience's reaction into account, does not prove the converse—that by taking their audience's reaction into account, entities necessarily engaged in expressive activity—particularly where the entities are solely compiling third-party speech, as opposed to engaging in pure speech of their own (like writing a book). For example, a compiler of third-party user posts that simply sorts posts chronologically has not engaged in expressive activity—even if the reason the compiler chooses to do so is because it thinks its audience wants chronological sorting—because the compiler has not expressed any message of its own, and simply "'giving [an audience] the content they appear to want,' probably is not expressive." *NetChoice II*, 152 F.4th at 1014; *see Moody*, 603 U.S. at 736 n.5; *id.* at 783 (Alito, J., concurring in the judgment) ("[C]ompilations that organize the speech of others in a non-expressive way (e.g., chronologically) fall 'beyond the realm of expression.'") (brackets omitted) (quoting *Hurley*, 515 U.S. at 569); *Mapother v. Dep't of Justice*, 3 F.3d 1533, 1540 (D.C. Cir. 1993) (a chronological organization "reflects no point of view"). Likewise, a compiler that simply sorts posts by overall popularity has not engaged in expressive activity, even if the compiler chooses to do so because it thinks its audience wants popularity sorting. Likewise, a compiler (like Plaintiffs here) that sorts posts by personalized, rather than overall, popularity (algorithmically-calculated numeric scores representing the probability that individual users will engage with the posts) has not engaged in expressive activity. In each of these cases, the compiler has expressed no message of its own beyond simply "reflect[ing] users' revealed preferences to them"—which is "probably is not expressive." *NetChoice II*, 152 F.4th at 1014; *accord NetChoice I*, 761 F. Supp. 3d at 1222; *Jackson*, 2025 WL 2399525, at *10.

Plaintiffs cite *Sorrell v. IMS Health Inc.*, 564 U.S. 552 (2011), to argue that speakers can

use information they have gathered about their listeners to tailor their messages to each listener. Meta MPI 13–14; TikTok MPI 13; YouTube MPI 14. In *Sorrell* the Supreme Court discussed how pharmaceutical sellers use information about doctors' prescribing habits to shape their messages to each doctor. *Sorrell*, 564 U.S. at 578. But that is not the issue here—whether a speaker can use information about its audience to tailor its message is a different question from whether the speaker has a message to begin with. The sellers in *Sorrell* had a clear message to doctors: *you should use our brand-name drugs*. The sellers did not merely reflect the interests of their audience back to them—to the contrary, the whole point was that the sellers wanted to convey their *own* message to *change* their audience's mind and get their audience to use more of their drugs over generic alternatives. *See id.* at 575–79. Here, Plaintiffs do the opposite, asserting only that their feeds show the user what the user is interested in, without claiming any message of Plaintiffs' own. This is not an expressive compilation and does not constitute protected speech.

B. Even If Plaintiffs' Addictive Feeds Were Protected Speech, Plaintiffs Have Not Met Their Burden of Showing That SB976 Is Likely Unconstitutional

Even if the Court were to conclude that the addictive-feed output of Plaintiffs' machine-learning feed-personalization algorithms were protected under the First Amendment, SB976 would be subject to at most intermediate scrutiny and is constitutional under that standard.

1. SB976 Is Content-Neutral and Subject to At Most Intermediate Scrutiny

Where, as here, "the primary distinction the statute makes does not depend on the message that might be conveyed," the statute is content-neutral and is subject to only intermediate scrutiny. *See, e.g.*, *Porter v. Martinez*, 68 F.4th 429, 441–43 (9th Cir. 2023). SB976 makes no distinction based on the message that might be conveyed. Indeed, to determine whether Plaintiffs' feeds comply with or violate SB976, one need not examine the actual content of any feed; rather, one need only look to whether Plaintiffs' algorithms use "information provided by the user, or otherwise associated with the user or the user's device," as one of their inputs in selecting what posts to display in that feed. § 27000.5(a). This is content-neutral. *See Porter*, 68 F.4th at 441–43 (where officials do not need to examine the content of communication and need observe only the surrounding circumstances to assess whether communication complies with

statute, statute is content-neutral and subject to only intermediate scrutiny).

The Ninth Circuit has already held that SB976 is content-neutral, other than with respect to its like-count provision (§ 27002(b)(3)), which is not at issue in these cases. *NetChoice II*, 152 F.4th at 1015. As the Court held, SB976 "regulates 'communication of all kinds,' indiscriminate of topic or content." *Id.* at 1016. The Court specifically rejected arguments that SB976's treatment of websites "limited to commercial transactions or to consumer reviews," § 27000.5(b)(2)(A), rendered SB976 content-based. *Id.* at 1015–16. The Court further specifically rejected arguments that SB976's reference to "a social media platform," § 27000.5(b)(1), rendered SB976 content-based—as the Court noted, SB976 used the term "'social media' platform" as a shorthand, but the actual Act "applies to websites whether they facilitate social interaction or other forms of content" and thus is not content-based. *Id.* at 1016.

Plaintiffs each nonetheless try to argue that SB976 is subject to strict scrutiny—each without mentioning *NetChoice II* in their respective sections. *See* Meta MPI 16–19; TikTok MPI 18–21; YouTube MPI 16–18. Their arguments fail to overcome *NetChoice II* or other precedent.

To start, Meta argues that SB976 "expressly singles out 'social media platform[s]' for regulation." Meta MPI 18. That is flatly wrong. As the plain language of SB976 makes clear, SB976 "includ[es], but [is] *not limited to*," social media platforms and does not single them out. § 27000.5(b)(1) (emphasis added); *see NetChoice II*, 152 F.4th at 1016.[9]

[9] Meta mischaracterizes Senator Skinner's remarks on SB976. Meta MPI 19 n.5 (citing *Sen. Floor Session* (Cal. May 20, 2024), https://tinyurl.com/bdej7cnr). Senator Skinner did not say that SB976 was "*de facto* 'directed' against 'social media'" or targeted at Facebook or any particular speaker or content. *Contra* Meta MPI 19 n.5. Rather, she discussed how SB976 "focus[es] not on content" but instead "regulates the addictive features embedded in these platforms" and used Facebook's original chronological feed as an example of what would constitute a non-addictive feed. *Sen. Floor Session* 2:21:30–2:21:47, 2:26:36–2:27:01; *see generally id.* 2:20:42–2:27:56. Meta also repeatedly mischaracterizes an Assembly Committee hearing report. Meta MPI 19, 21, 23 (citing Meta MPI Ex. 12). The report does not say that "the

Meta and TikTok both argue that SB976 is speaker-based because it distinguishes between compilations of third-party speech (posts "generated or shared by users," § 27000.5) and first-party speech and then regulates only the former. Meta MPI 18; TikTok MPI 19. But the First Amendment itself distinguishes between third-party and first-party speech. *See Project Veritas*, 125 F.4th at 955 (First Amendment "bears less heavily when speakers assert the right to make other people's speeches"); *cf.* 47 U.S.C. § 230 (also distinguishing between third-party and first-party speech). In any event, the law is clear that "[a] speaker preference is problematic only if it reflects a content preference." *NetChoice II*, 152 F.4th at 1017. No distinction in SB976 between compilations of third-party user posts and first-party speech reflects any content preference.[10]

Meta and TikTok both claim that SB976's regulation of their addictive feeds represents a break from our Nation's "history and tradition." Meta MPI 17–18; TikTok MPI 19–20. But they fail to explain how imposing content-neutral regulations on conduct that involves no meaningful expressive human decision-making represents a break from "history and tradition." To the contrary, there is a long history and tradition whereby non-expressive conduct is deemed outside the protection of the First Amendment, *see, e.g.*, *City of Dallas v. Stanglin*, 490 U.S. 19, 24–25 (1989), and content-neutral regulations are subject to only intermediate, not strict, scrutiny, *see, e.g.*, *Hill v. Colorado*, 530 U.S. 703, 719–25 (2000). Meta's and TikTok's bare invocation of "history and tradition" fails to rebut that at most intermediate scrutiny should apply to a content-neutral law like SB976. *Cf. NetChoice II*, 152 F.4th at 1015.

---

personalized-feed restrictions are 'about [the] delivery of user-created content,' which they describe as 'harmful material to youth.'" *Contra* Meta MPI 19 (brackets in Meta's original). What it actually says is that SB976 "is *only* about *algorithmic* delivery of user-created content" and that children need "*protection from the harms of addictive feeds and algorithms, which algorithms rely on addicting children and which can send extremely* harmful material to youth." Meta MPI Ex. 12 at 11 (language omitted by Meta emphasized).

[10] Meta also argues that SB976 is speaker-based because it exempts commercial-transaction and consumer-review websites. Meta MPI 20 n.8. Meta fails to mention that the Ninth Circuit has already expressly rejected that argument. *NetChoice II*, 152 F.4th at 1015–16.

YouTube takes a different tack, arguing that SB976 lacks content-neutrality because it "defin[es] regulated speech by its function or purpose." YouTube MPI 17 (citing *Reed v. Town of Gilbert*, 576 U.S. 155, 163–64 (2015)). But this argument "stretches *Reed*'s 'function or purpose' language too far.'" *City of Austin v. Reagan Nat'l Advert. of Austin, LLC*, 596 U.S. 61, 74 (2022). *Reed* did not hold "that any classification that considers function or purpose is *always* content based." *Id.* (emphasis in original). Rather, it holds only that "subject-matter discrimination is facially content based," whether "overt" or "subtler forms of discrimination that achieve identical results based on function or purpose[.]" *Id.*; *accord NetChoice I*, 761 F. Supp. 3d at 1224 ("[I]t is not quite accurate to say that all content-*based* laws face strict scrutiny. Rather, it is content-*discriminatory* laws that face strict scrutiny.") (emphasis in original). SB976 does not discriminate based on subject matter. Rather, it targets only the mechanism by which content is presented—i.e. personalized, addictive, algorithmically-driven feeds. SB976 is therefore content-neutral—as the Ninth Circuit has already held. *NetChoice II*, 152 F.4th at 1015.

Finally, YouTube claims that a hypothetical law prohibiting librarians from recommending books to patrons based on that patron's interest "would plainly be content-based, even though that restriction would not single out books about a particular topic." YouTube MPI 18. But YouTube provides no citation for this assertion. Moreover, a librarian's book recommendations are in actuality an expression of his or her own editorial judgment, which, as YouTube recognizes, includes "the librarian's views about the quality of the books and their fit for the patron." *Id.* SB976 does not restrict Plaintiffs' ability to express their *own* views about the quality of content in their personalized feeds—content of which Plaintiffs have effectively disclaimed endorsement. *See supra* pp. 36–37 (citing 47 U.S.C. § 230). Rather, as explained above, SB976 regulates only the mechanism by which material is presented in addictive feeds, and it does so evenhandedly, regardless of that material's content. This is similar to, for example, regulations on solicitation, which are content-neutral (even if one must read the content of a message to determine whether it *is* a solicitation) because they "do not inherently present 'the potential for becoming a means of suppressing a particular point of view,' so long as they do not discriminate based on topic, subject matter, or viewpoint." *City of Austin*, 596 U.S. at 72; *Porter*, 68 F.4th at 442.

Because SB976 is content-neutral, it is subject to at most intermediate scrutiny.

2. SB976 Satisfies Intermediate Scrutiny

To satisfy intermediate scrutiny, a law must (1) "serve a significant governmental interest" and (2) must be narrowly tailored while "leav[ing] open ample alternative channels for communication of the information." *Project Veritas*, 125 F.4th at 952. The law "need not be the least speech-restrictive means of advancing a governmental interest." *Id.* at 956 (quotation marks omitted). Additionally, "[i]t is well established that an alternative channel need not be ideal, but merely adequate." *Id.* at 957. Courts "generally will not strike down a governmental action for failure to leave open ample alternative channels unless the government enactment will foreclose an entire medium of public expression across the landscape of a particular community or setting." *Id.* (ellipsis omitted). As applied to Plaintiffs, SB976 satisfies intermediate scrutiny here.

a. SB976 Serves a Significant Governmental Interest

As this Court has previously found, SB976 serves "an important government interest: the protection of children's health." *NetChoice I*, 761 F. Supp. 3d at 1227; *see also NetChoice II*, 152 F.4th at 1017 (affirming that State's interest in "protecting minors' mental health by reducing screentime and habit-forming platform usage" supported SB976's private-mode default requirements); *Sable Commc'ns of Cal., Inc. v. FCC*, 492 U.S. 115, 126 (1989) (government has "a compelling interest in protecting the physical and psychological well-being of minors").

Plaintiffs argue that the State's interest in SB976 is in restricting the ideas to which children may be exposed, and thus the State's interest is "not unrelated to the suppression of free speech." Meta MPI 20–21; YouTube MPI 19. This is wrong. The State's interest in SB976 is to prevent harms to children from the *effects* of addictive feeds, not to restrict any ideas. The Supreme Court has confirmed the constitutionality of such interests. For example, in *City of Renton v. Playtime Theaters, Inc.*, 475 U.S. 41 (1986), the Court examined a law that prohibited adult theaters from locating in certain parts of the city. *Id.* at 43–44. The law singled out certain speakers (adult theaters) and did so based on their speech (adult movies). *See id.* at 47. But the city's interest was not "the *content* of the films" but "the *secondary* effects of such theaters on the surrounding community" (increased crime, lower property values, etc.). *Id.* at 47–48 (emphasis in original).

The Court held that this constituted a permissible substantial government interest that justified the city's law. *Id.* at 48–50. So too here. The State's interest is in the secondary effects of addictive feeds, e.g., sleep deprivation, poor mental health, diminished attention capacity, ADHD, anxiety, depression, and so forth. *See supra* pp. 20–25. These constitute significant governmental interests unrelated to restricting any ideas to which children may be exposed.

SB976 does not prevent any child (or adult) from searching for ideas or topics that he or she may be interested in. And conversely, SB976 regulates all addictive feeds equally, regardless of what ideas might be expressed therein. Algorithmically-generated addictive feeds full of posts promoting, for instance, the virtues of staying in school or saying "no" to drugs, are as equally subject to SB976 as addictive feeds full of posts extolling dropping out or doing drugs. That is because the State's interest behind SB976 is not in suppressing (or promoting) any idea or topic, but rather in curbing addictive internet usage by regulating the *mechanism* of addictive feeds and the use of machine-learning algorithms to feed children more and more of whatever the algorithms calculate will keep children "engaged" and on the platform for as long as possible. SB976 § 1; *accord, e.g.*, Telzer Decl. ¶ 69 (addiction is "not tied to the specific type of content on the platform but rather to the features that facilitate compulsive use"); *see id.* ¶¶ 47–49, 144–48.

Plaintiffs also argue that the State has not shown a causal connection specifically between the personalized feeds regulated by SB976 and the harms to children the State seeks to remedy. Meta MPI 21; TikTok MPI 21; YouTube MPI 18–19. But this "impose[s] on the [State] an unnecessarily rigid burden of proof." *City of Renton*, 475 U.S. at 50. In enacting SB976, the Legislature found that "the algorithmic delivery of content and other design features [] pose a significant risk of harm to the mental health and well-being of children and adolescents" and, among other things, noted a U.S. Surgeon General's report that children and teens who use social media at least three hours a day have double the risk of poor mental health outcomes as those who do not. SB976 § 1. "The First Amendment does not require a [State], before enacting such an ordinance, to conduct new studies or produce evidence independent of that already generated . . . , so long as whatever evidence the [State] relies upon is reasonably believed to be relevant to the problem that the [State] addresses." *City of Renton*, 475 U.S. at 51–52; *accord, e.g.*, *Porter*, 68

F.4th at 446 ("California does not need to produce new empirical evidence to justify [its law].").

Plaintiffs cite *Brown v. Entertainment Merchants Association*, 564 U.S. 786 (2011), to argue that the State needs to prove a direct causal link between algorithmic feeds and harms to children, because (in their view) evidence of mere correlation is insufficient. Meta MPI 21–22; YouTube MPI 18–19. But, among other things, *Brown* involved a permanent (not preliminary) injunction, entered after full discovery and summary judgment. *See Brown*, 564 U.S. at 790. Additionally, the problem with the studies in *Brown* was *not* strictly that they showed correlation instead of causation, but that they "suffer[ed] from significant, admitted flaws in methodology" while showing "at best *some* correlation" between violent video games and "minuscule real-world effects." *Id.* at 800–01 (emphasis added). And following *Brown*, the Supreme Court affirmed that even under strict scrutiny (much less intermediate scrutiny), States do not need "proof by documentary record" to establish even a compelling governmental interest. *Williams-Yulee v. Fla. Bar*, 575 U.S. 433, 447 (2015); *accord, e.g.*, *Fla. Bar v. Went for It, Inc.*, 515 U.S. 618, 628 (1995) (States can "justify speech restrictions by references to studies and anecdotes pertaining to different locales altogether, or even, in a case applying strict scrutiny, to justify restrictions based solely on history, consensus, and 'simple common sense'") (citations omitted); *French v. Jones*, 876 F.3d 1228, 1241 (9th Cir. 2017) (the State "need not present empirical evidence . . . for its rule to survive strict scrutiny"). Neither this Court nor the Ninth Circuit has required the evidentiary standard that Plaintiffs try to impose here. In fact, in upholding SB976's requirement that child accounts must be set to private mode by default, § 27002(b)(5), both this Court and the Ninth Circuit found that harm to children from excessive internet usage generally was sufficient. *NetChoice I*, 761 F. Supp. 3d at 1227; *NetChoice II*, 152 F.4th at 1017.

In any event, the State *has* presented abundant evidence demonstrating a direct causal (not merely correlative) link between addictive feeds and harm to children. Not only are social-media usage and addictive feeds linked to an array of harms to children, including sleep problems, ADHD, anxiety, and depression, they *cause* and *exacerbate* those harms. *See supra* pp. 20–25.

b. SB976 Is Narrowly Tailored

SB976 is narrowly tailored to curbing the harms of social-media addiction and excessive

use: it places limits on platforms only in connection with the addictive-feed *mechanisms* that Plaintiffs utilize to try to keep children on their platforms for as long as possible. And SB976 also leaves open ample alternative channels for communication. It does not block or restrict any ideas or viewpoints, does not prevent minors from signing up for any site, and does not bar them from viewing any content, searching for any specific media, or subscribing to any creators they independently want. It does not prevent parents from allowing their children to access personalized feeds. And it does not prevent Plaintiffs from engaging in their own first-party speech or from including or prioritizing any message of their own that they want in any feed. Despite all this, Plaintiffs attack SB976 as underinclusive and overinclusive. Both attacks fail.

Regarding underinclusiveness, the Supreme Court has held that "the First Amendment imposes no freestanding 'underinclusiveness limitation.'" *Williams-Yulee*, 575 U.S. at 449. Nevertheless, Plaintiffs claim that SB976 is underinclusive because it only regulates platforms that compile feeds of third-party content, as opposed to platforms with feeds of the platforms' own first-party speech. Meta MPI 22–23; TikTok MPI 23; YouTube MPI 21. But as discussed above, the First Amendment itself draws a distinction between a speaker's own speech and "other people's speeches." *Project Veritas*, 125 F.4th at 955. In any event, "[a] State need not address all aspects of a problem in one fell swoop; policymakers may focus on their most pressing concerns." *Williams-Yulee*, 575 U.S. at 449; *accord, e.g.*, *City of Renton*, 475 U.S. at 52–53. Plaintiffs also claim SB976 is underinclusive because it allows minors to continue to access addictive feeds if their parents grant approval, rather than banning addictive feeds outright. Meta MPI 22; YouTube MPI 20. But as the Supreme Court has held, "[w]e will not punish [a State] for leaving open more, rather than fewer, avenues of expression, especially when there is no indication that the selective restriction of speech reflects a pretextual motive." *Williams-Yulee*, 575 U.S. at 452. The fact that SB976 allows for parental opt-ins does not suggest a pretextual motive—rather, it reflects the State's effort to limit SB976 so it does not go beyond what is necessary to effectuate its purpose. After all, it is "counterintuitive to argue that a law violates the First Amendment by abridging *too little* speech." *Id.* at 448 (emphasis in original). Rather, courts "uph[o]ld laws—even under strict scrutiny—that conceivably could have restricted even

greater amounts of speech in service of their stated interests." *Id.* at 449.

Regarding overinclusiveness, Plaintiffs claim that they have adopted tools to protect children and that parents also have third-party tools to help them oversee their children online. Meta MPI 23; TikTok MPI 22; YouTube MPI 19–20. To the extent that those tools complement SB976 and its goal of protecting children's health, the Attorney General encourages them. But that does not mean that they are "less restrictive alternative[s]" or necessarily "serve the Government's purpose." *United States v. Playboy Ent. Grp., Inc.*, 529 U.S. 803, 813 (2000). Plaintiffs' tools do not include an option to disable their addictive-feed mechanisms.[11] The tools Plaintiffs do offer are cumbersome, impractical, and at best address only platform usage generally, not the pernicious effect of Plaintiffs' addictive-feed algorithms. *See* Golbeck Decl. ¶¶ 69–80; Hilbert Decl. ¶¶ 95–103. And the third-party tools Plaintiffs reference, such as site-blockers to block minors' access to Plaintiffs' platforms (or the internet as a whole), are *more* restrictive than SB976, because rather than allowing children to access speech (i.e., the user posts on Plaintiffs' platforms) without Plaintiffs' addictive-feed mechanisms, they block access altogether. Regulation like SB 976 that specifically targets addictive-feed *mechanisms*, while leaving open access to speech, is narrowly tailored and not overinclusive.

Plaintiffs also argue that SB976 is overinclusive because it applies to all children without distinguishing between different ages or each child's individual level of internet addiction. Meta MPI 24; YouTube MPI 21. But the State's interest in safeguarding the mental and physical health of minors applies as much to 17-year-olds as it does 13-year-olds. *See*, *e.g.*, Radesky Decl. ¶¶ 25 et seq. (discussing harms to adolescents up to and beyond age 17); Christakis Decl. ¶¶ 65 et seq. (same); Telzer Decl. ¶¶ 44 et seq. (same); Twenge Decl. ¶¶ 8 et seq. (same); *cf. Morse*, 551 U.S. at 407 (state's interest in curbing drug use extended to 12th graders). SB976 would not be properly tailored to achieve the State's interest if it did not cover all children in California.

[11] Some Plaintiffs appear to offer options to disable their addictive-feed mechanisms to users in Europe—but not in the United States. *See* Ex. 1 at 107. If they do offer such an option in the United States (which they do not assert in their filings), the option is so deeply obscured so as to be unavailable as a practical matter to users. *See* Hilbert Decl. ¶¶ 95–96.

3. Even If Strict Scrutiny Applied, SB976 Satisfies Strict Scrutiny

Even assuming (counterfactually) that strict scrutiny applied, SB976 is constitutional. "Under strict scrutiny, restrictions may be justified only if the government proves that they are narrowly tailored to serve compelling state interests." *In re Nat'l Sec. Letter*, 33 F.4th 1058, 1070 (9th Cir. 2022). Restrictions need only "be narrowly tailored," not "perfectly tailored." *Id.* at 1073. "Accordingly, a reviewing court should decline to wade into the swamp of calibrating the individual mechanisms of a restriction." *Id.* (quotation marks and brackets omitted). "Ultimately, the Supreme Court has emphasized that strict scrutiny is not 'fatal in fact.'" *Id.*

SB976 satisfies strict scrutiny here. The State's interest in protecting children from the harm of internet addiction and addictive feeds is not only significant, it also is compelling. *See, e.g.*, *Sable*, 492 U.S. at 126 (government interest in protecting physical and psychological well-being of minors is "compelling"). Plaintiffs' evidentiary arguments fail—as discussed above, the State "need not present empirical evidence . . . for its rule to survive strict scrutiny," *French*, 876 F.3d at 1241, and in any event, the State has presented abundant evidence, *see supra* pp. 20–25. Further, as discussed above, SB976 is narrowly tailored. *Cf. French*, 876 F.3d at 1237–41.

C. YouTube Has Not Established That SB976 Is Unconstitutionally Vague

YouTube alone argues that SB976's provision for default settings that can be changed by a "verified parent," § 27002(b), is vague. YouTube MPI 23–24. But its arguments are unavailing.

A statute is impermissibly vague only if it "fails to provide a reasonable opportunity to know what conduct is prohibited, or is so indefinite as to allow arbitrary and discriminatory enforcement." *Arce v. Douglas*, 793 F.3d 968, 988 (9th Cir. 2015). "[T]he vagueness doctrine does 'not doubt the constitutionality of laws that call for the application of a qualitative standard to real-world conduct; the law is full of instances where a man's fate depends on his estimating rightly some matter of degree.'" *United States v. Bronstein*, 849 F.3d 1101, 1108 (D.C. Cir. 2017) (nested quotation marks and ellipses omitted).

In *Matsumoto v. Labrador*, 122 F.4th 787 (9th Cir. 2024), the plaintiffs challenged as vague a statute prohibiting "recruiting" a pregnant minor with the intent to conceal the abortion from the minor's parents or guardian. *Id.* at 794–95. The statute did not define "recruiting," so the Ninth

Circuit applied the ordinary meaning of "recruit": "to seek to persuade, enlist, or induce someone to join an undertaking or organization, to participate in an endeavor, or to engage in a particular activity or event." *Id.* at 808. The parties noted that recruiting "may" or "might" encompass a number of different activities ("'recruiting' may [] include subsidizing or fully funding," "[l]egal advice, too, might constitute recruiting," etc.). *Id.* at 809–10. But the Court rejected the plaintiffs' argument that "recruiting" was unconstitutionally vague, holding that it was "an imprecise but comprehensible normative standard" and was not so open-ended so as to specify "no standard of conduct at all." *Id.* (ellipsis omitted); *accord, e.g.*, *Cal. Tchrs. Ass'n v. State Bd. of Educ.*, 271 F.3d 1141, 1153–54 (9th Cir. 2001) (citing cases rejecting vagueness challenges to terms like "approach," "consent," and "reasonably believe").

YouTube's claim here that "verified" is vague likewise fails. Like "recruiting," "verified" has an ordinary plain meaning: to establish the truth, accuracy, or reality of something. Merriam Webster Dictionary, *Verify* (2026), https://tinyurl.com/5yv5pw2p. That there may at times be questions about application of that standard does not mean that there is "no standard of conduct at all" and does not render the statute unconstitutionally vague. *Matsumoto*, 122 F.4th at 805 (ellipsis omitted); *see Bronstein*, 849 F.3d at 1107 (fact that term "does not mean the same thing to all people, all the time, everywhere" and "may vary depending upon whom you ask" does not render statute vague) (brackets omitted); *Cal. Tchrs.*, 271 F.3d at 1151–52 (generally same).

II. PLAINTIFFS HAVE NOT MET THEIR BURDEN TO ESTABLISH IRREPARABLE HARM

Meta and TikTok assert that they have established irreparable harm merely by bringing a First Amendment claim. Meta MPI 24; TikTok MPI 24. But as discussed above, Plaintiffs have failed to establish a likelihood of success on the merits of their First Amendment claim and thus have failed to establish irreparable injury. Furthermore, Plaintiffs do not assert or put forth evidence of any idea, thought, opinion, or viewpoint that SB976 prevents them from expressing, nor do they identify any human being whose speech or expression has been affected by SB976. As applied to Plaintiffs here, SB976 at most places restrictions on output generated by machine-learning algorithms, which do not even understand the content of the output they are generating. Plaintiffs thus have not established irreparable harm to any actual human being.

YouTube additionally argues that it will be forced to expend some unspecified amount of money or resources to comply with SB976. YouTube MPI 25. But YouTube does not provide even an estimate, much less proof, of this supposed cost. "Speculative injury does not constitute irreparable injury." *Goldie's Bookstore, Inc. v. Super. Ct.*, 739 F.2d 466, 472 (9th Cir. 1984); *see also NetChoice I*, 761 F. Supp. 3d at 1223 (noting how "[a] covered company could just remove the user activity factors from the recommendation algorithms driving its media feeds"). In any event, "ordinary compliance costs are typically insufficient to constitute irreparable harm." *Council for Responsible Nutrition v. James*, 159 F.4th 155, 171 (2d Cir. 2025).

III. PLAINTIFFS HAVE NOT MET THEIR BURDEN TO ESTABLISH THAT THE BALANCE OF EQUITIES OR PUBLIC INTEREST WEIGH IN THEIR FAVOR

"[S]imply raising a serious First Amendment claim is not enough to tip the hardship scales" in favor of plaintiffs. *Stormans, Inc. v. Selecky*, 586 F.3d 1109, 1138 (9th Cir. 2009) (brackets omitted). Among other things, courts should take into account the effect of an injunction on third parties, including on "the health of state residents." *Id.* at 1139 (quotation marks omitted). Enjoining SB976 would prevent the government from pursuing its paramount public interest in protecting children's health. *See Sable*, 492 U.S. at 126. As explained above, there is an abundance of evidence demonstrating a causal link between Plaintiffs' addictive feeds and an ongoing array of harms to children. *See supra* pp. 20–25. SB976's regulation of addictive feeds would begin to alleviate those harms and lead to improvements in children's mental health and well-being. *See supra* p. 25. Conversely, enjoining SB976 would inflict irreparable harm on California by preventing enforcement of a statute enacted by the legislative representatives the people elected to protect their children. *See Am. Fed'n of Gov't Emps., AFL-CIO v. Trump*, 148 F.4th 648, 656 (9th Cir. 2025) ("[A]ny time a State is enjoined by a court from effectuating statutes enacted by representatives of its people, it suffers a form of irreparable injury."); *Stormans*, 586 F.3d at 1140 ("court should give due weight to the serious consideration of the public interest in this case that has already been undertaken by the responsible state officials").

CONCLUSION

The Court should deny Plaintiffs' MPIs.

Dated: February 20, 2025

Respectfully submitted,

ROB BONTA
Attorney General of California
TODD GRABARSKY
LARA HADDAD
Supervising Deputy Attorneys General

s/Shiwon Choe
ASHA ALBUQUERQUE
WILLIAM BELLAMY
SHIWON CHOE
CHRISTOPHER J. KISSEL
SAMUEL SOKOLSKY
MEGHAN H. STRONG
ZELDA VASSAR
Deputy Attorneys General
Attorneys for Defendant Rob Bonta, in his official capacity as Attorney General of California