Before the
FEDERAL TRADE COMMISSION
Washington, DC 20580

Commercial Surveillance ANPR, R111004	)	Comments from Public Knowledge and the Yale Law School Tech Accountability & Competition Project[1], a division of the Media Freedom & Information Clinic[2]
	)	
	)	
Docket Number: FTC-2022-0053-0001	)	
	)	

[1] Authors: Sara Collins, Nicholas Garcia, Alex Petros, and Eva Rhule of Public Knowledge, as well as Kart Kandula and Eleanor Runde of the Yale Law School Technology Accountability & Competition Project. Special Thanks to Professor Arvind Narayanan of Princeton University, a leading scholar exposing the frequent mismatch between the promise of AI and the troubling outcomes it can generate. Professor Narayanan provided feedback on Part IV of these comments, in which we explain (based largely on the professor's own research) why AI deserves special scrutiny from the Commission.

[2] These comments do not represent the views of the Yale Law School or Yale University.

Table of Contents

Public Knowledge is writing this comment to urge the Federal Trade Commission to implement new trade regulations concerning the collecting, aggregation, use, sharing, and retention of consumer data. We look forward to working with the Commission to create a rule that addresses the harmful practices of commercial surveillance while protecting civil rights, reducing harms to consumers, and promoting a vibrant digital ecosystem.

As recounted in the announcement for the proposed rulemaking, for the past two decades the Commission has done privacy work through either bringing litigation against companies whose practices violate Section 5 or through general policy statements. While this approach is generally useful for stopping bad actors and giving general guidance to industry, it is not able to address prevalent practices which harm consumers. Furthermore, the Commission's privacy cases have tended to focus on whether a company gave proper "notice" to a user about what would happen to their data, so therefore they couldn't make an informed "choice." Commissioners Khan[3] and Slaughter[4] in statements issued after the announcement of this proceeding all remarked that the notice and choice model is a consumer protection failure. This rulemaking is an opportunity for the Commission to change conduct, rather than continue to rely solely on transparency measures. Creating rules also allows the Commission to seek civil penalties against first-time violators. Under the current model in order to get civil penalties, the Commission must first get a written decision showing that a specific practice was unfair or deceptive; then anyone who engages in that practice after the decision is published is on notice that engaging in that kind of conduct is not allowed and will therefore carry civil penalties.[5] This

[3] Statement of Commissioner Lina M. Khan Regarding the Commercial Surveillance and Data Security Advance Notice of Proposed Rulemaking (Aug. 11, 2022), at 3-4.

[4] Statement of Commissioner Rebecca Kelly Slaughter Regarding the Commercial Surveillance and Data Security Advance Notice of Proposed Rulemaking (Aug. 11, 2022) at 1-2.

[5] 15 U.S.C. §45(m)(1)(B

means the deterrent effect against companies engaging in novel data extractive practices have little to no incentive to think about the harm it may do to their users or customers. Finally, the Commission is limited in how many cases it can bring a year. Having rules that are clear and apply equally to existing and emergent data practices gives companies a way to stay compliant, while also deterring harmful behavior.

To assist in this proceeding, we are submitting comments that address six different areas of interest to the Federal Trade Commission. First, we assert that the Commission has the necessary authority and expertise to promulgate privacy and data security rules. Relying on industry to police itself through self-regulation will not create meaningful change in the digital economy. Second, we catalog not just the different ways companies collect, share, and retain data, but also how those practices harm consumers. Further, we show that even when the harms come to light, companies have little incentive to change their behavior. Third, we caution the Commission from over relying on either consent or anonymization techniques to protect individual privacy. It is easy to manipulate consumers into providing consent and while data holders may claim that the information is anonymized, oftentimes that isn't the case. If the Commission chooses to allow consent or anonymization to be used as a legitimate basis for collection or sharing, strict rules must be in place to ensure that consent is meaningful, and that data is truly anonymized. In our fourth section, which is meant as a response to the Commission's questions on Automated Decision-Making Systems, we propose a framework that requires companies to only sell artificial intelligence (AI) systems that accurately and effectively perform the job that is advertised. This framework draws heavily on the Commission's work on truth in advertising and provides a useful starting point for any discussion of AI regulation. Fifth, we enumerate the benefits of implementing specific remedies in the new trade regulations, which should include data deletion and algorithmic disgorgement. Finally, we provide recommendations on fruitful avenues for substantive rulemaking.

I. COMMISSION'S AUTHORITY TO COMMENCE A RULEMAKING

The Commission has clear authority under Section 18 to commence a rulemaking on commercial surveillance and data security. Section 45(a)(1) declares "[u]nfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce" unlawful and empowers the Commission with rulemaking authority to effectuate this mission. This proposed rulemaking is well in line with both the statutory text and the FTC's overall mission of "protecting the public from deceptive or unfair business practices and from unfair methods of competition."[6]

Commercial surveillance is very well captured by the term "unfair or deceptive acts or practices in or affecting commerce." It is deceptive in that most consumers have no idea the breadth and depth of their intimate data collected through seemingly innocuous activities such as web browsing. The FTC need not show that deceptive claims are expressed explicitly to exercise its consumer protection authority; rather, a showing of "deceptive effect" is sufficient.[7] Furthermore, these practices are unfair to the average consumer, because it is nearly impossible for consumers not to be surveilled while they surf the internet. The billions spent on these practices puts this well in line with affecting commerce. Some of the most profitable companies in the world, such as Facebook and Google, have turned targeted advertising into an incredibly lucrative market. The fuel for those targeted ads? User data from commercial surveillance.

[6] *Mission*, Federal Trade Commission, (last visited Oct. 4, 2022), https://www.ftc.gov/about-ftc/mission.
[7] Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 667-68 (2014).

The Commission's clear authority was further elucidated in the Magnuson Moss Warranty-Federal Trade Commission Improvements Act in 1975.[8] The Commission is well in line with its statutory responsibilities under both the FTC Act and Magnuson-Moss at this preliminary stage. The FTC published an advanced notice of proposed rulemaking (ANPR) on August 22, 2022, detailing the Commission's proposed area of focus and objectives. Public comment was solicited both through this docket and a virtual public forum on September 8, 2022.[9]

Privacy Harms Are a Prevalent, "Widespread Pattern" Throughout Economy

The Commission may only issue rules for practices for which it has reason to believe are "prevalent." The statute further defines prevalence as practices where the Commission "has issued cease and desist orders regarding such acts or practices" or if there is "a widespread pattern" of the practices.[10] The Commission's long history of privacy enforcement orders clearly indicates that both prongs of this prevalence standard are easily met.

Recent history showcases digital companies violating user privacy and various FTC attempts to rein them in. In 2019, the Commission leveled a then-record $5 billion fine on Facebook for violating a previous privacy order.[11] The FTC found Facebook had deceived its users on its data-sharing practices in clear violation of a 2012 FTC consent order.[12] This episode

[8] 15 U.S.C. §§ 2301-2312.

[9] *Commercial Surveillance and Data Security Public Forum*, Federal Trade Commission, (last visited Oct. 4, 2022), https://www.ftc.gov/news-events/events/2022/09/commercial-surveillance-data-security-anpr-public-forum.

[10] 15 U.S.C. § 57a(b)(3).

[11] *FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook*, Federal Trade Commission, (Jul. 24, 2019) https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook.

[12] Id.

also underlines the ineffectiveness of case-by-case enforcement and the need for a strong rule applying to the entire industry.[13]

The FTC recently sued data broker Kochava for selling user geolocation data used to track users' visits to sensitive locations—from healthcare providers to addiction treatment centers.[14] While the case is laudable, Kochava is just one broker in a massive user tracking economy. Even if the case is successful, new Kochavas will spring up Hydra-like in its place. The answer must be an industry-wide rulemaking.

Self-Regulation Has Not Worked, And Will Not Work Here

Self-regulation simply does not work in most contexts, and that is certainly true for this one. Companies are profit-maximizing entities and will act in the best interests of their shareholders, not the public's interest. Commercial surveillance and data exploitation, despite its potential societal pitfalls, is immensely profitable. Data, especially given its increasing returns to scope and scale, is the lifeblood and currency of digital platforms. Left unchecked, we can expect companies to hoover up all the user data they can get their hands on.

Self-regulation is anathematic to the entire American governmental system. As Federalist 51 so eloquently lays out, "ambition must be made to counteract ambition."[15] As the Founders recognized, in order to balance power, you need to ensure that institutions act as a counterbalance to one another. In doing so, you temper the ability of any one faction to exert an outsized influence. Companies, as can be expected, act in their own self-interest. To ensure

[13] In the Matter of Facebook, Inc., C-4365, 2012 FTC LEXIS 135 (F.T.C. April 27, 2020); In the Matter of Twitter, Inc., C-4316, 202 FTC LEXIS 162 (F.T.C. May 26, 2022); In the Matter of Uber Technologies, Inc., C-4662, 2018 LEXIS 152 (F.T.C. October 25, 2018) (revising the consent agreement and expanding the settlement with Uber).

[14] *FTC Sues Kochava for Selling Data that Tracks People at Reproductive Health Clinics, Places of Worship, and Other Sensitive Locations*, Federal Trade Commission, (Aug. 29, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-sues-kochava-selling-data-tracks-people-reproductive-health-clinics-places-worship-other.

[15] James Madison, *The Federalist No. 51*, (February 8, 1788), *available at* https://avalon.law.yale.edu/18th_century/fed51.asp.

consumers are protected from their unbridled pursuit of self-interest, regulators must be powerful and independent from the companies they seek to regulate. That counterbalance produces the protections consumers need. Self-regulation cedes too much under the guise of balancing interests.

History is rife with examples of failed self-regulation. The Safe Harbor programs in COPPA have drawn scrutiny from Congress for their ineffectiveness.[16] For a company to be part of the Safe Harbor program, one of the six approved Safe Harbor organizations must review that company's privacy policies and practices, and ensure they comply with COPPA. However, the organizations certifying compliance with COPPA don't just make money from the process of certification, but some also have COPPA consulting services as well.[17] This creates a clear conflict of interest. The Commission has already removed a company from the self-regulatory list after finding clear conflict-of-interest issues given the company was funded by those it sought to regulate.[18] Effective regulation needs to be run by entities with solely the public weal as their goal, such as the Commission.

COPPA is not alone in privacy self-regulatory failures. The "Student Privacy Pledge" allowed education technology companies to voluntarily pledge to uphold data protection rules.[19] Unsurprisingly, relying on the remote possibility that an FTC enforcement action could arise for

[16] *See Reps. Castor, Schakowsky Request Answers from COPPA's Safe Harbor Programs*, U.S. Representative Kathy Castor, (Jan. 10, 2022), https://castor.house.gov/news/documentsingle.aspx?DocumentID=403770.

[17] Rohit Chopra, Statement of commissioner Rohit Chopra - Regarding Miniclip and the COPPA Safe Harbors Commission File No. 1923129 (2020), https://www.ftc.gov/system/files/documents/public_statements/1575579/192_3129_miniclip_-_statement_of_cmr_chopra.pdf.

[18] *Aristotle Removed from List of FTC-Approved Children's Privacy Self-Regulatory Programs*, Federal Trade Commission, (Aug. 4, 2021), https://www.ftc.gov/news-events/news/press-releases/2021/08/aristotle-removed-list-ftc-approved-childrens-privacy-self-regulatory-programs.

[19] Kristal Kuykendall, *Illuminate Education Booted from Student Privacy Pledge*, The Journal, (Aug. 9, 2022), https://thejournal.com/articles/2022/08/09/illuminate-education-booted-from-student-privacy-pledge-referred-for-potential-ftc-state-ag-action.aspx.

a company if it was found to not be conforming to its public statements, meant the pledge was relatively toothless and at least one company was removed after a massive data breach endangering a large volume of student data.[20] The Commission must take notice of the fact that there is far too much at stake for it to continue to leave companies to their own devices when it comes to consumer privacy. It's time for a commercial surveillance rulemaking.

II. DATA COLLECTION AND USE IS PERVASIVE, NON-TRANSPARENT, AND HARMFUL TO CONSUMERS

Since the inception of the internet more than two decades ago, the exponential increase in the volume of consumer data collected through the internet and the subsequent exploitation of that consumer data to target advertisements, aid in financial and insurance discrimination, and put consumers at greater risk is astonishing. Companies have internalized the exploitative value of consumer data and many internet-based companies as well as data brokers rely on that unbridled exploitation to return shareholder value. Businesses, therefore, will pursue an ever-increasing trove of consumer data through active and passive collection techniques to make money for their shareholders and that incentive leads businesses to obfuscate their data collection practices, misleading consumers or being opaque in their public wording about what exactly use of their service means for the consumer's data.

As then FTC Commissioner Chopra testified at a hearing in 2019, data has unique features that are unlike any other asset in the economic marketplace.[21] Those features are:

First, it is not a finite resource, like precious metals or minerals.

Second, data is not "consumed" in the traditional sense… it can be copied and shared.

[20] Id.

[21] Online Platforms and Market Power, Part 3: The Role of Data and Privacy in Competition: Hearing Before the Subcomm. on Antitrust, Commercial, and Administrative Law, 116 Cong. 2 (2019) (Statement of Rohit Chopra). https://www.ftc.gov/system/files/documents/public_statements/1549812/chopra_-_testimony_at_hearing_on_online_platforms_and_market_power_part_3_10-18-19.pdf

Third, data gets more valuable as you collect more of it.[22]

These three features mean companies have little incentive to limit their data collection and consumers cannot be expected to effectively fight back against the push to collect more data.

How Companies Collect Data from Users

Data collection mechanisms can be put into two buckets: active or passive. Active data collection happens when the user of the application or website provides the information directly. This can include filling out a registration form, purchasing an item, or giving a voice command to a smart speaker. When a company collects information actively from a user, that person then generally understands the company is collecting data they provide. This gives the user some modicum of notice on what is being collected from them, because they are asked to provide the information directly. With that knowledge, users can then judge whether they want to share that information with the website or application or take other countermeasures like providing false information to obscure their identity. While it may be slightly easier for users to understand what data is being collected from and about them when companies are actively collecting their data, it does not provide any insight into how the data may be used later. And once that data is provided, there is no way for the user to stop it from being used further.

Passive data collection is extremely difficult for an individual user to manage because they don't know exactly what is being collected or with whom it is being shared. While the application may ask for permission to access certain broad types of information, like location, the user doesn't know precisely what is being shared. At best, the user could read a dense, complex privacy policy that obscures, obfuscates, and downplays any potential data collection.[23] Or, even

[22] Id.

[23] E.g., Kevin Litman-Navarro, *We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.*, New York Times (Jun. 12, 2019), https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html?mtrref=www.google.com&gwh=C394A6BFF9F7044C1C548527E40481FC&gwt=regi&assetType=R

worse, be told they can change their privacy settings, but can only access those settings by navigating through a series of confusing menus and submenus, only to be confronted with options that are unclear about what exactly they do.[24] Below are case studies showing how active and passive data collection harms users and their communities.

Data Collection and Use Case Studies

Geo-location Information

Last year, a Catholic Substack, The Pillar, was able to out a gay priest using location data collected by the dating app Grindr.[25] While the location data was not identified as being the outed priest, using publicly available information and other data purchased from data brokers, The Pillar was able to identify him with a high degree of confidence. This public outing forced the priest to resign from his role in the Catholic Church. While this could be considered a tragic one-off story, Grindr had been fined for selling access to user location data and told of this vulnerability for years but chose to do nothing.[26] Grindr is not the only dating app to share

EGIWALL; Marcus Moretti & Michael Naughton, *Why Privacy Policies Are So Inscrutable*, The Atlantic (Sept. 5, 2014), https://www.theatlantic.com/technology/archive/2014/09/why-privacy-policies-are-so-inscrutable/379615/.

[24] See, e.g., Ben Stegner, *How Dark Patterns Mislead You Into Making Bad Privacy Choices*, Make Use Of (Nov. 4, 2021), https://www.makeuseof.com/tag/dark-patterns-bad-privacy-choices/; Hana Habib et al., *An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites*, USENIX Symposium on Usable Privacy and Security (Aug. 11-13, 2019), https://www.usenix.org/system/files/soups2019-habib.pdf.

[25] Joseph Cox, *The Inevitable Weaponization of App Data Is Here*, Vice (Jul. 21, 2021), https://www.vice.com/en/article/pkbxp8/grindr-location-data-priest-weaponization-app.

[26] Alex Lomas, *Dating apps that track users from home to work and everywhere in-between*, Pen Test Partners (Aug. 11, 2019), https://www.pentestpartners.com/security-blog/dating-apps-that-track-users-from-home-to-work-and-everywhere-in-between/ (Finding Grindr's previous statement that their location information was not being stored "precisely" and being "more akin to a 'square on an atlas'" was wrong and that "Grindr location data was able to pinpoint our test accounts down to a house or building, i.e. exactly where we were at that time" and attempts to contact Grindr were met with no response); @Seppevdpll, *It is Still Possible to Obtain the Exact Location of Millions of Men on Grindr, Queer Europe* (Sept. 13, 2018), https://www.queereurope.com/it-is-still-possible-to-obtain-the-exact-location-of-cruising-men-on-grindr/ (Recounting Grindr's acknowledgement of flawed features dating back to 2014 but today "anyone can still use Grindr's servers to collection the location, sexual position and HIV status of cruising men."); Andy Greenberg, *Gay Dating Apps Promise Privacy, But Leak Your Exact Location*, WIRED (May 20, 2016), https://www.wired.com/2016/05/grindr-promises-privacy-still-leaks-exact-location/ (Quoting a Grindr spokesperson responding to WIRED's research saying only that "Grindr takes our users safety extremely seriously, as well as their privacy…we are working to develop increased security features for the app); *Out of Control*, Forbruker Rådet (Jan. 14, 2020), https://www.forbrukerradet.no/out-of-control/#.

sensitive user data widely. OKCupid shared users' personal information regarding sexuality, drug use, and political views with up to 300 different third-party advertising and analytics firms.[27]

Health Information

Since the Dobbs decision, there has been heightened scrutiny of how period tracking apps collect, use, and retain data.[28] And for good reason: just last year, the FTC settled with Flo, a popular period tracking app, for lying about its privacy practices to users.[29] Flo promised that data would only be shared to "provide services in connection with the app,"[30] but that was not the case. Flo was sharing health related information, like a user's pregnancy status, with a variety of third-party marketing and analytics firms. To add insult to injury, the health information being shared with third parties was not being securely transferred. An investigation by the Wall Street Journal made it clear that outside attackers could see the unencrypted health information being transferred from Flo to its third-party marketers, like Facebook.[31]

Since the settlement with the FTC, Flo has improved its privacy practices. It has since launched an "anonymous mode," which removes all personal and technical identifiers from a profile.[32] Once a user selects "anonymous mode" Flo is not able to reassociate a profile with a specific user, even if an outside entity, like law enforcement, seeks that information. While it is

[27] Natasha Singer & Aaron Krolik, *Grindr and OkCupid Spread Personal Details*, New York Times (Jan. 13, 2020), https://www.nytimes.com/2020/01/13/technology/grindr-apps-dating-data-tracking.html.

[28] Rina Torchinsky, *How period tracking apps and data privacy fit into a post-Roe v. Wade climate*, NPR (Jun. 24, 2022), https://www.npr.org/2022/05/10/1097482967/roe-v-wade-supreme-court-abortion-period-apps).

[29] Id.

[30] In the Matter of Flo Health, Inc., Docket C-4747, Complaint (Jan. 13, 2021), https://www.ftc.gov/system/files/documents/cases/flo_health_complaint.pdf.

[31] Dam Schechner, *You Give Apps Sensitive Personal Information. They Tell Facebook.*, Wall Street Journal (Feb. 22, 2019), https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636.

[32] Amina Kilpatrick, *Period tracker app Flo developing 'anonymous mode' to quell post-Roe privacy concerns*, NPR (Jun. 30, 2022), https://www.npr.org/2022/06/30/1108814577/period-tracker-app-flo-privacy-roe-v-wade.

commendable that the FTC was able to use its enforcement authority to get Flo to change its practices, it is not the only period tracker on the market. The Norwegian Consumer Council found that MyDays was sharing sensitive user data (in this case location information) with a host of data brokers.[33] MyDays, now MyDaysX, has not stopped this practice.[34]

The underlying reason for all this opaque data sharing is targeted advertising. Targeted advertising is fueled by personal data.[35] The theory of targeted advertising is that by having as much information as possible about your potential customers, you can more effectively spend your advertising dollars. This is why we have seen privacy scandals in almost every sector — from makeup and beauty to weather forecasting, to radio apps.[36] While targeted advertising, in the right circumstances,[37] does increase sales that doesn't mean the harms to consumers,[38] like an encroachment on autonomy or discrimination, are worth the unfettered data collection and use.

Data Brokers

Amongst its many services, Oracle is a data broker, meaning it acts as a data aggregator and seller of consumer information. It collects data from across the internet and beyond—using a combination of its own proprietary cookies, JavaScript code, and tracking pixels, as well as buying other publicly available information—to create sophisticated profiles that marketers can use for their own targeted advertising campaigns.[39] Oracle goes one step further though: it sells

[33] *Out of Control*, *supra* note 26.

[34] *Privacy Policy*, MyDaysX (last updated Aug. 12, 2022), https://blog.mydaysx.club/privacy-policy/.

[35] Anna Dorothea Ker, *The Big Business of Ad Tech*, The Privacy Issue (Jan. 22, 2020) https://theprivacyissue.com/data-tracking/big-business-ad-tech.

[36] Thomas Brewster, *A Load of Apple iPhone Apps Are 'Covertly' Selling Your Location*, Forbes (Sept. 7, 2018), https://www.forbes.com/sites/thomasbrewster/2018/09/07/a-load-of-apple-iphone-apps-are-covertly-selling-your-location/?sh=5a7e7cc85832.

[37] Louise Matsakis, *Online Ad Targeting Does Work–As Long As It's Not Creepy*, WIRED (May 11, 2018), https://www.wired.com/story/online-ad-targeting-does-work-as-long-as-its-not-creepy/?redirectURL=/story/online-ad-targeting-does-work-as-long-as-its-not-creepy/.

[38] Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 Boston Univ. L. Rev. 793 (2022)

[39] Oracle Data Cloud, *The power of the Oracle ID Graph*, YouTube (2017), https://www.youtube.com/watch?v=63ZsFLIUcNw.

proxy data for sensitive data like race, political leanings, location, and medical information to get around privacy controls.[40] Thus, when Facebook tries to ensure that race based ads for credit or housing[41] are not present on its platform, Oracle can provide the data and technology to circumvent those consumer protections.[42] These data points aren't only used for advertising, but for assessing "risk" as well. LexisNexis Risk Solutions offers a scoring product to health insurance providers that claims to predict a patient's medical costs.[43] As described, the product uses "442 non-medical personal attributes" to predict a patient's overall health risk, how likely they are to visit an emergency room, need prescriptions, and even their "motivation to stay healthy."[44] While insurers say that all of this information is used to help their patients receive care and lower costs, this kind of scoring makes it much easier to push certain patients into high cost plans or find patients who are unlikely to need much care but will certainly pay their premiums on time. These third parties, who almost never interact with consumers directly, can leverage vast amounts of data against consumers because there is very little stopping first party data collectors, like apps and websites, from collecting as much information about their users as possible and then selling and sharing it freely.

Device Data Collection

Information from apps and websites are not the only ways that companies collect data about their customers. Devices like phones, set-top boxes, routers, and Information of Things

[40] Complaint, *Katz-Lacabe v. Oracle Am., Inc.*, Case No. 3:22-cv-04792.

[41] *Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms*, U.S. Department of Jusice (Jun. 21, 2022), https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known.

[42] *Katz-Lacabe*, *supra* note 40.

[43] Marshall Allen, *Health Insurers Are Vacuuming Up Details About You–And It Could Raise Your Rates*, ProPublica (Jul. 17, 2018), https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

[44] Id.

(IOT) devices generally all provide ways to surreptitiously collect data without users' being aware of it. Cable providers use set-top boxes to gather their customers' personal information, sell or share it with third parties, so the information can be used for targeted advertising.[45] This information is often combined with loyalty program data as well, all without the customer being the wiser.[46]

Broadband Network Operators

Today, many of the largest cable providers also are prominent internet service providers. The FTC has noted in its own report of the industry that:

- *Many of the ISPs… amass large pools of sensitive consumer data.*
- *Many of the ISPs… gather and use data in ways consumers do not expect and could cause them harm.*
- *Although many of the ISPs… purport to offer consumers choices, these choices are often illusory.*
- *Many ISPs can be at least as privacy-intrusive as large advertising platforms.*[47]

The lack of privacy oversight in this sector of the economy has led ISPs on a privacy "race to the bottom"[48] so that they too could participate in the lucrative AdTech market. This voracious data appetite is not relegated to just companies offering cable or internet; wireless carriers (some of whom also participate in the cable and internet market) also have been found selling location

[45] *Big Data is Watching*, Center for Digital Democracy (Aug. 5, 2016), https://www.democraticmedia.org/content/big-data-watching-growing-digital-data-surveillance-consumers-isps-and-other-leading-video.

[46] Steven Perlberg, *Targeted Ads? TV Can Do That Now Too*, Wall Street Journal (Nov. 20, 2014), http://www.wsj.com/articles/targeted-ads-tv-can-do-that-now-too-1416506504.

[47] Federal Trade Commission, *A Look At What ISPs Know About You* (Oct. 21, 2021).

[48] Rebecca Kelly Slaughter, *Remarks Regarding the FTC ISP Report*, Federal Trade Commission (Oct. 21, 2021), https://www.ftc.gov/system/files/documents/public_statements/1597814/slaughter_isp_report_statement_10-20-21.pdf.

data as well.[49] A Motherboard investigation[50] showed that anyone from bail bondsmen[51] to bounty hunters[52] could purchase location data history if they had a person's cell number and a couple hundred dollars. This has spurred the Federal Communications Commission to conduct its own investigation into the wireless industry's general data privacy practices.[53] However, like we've seen raised in the Kochava litigation,[54] industry is quick to say that they have consent from their customers to use this data.[55]

[49] Press Release, FCC Proposes Over $200M in Fines for Wireless Location Data Violations, Fed. Commc'n Comm'n (Feb. 28, 2020), https://www.fcc.gov/document/fcc-proposes-over-200m-fines-wireless-location-data-violations (choose "pdf" next to "News release")

[50] Joseph Cox, *I Gave a Bounty Hunter $300. Then He Located Our Phone.*, Vice (Jan. 8, 2019), https://www.vice.com/en/article/nepxbz/i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tmobile.

[51] Jennifer Valentino-DeVries, *Service Meant to Monitor Inmates' Calls Could Track You, Too*, New York Times (May 10, 2018) https://www.nytimes.com/2018/05/10/technology/cellphone-tracking-law-enforcement.html.

[52] Joseph Cox, *Hundreds of Bounty Hunters Had Access to AT&T, T-Mobile, and Sprint Customer Location Data for Years*, Vice (Feb. 6, 2019), https://www.vice.com/en/article/43z3dn/hundreds-bounty-hunters-att-tmobile-sprint-customer-location-data-yearsJ.

[53] Press Release, Rosenworcel Probes Mobile Carriers on Data Privacy Practices, Fed. Commc'n Comm'n (Jul. 19, 2022), https://www.fcc.gov/document/rosenworcel-probes-mobile-carriers-data-privacy-practices (choose "pdf" next to "News release")

[54] John D. McKinnon, *Idaho Company Sues FTC, Claiming Agency Threatened Suit Over Its Tracking Data*, Wall Street Journal (Aug. 15, 2022), https://www.wsj.com/articles/idaho-company-sues-ftc-claiming-agency-threatened-suit-over-its-tracking-data-11660608782 (Reporting Kochava's claims that it's data is anonymized and they have consent).

[55] *Out of the fifteen carriers who submitted letters to the FCC, five of the carriers claimed to have user consent when sharing customer geolocation information with third parties, including all of the major cell phone carriers.* See Letter from Joan Marsh, AT&T Exec. Vice President, to Jessica Rosenworcell, Fed. Commc'n Comm'n Chairwoman (Aug. 3, 2022) (on file with the Fed. Commc'n Comm'n); Letter from Darah Franklin, Counsel to Google N. Am. Inc., to Jessica Rosenworcell, Fed. Commc'n Comm'n Chairwoman (Aug. 3, 2022) (on file with the Fed. Commc'n Comm'n); Letter from Roberta Kraus, Gen. Counsel to Lycamobile, to Jessica Rosenworcell, Fed. Commc'n Comm'n Chairwoman (Aug. 16, 2022) (on file with the Fed. Commc'n Comm'n); Letter from Kathleen Ham, Senior Vice President, Gov't Aff., T-Mobile USA, Inc., to Jessica Rosenworcell, Fed. Commc'n Comm'n Chairwoman (Aug. 3, 2022) (on file with the Fed. Commc'n Comm'n); Letter from William Johnson, Senior Vice President, Fed. Regul. & Legal Aff. at Verizon, to Jessica Rosenworcell, Fed. Commc'n Comm'n Chairwoman (Aug. 3, 2022) (on file with the Fed. Commc'n Comm'n).
None of the carriers who reported sharing information with third parties cited any kind of anonymization practices as a method of safeguarding current and former subscriber geolocation data. However, most of the answers to that question were very vague citing methods like "technical security measures" that could include anonymization practices but didn't explicitly say so.

III. RELYING ONLY ON NOTICE AND CONSENT OR ANONYMIZATION DOES NOT PROTECT CONSUMERS

In developing our Nation's approach to privacy, we have left it largely up to an ad hoc, ex post structure that relied on public representations by companies and then pursuit of some violations of those representations by the FTC under its Section 5 enforcement authority. Such reliance at this juncture in the maturation of the digital economy is not tenable. There must be baseline rules that guide companies' collection and use of consumer data, even when that data is anonymized. We know, as referenced earlier, that the economic incentives for collection and use are not aligned with respecting privacy and cause harm. As set out below, we will demonstrate that the reliance on "notice and choice" regimes and data anonymization are not tools that can cure the problem faced by excessive commercial surveillance. Instead, we need the FTC to undertake a rulemaking to develop specific rules to control for the abuses that we have seen in our current "system" of protecting consumer data.

Notice and Consent is Not a Cure

As an initial matter, companies generally do not need to get consent in order to collect or process personally identifiable data. This is because the FTC has fostered a privacy enforcement regime based on traditionally relied upon violations of "notice and consent" frameworks based on its Section 5 deception authority as the primary privacy enforcement mechanism.[56] Essentially, the logic of this practice is that consumers must have "notice" of a company's privacy practices, usually through a privacy policy. Then, the theory goes, the consumer can make an "informed choice" about whether to use a service. If the company is not following their

[56] Federal Trade Commission, Comment Letter In the Matter of Developing the Administration's Approach to Consumer Privacy (Nov. 9, 2018), https://www.ftc.gov/system/files/documents/advocacy_documents/ftc-staff-comment-ntia-developing-administrations-approach-consumer-privacy/p195400_ftc_comment_to_ntia_112018.pdf

own privacy policy, then sufficient notice has not been given to the consumer and the notice is deceptive.

Notice and Consent Is a Fallacy

In theory, the notice and consent approach outlined above should give consumers greater control over their data and online lives. In practice, however, notice and choice does not and cannot work. This is because the framework assumes a few things that we know are not true: 1) that users read privacy policies;[57] 2) that they understand what the policies say;[58] and 3) that they have a practical choice about whether or not to use a website or application under those conditions.[59] These assumptions are wrong, and, therefore, the "notice and choice" framework simply can't protect internet users.

Instead, what happens is that users drown in information that they can't be expected to read and understand because, not only is it often presented in an opaque manner meant to confuse, but it varies on each website and app. This allows companies to siphon data from those very same users that they then exploit. Such a framework is perfect if the policy objective is to allow for the unbridled collection and use of consumer data. It is not optimal if the goal of policy is to provide consumers protection.

Additionally, seeking consent from a consumer when there are virtually no limits on what data can be collected, and little to no restrictions on its use, leaves consumers facing a Hobbesian choice because not using the internet's most dominant websites or many of the most popular

[57] Christopher Koopman, *No one reads online privacy policies*, The Benchmark (Feb. 15, 2019), https://medium.com/cgo-benchmark/no-one-reads-online-privacy-policies-42b067701efc.

[58] Litman-Navarro, *supra* note 23.

[59] Katharine Kemp, *Concealed data practices and competition law: why privacy matters*, European Competition Journal (Oct. 15, 2020), https://www.tandfonline.com/doi/pdf/10.1080/17441056.2020.1839228?needAccess=true.

apps is not a meaningful choice; It is not 1998 anymore and the use of the internet is critical for participation in modern commerce, culture, and living.

When companies claim to have received consent from users, oftentimes they are citing terms in their privacy policy like "by using this site you consent to the collection, use, and disclosure of your information in accordance with the Privacy Policy." By any metric, this is not meaningful consent, and this practice cannot be the foundation of how data is collected and used.

Informed, Affirmative and Express Consent

Consent, if it is going to be used by companies to legitimize data collection practices, at the very least must be informed, affirmative, and express. Under the European General Data Protection Regulation (GDPR) this means that consent is "freely given, specific, informed and [an] unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."[60] Companies are also not allowed to require a person to consent to data processing in order to use an application or service.[61] These requirements must be the floor, not the ceiling, when designing a consumer friendly consent regime.

The cases brought under the GDPR have expanded the understanding of what is and is not informed, affirmative, express consent. Just this past year France's data protection authority, CNIL, fined Google and Facebook for not making it as easy to refuse all cookies as it was to accept all cookies.[62] This case built on a 2019 case against Google, which said the consent for

[60] Regulation 2016/679, of the European Parliament and Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 4, 5. Article 4(11)

[61] Id. at Article 7.

[62] Kelvin Chan, *France fines Google, Facebook millions over tracking consent*, ABC News (Jan. 6, 2022), https://abcnews.go.com/International/wireStory/france-fines-google-facebook-millions-tracking-consent-82108667; *Cookies: Google fined 150 million euros*, CNIL (Jan. 6, 2022), https://www.cnil.fr/en/cookies-google-fined-150-million-euros.

data processing was not valid because "essential information was disseminated across several documents" and required five to six steps to view that information.[63] Without that information being easily accessible, the consent could not be considered informed. These cases taken together describe a possible path forward for the FTC when creating rules governing consent. Users must be able to make an informed choice the moment when the data is going to be collected, and that choice must not be hindered by a company's design decisions.

There are some data collection activities where it would be very difficult to obtain consent from a user. For example, connected devices like smart speakers, lights, or appliances often do not have interfaces that are designed for reading and understanding lengthy privacy policies or consent notices. And even if they do ask for the consent of users through an associated app or website, other users – like guests in a home – would never have the opportunity to opt in or opt out of the associated data collection. That makes it difficult for connected devices to truly "inform" their users of what they are doing with their data. This problem is compounded for devices that collect data in public spaces or in highly trafficked areas of private spaces (like homes or businesses). Individuals in these spaces might not even know these devices are in use, and even if they do, these devices would not be able to get informed, affirmative, and express consent from every data subject. These kinds of devices and use cases should never be allowed to rely on consent, because their design does not allow for informed, affirmative, express consent.

As a general matter, designing a consent-based system that also advances consumer protection goals is difficult. The relationship between consumers and the businesses that collect their data is asymmetric. Consumers rely on companies to inform them of when their data is

[63] Chris Fox, *Google hit with £44m GDPR fine over ads*, BBC News (21 Jan. 2019), https://www.bbc.com/news/technology-46944696.

collected, for what purpose, and how that may affect them in the future.[64] This imbalance of power is further exacerbated because businesses control how and when all of that information is relayed to the consumer.[65] If the FTC decides to allow companies to collect personal data based on the consent of the user, there must be strict controls on how consent is given. Companies cannot be allowed to turn consent into a "useless exercise,"[66] but rather the Commission must create rules that limit consent to when a user is informed of what data is collected and how it can be used and is free to say "no" without interference.

Data Anonymization, as a Privacy Technique, Must Be Viewed Skeptically

Data anonymization is the process of hiding information in a dataset that would uniquely identify the individuals that data is collected from. The process is meant to ensure that data collected from individuals cannot be used to identify them personally or be linked to one of their individual identifiers. Data protection laws like the GDPR and the Health Insurance Portability and Accountability Act (HIPAA) restrict the sharing of personal data to protect an individual's identity,[67] thereby encouraging data anonymization practices as a means to comply with these laws so companies can share and sell data for statistical purposes, research, or advertising.[68] While data anonymization may have its uses within a privacy framework, it is not an overarching solution. Oftentimes, when a company says data has been "anonymized" it takes a trivial amount of effort to reidentify the data set. Therefore, the Commission must create rules that identify

[64] Alessandro Acquisti et al., *The Economics of Privacy*, 54 J. of Economic Literature 442, 442-92 (2016).

[65] Woodrow Hartzog, Privacy's Blueprint (Harvard University Press, illustrated ed. 2018).

[66] Joe Nocera, *How Cookie Banners Backfired*, New York Times (Jan. 29, 2022), https://www.nytimes.com/2022/01/29/business/dealbook/how-cookie-banners-backfired.html.

[67] Chris Foot, *Data anonymization best practices protect sensitive data*, TechTarget (Dec. 16, 2020), https://www.techtarget.com/searchdatamanagement/feature/Data-anonymization-best-practices-protect-sensitive-data.

[68] J.M. Hendrickx et al., *Estimating the success of re-identifications in incomplete datasets using generative models*, Nature Communications (Jul. 23, 2019), https://www.nature.com/articles/s41467-019-10933-3#citeas.

which anonymization techniques are permissible, if there are data types that are not suited for anonymization, and how companies must guard against reidentification.

Data Suppression Techniques

Data anonymization techniques obscure identifiers within a dataset such as names, social security numbers, birthdates, and addresses.[69] Companies employ various methods to obscure identifying data points. One type of anonymization is attribute suppression. Attribute suppression removes a category of information, such as social security numbers, from the dataset entirely, making it a stronger type of anonymization as that information is no longer accessible.[70] Record suppression works very similarly: it deletes an entire record from a dataset, for example information collected on a particular individual, so it is mostly used to remove outliers in a dataset.[71] If dataset owners want to remove attributes or records but still keep a placeholder in that space, they can use character masking which replaces some characters with an "*" or "x."[72] Datasets can also be anonymized through pseudonymization, or coding, which swaps identifiers with either random or deterministically generated fictional values.[73] When pseudonymizing datasets, the original values can either be removed permanently or stored (and sometimes encrypted) so they can be accessed later, but the latter approach is less secure than removal because it is more susceptible to a data breach.[74] Generalization is the process of anonymizing datasets by making certain values less precise by converting exact values to a range, for example

[69] Foot, *supra* note 67.

[70] Personal Data Protection Commission Singapore, *Guide to Basic Data Anonymisation Techniques*, International Association of Privacy Professionals 1, 12-26 (Jan. 25, 2018), https://iapp.org/media/pdf/resource_center/Guide_to_Anonymisation.pdf.

[71] Id.

[72] Id.

[73] Id.

[74] Id.

changing someone's age to an age range.[75] If the range is too small, re-identification of generalized datasets is fairly easy.[76] Summarizing the values via data aggregation is another method of anonymization that can be used to decrease preciseness of data.[77]

Data Swapping Techniques

Datasets can also be anonymized using swapping (a.k.a. shuffling or permutation): rearranging the data so that individual values do not correspond to the original records; or data perturbation: slightly modifying quasi-identifiers in the original dataset, using rounding or adding random noise.[78] Injecting a predetermined amount of noise in a dataset is called differential privacy, which is a more robust method of anonymization.[79] Data scientists can still work backwards to obtain original records since they know how much noise was injected in the dataset, but it takes multiple queries to get enough variation in the records to work out individually identifiable information, which reduces the risk of a privacy breach for each individual in the dataset.[80] Generating synthetic datasets separately from the original data is another method of anonymization, but like pseudonymization, makes the original values more susceptible to a data breach.[81] Some companies use software products developed by companies like Salesforce or Aircloak to anonymize their datasets for them.[82] Many of these anonymization

[75] Id.

[76] Id.

[77] Id.

[78] Id.

[79] *What is Differential Privacy?*, MIT Ethical Technology Initiative, http://eti.mit.edu/what-is-differential-privacy/ (last visited Sept. 30, 2022).

[80] Id.

[81] Id.

[82] Foot, *supra* note 67; Nicolas Sartor, *A Brief History of Data Anonymization*, Aircloak, (Sept. 23, 2019), https://aircloak.com/history-of-data-anonymization/.

methods have not changed since the 1990s, with some techniques, such as the addition of noise, originating back to the 1970s.[83]

Failures of Data Anonymization

Despite attempts to make datasets anonymous, numerous studies show that companies rarely prevent "anonymous" data from being re-identified. Researchers have known about data anonymization failures since the early 1980s.[84] In discussing the tension between data utility and privacy, one data scientist predicted there can be no guarantee that sensitive information will not be revealed when using data.[85]

Using a model that can predict re-identification accuracy, one study found on average fifteen characteristics from one anonymized dataset is enough to re-identify 99.98% of Americans.[86] Most datasets contain well over fifteen characteristics and have hundreds of characteristics per individual; one dataset accidentally published by analytics company Alteryz in 2017 contained 248 categories of data for 123 million American households.[87] The study also found that even when population data are very similar, individuals can still be re-identified, de-bunking the common justification that data from populations with low uniqueness are sufficiently de-identified to be anonymous.[88]

Location datasets can easily be de-anonymized to match individuals to their data characteristics. A study tracking customer location data from a European cell phone provider in a single day found that just three anonymous locations were sufficient to match 80% of customers,

[83] Sophie Bushwick, *"Anonymous" Data Won't Protect Your Identity*, Scientific American, (Jul. 23, 2019), https://www.scientificamerican.com/article/anonymous-data-wont-protect-your-identity/; Sartor, *supra* note 82.

[84] Id.

[85] Id.

[86] Hendrickx, *supra* note 68.

[87] Tracey Lien, *Alteryx data breach exposed 123 million American households' information*, Los Angeles Times (Dec. 22, 2017), https://www.latimes.com/business/technology/la-fi-tn-alteryx-data-breach-20171222-story.html.

[88] Hendrickx, *supra* note 68.

and four locations matched 95%.[89] Researchers noted this kind of matching could be facilitated with a host of other types of location information, not just geolocation information from phone providers.[90] Passive location information collection that could be used includes data from car insurance companies tracking safe driving behaviors of a vehicle, road-side cameras, photo metadata, and records of IP addresses every time a new webpage is visited.[91] Active location information collection can come from geo-tagging photos on social media platforms and posting reviews on websites like TripAdvisor.[92] Any of these anonymous locations could be one of the three or four pieces of information that can be used to match the customer it came from.

Another de-anonymization study focusing on location data found that just by combining two datasets tracking individuals traveling to three to four locations in one day, researchers could match the individual to their destinations 16.8% of the time by tracking their movement over one week.[93] When the dataset was expanded to contain information over four weeks, the matching success rate was increased to 55%.[94] The collection of mobility traces is a growing trend and as it continues, the ability to match individuals from anonymized datasets based on their geolocation data is only going to increase.[95]

Not only have researchers been able to match individuals to their respective attributes in "anonymized" location data, researchers at the University of Melbourne discovered how easy re-identification could be used to find information on specific individuals in a dataset. The

[89] Apostolos Pyrgelis et al., *There goes Wally: Anonymously sharing your location gives you away*, 1, 7 (Nov. 15, 2018), https://www.semanticscholar.org/reader/0ca4c04f06969bc9910c0828f45889040635825f.
[90] Id. at 1.
[91] Id.
[92] Id.
[93] Daniel Kondor et al., *Towards Matching User Mobility Traces in Large-Scale Datasets*, (Sept. 24, 2018), https://ieeexplore-ieee-org.proxygt-law.wrlc.org/stamp/stamp.jsp?tp=&arnumber=8470173.
[94] Id.
[95] Id.

researchers were able to re-identify themselves and others using data publicly released from 15 million Myki smart cards, the ticketing system for public transport in Victoria, Australia.[96] Two billion touch-on and touch-off events over a period of three years were publicly released.[97] Each event contained a card identifier (an anonymized card identifier number), the card type, the time, and the location of the event.[98] By searching for just two of the times the researchers knew they had personally used their Myki cards, they were able to identify their own card.[99] They further identified the Myki cards of a co-traveler with one of the researchers by searching for other cards who traveled at the same time, with the only other additional piece of information being the co-traveler was an adult commuter.[100] They were even able to pick out the Myki card for a stranger, a member of the Victorian Legislative Assembly, simply by searching for state parliamentarian travel passes (there are 424 total) that visited a certain train station in his district more than once.[101] This study shows how the ease of re-identification can have serious consequences far beyond simply matching individuals to their attributes in a dataset: "ex-partners, one-time acquaintances, or other parties can determine places of home, work, times of travel, co-travelling patterns."[102] The researchers were particularly concerned with how this would impact vulnerable groups such as people trying to escape abuse or unwanted attention.[103]

Location data isn't the only area where anonymization techniques fail. A Princeton study found that de-anonymizing web browsing data from Google Chrome could be used to identify

[96] Chris Culnane et al., *Stop the Open Data Bus, We Want to Get Off*, 1, 2 (Aug. 15, 2019), https://arxiv.org/pdf/1908.05004.pdf.

[97] Id.

[98] Id.

[99] Id. at 3.

[100] Id. at 3-4.

[101] Id. at 4.

[102] Id. at 2.

[103] Id.

users' Twitter profiles.[104] Researchers asked active Twitter users to volunteer their browser history anonymously to try to identify their Twitter profile by matching the profile feed to commonalities in browsing history.[105] Out of 374 Twitter users, 72% of users with a browsing history of between 50-75 URLs were successfully identified by the researchers' algorithm.[106] Researchers then looked to see if four common third-party trackers—Google, Facebook, ComScore, and AppNexus—could use their data from tracking browsing history to do the same thing.[107] They found each third-party tracker was pervasive enough to de-anonymize the data with a similar success rate; Google's was the highest at 70-80% for 50-75 URLs.[108]

The FTC has already recognized the dangers of anonymization practices in their complaint against Kochava for violating Section 5(a) of the FTC Act, 15 U.S.C. § 45(a). The complaint said the location data Kochava sold was not anonymous as it contained each device's mobile advertising ID (MAID).[109] They said when combined with the time stamped coordinates of the data, the MAIDs could be used to identify the owner of the device.[110] Personal information, such as the individuals' home addresses, can be learned about the individuals from combining these two characteristics.[111]

Data anonymization practices, if utilized correctly, can lessen the harms that sharing data to third parties has because they are meant to prevent the data from being traced back to the individuals it was collected from. However, the handful of studies represented here show just a

[104] Jessica Su et. al, *De-anonymizing Web Browsing Data with Social networks*, (2017), https://www.cs.princeton.edu/~arvindn/publications/browsing-history-deanonymization.pdf.
[105] Id.
[106] Id.
[107] Id.
[108] Id.
[109] Complaint, *FTC v. Kochava Inc.*, Case No. 2:22-cv-377.
[110] Id.
[111] Id.

slice of the significant failures in these anonymization practices across various types of datasets that expose consumers to risk of data related attacks. Data anonymization practices, even the more secure methods, do not adequately protect individuals from having their data re-identified. If the FTC chooses to have lower standards for anonymized data or makes rules that would not subject anonymized data to the same requirements as personally identified data, there must be standards for how those anonymization practices should work to lessen the risk that they pose to individuals. This means creating both technical rules, like allowable anonymization techniques and a standard for when data has been re-identified, as well as policy protections, like requiring companies that sell or share anonymized data to have contractual prohibitions against reidentification and stating which data types or categories will never be considered anonymized.

IV. ARTIFICIAL INTELLIGENCE DESERVES SPECIAL SCRUTINY FROM THE COMMISSION

To effectively regulate any automated decision-making system, the Commission must first address the glut of algorithms and AI systems that are currently on the market which do not work. Researchers[112] have compiled evidence of defective AI systems and significant errors in algorithmic performance across industries, including in online content moderation,[113] education systems,[114] health care,[115] and medical insurance.[116] These errors cause real-world harms. At the

[112] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz & Andrew D. Selbst, *The Fallacy of AI Functionality*, 2022 ACM Conference on Fairness, Accountability, and Transparency (2022).

[113] Jesse O'Neill, *Facebook cracks down on discussing 'hoes' in gardening group*, NY Post (July 20, 2021), https://nypost.com/2021/07/20/facebook-cracks-down-on-discussing-hoes-in-gardening-group/.

[114] Mark A. Paige and Audrey Amrein-Beardsley, *"Houston, We Have a Lawsuit": A Cautionary Tale for the Implementation of Value-Added Models for High-Stakes Employment Decisions*, 49 Educational Researcher 350 (2020).

[115] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 Science 447 (2019); Karoline Freeman, Julia Geppert, Chris Stinton, Daniel Todkill, Samantha Johnson, Aileen Clarke & Sian Taylor-Phillips, *Use of Artificial Intelligence for Image Analysis in Breast Cancer Screening Programmes: Systematic Review of Test Accuracy*, 2021 BMJ 374.

[116] Rashida Richardson, Jason M. Schultz & Vincent M. Southerland, *Litigating Algorithms: 2019 US Report*, AI Now Institute (Sept. 2019).

same time, inaccurate perceptions of current AI capabilities confuse consumers, rendering them unable to predict, identify, or remedy the harms that defective AI products cause.

The FTC already holds the necessary authority to protect consumers from dysfunctional and underperforming algorithmic decision-making systems. We believe that, given the increasing ubiquity of AI-enabled systems, and their unique opacity to consumers, these issues merit special attention in the upcoming rulemaking.

Defective Performance of Artificial Intelligence-Enabled Systems

Claims made about the capabilities and efficacies of algorithms do not match up with reality. As one of us has written elsewhere, AI is too often unproven.[117] In reality, algorithmic decision-making systems can, and frequently do, suffer from a wide array of deficiencies and biases. There are a variety of reasons that AI systems can "fail." A system may be assigned a task that is conceptually or practically impossible; it may be the product of an imprecise translation of business objectives;[118] it may interact with the real world in unexpected ways, with too few risk-mitigation measures; or it may suffer from overstated or misrepresented capabilities.[119] Additionally, the data on which an algorithm relies may preclude proper performance, if the data is inaccurate, incorrectly coded, fragmented, or incomplete.[120]

[117] Sara Collins, *21st Century Snake Oil: The Consequences of Unregulated, Unproven AI*, Tech Policy Press (October 13, 2021).

[118] Samir Passi & Solon Barocas, *Problem Formulation and Fairness*, Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (2019).

[119] Raji, et al., *supra* note 112.

[120] Sharona Hoffman & Andy Podgurski, *Big Bad Data: Law, Public Health, and Biomedical Databases*, 41 J. Law Med. Ethics 56 (2013); Milena A. Gianfrancesco, Suzanne Tamang, Jinoos Yazdany & Gabriela Schmajuk, *Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data*, 178 JAMA Intern. Med. 1544 (2018). For one example of the harms that can be caused by such issues, see the "unsafe and incorrect" recommendations for cancer patients generated by IBM's artificial intelligence unit, "Watson." Eliza Strickland, *IBM Watson Heal Thyself: How IBM Watson Overpromised and Underdelivered on AI Health Care*, IEEE Spectrum, https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care; Casey Ross, Ike Swetlitz, Rachel Cohrs, Ian Dillingham, Nicholas Florko & Maddie Bender, *IBM's Watson Supercomputer Recommended 'Unsafe and Incorrect' Cancer Treatments, Internal Documents Show*, STAT News (July 25, 2018), https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-

While “prediction” is used loosely in the world of computer science, it carries a specific connotation in conversations among consumers and policymakers: the forecasting of future events. While advances in AI have led to breakthrough performances in a variety of tasks, there are significant limits on the ability of algorithms to predict—in the colloquial sense—future outcomes.[121] For instance, algorithmic models have performed poorly, and no better than two-variable regression models, in predicting recidivism.[122] In one study of life outcomes, even highly complex models working off of very rich datasets generated objectively inaccurate predictions.[123] In a society that mistakenly believes that bigger (or more variables) is better, we may be incentivizing the construction of models that are *worse* predictors than the simple ones we already have, yet are trusted more.[124] And this says nothing of the range of other concerns that plague large models such as privacy and racial inequality.[125]

But none of this is clear to the average consumer. Nor is the potential for bias in AI systems. These systems are prone to perpetuating, and sometimes even amplifying, historical prejudices.[126] As one technologist put it: “Machine learning is like money laundering for bias. It’s a clean, mathematical apparatus that gives the status quo the aura of logical inevitability.”[127]

incorrect-treatments/?utm_source=STAT+Newsletters&utm_campaign=beb06f048d-MR_COPY_08&utm_medium=email&utm_term=0_8cab1d7961-beb06f048d-150085821.

[121] Arvind Narayanan & Matt Salganik, “Limits to prediction: pre-read,” Princeton University (Sept. 1, 2020);

[122] Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 Science Advances (2018).

[123] Matthew J. Salganik, et al., *Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration*, 117 Proceedings of the National Academy of Sciences 15 (2020).

[124] Jung, et al., *supra* note 117.

[125] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, *On the dangers of stochastic parrots: Can language models be too big?*, ACM Conference on Fairness, Accountability, and Transparency (2021).

[126] A. Caliskan, J. J. Bryson, A. Narayanan, *Semantics derived automatically from language corpora contain human-like biases*. Science 356, 183–186 (2017).

[127] Maciej Ceglowski, *The Moral Economy of Tech*, Society for the Advancement of Socio-Economics (SASE) Conference (June 26, 2016).

These problems are causing real-world harms, including the exacerbation of existing inequities. Algorithms used to allocate health care have been shown to make different decisions for groups of different races.[128] Algorithms and associated data mining techniques used for employment decision-making risk replicating the biased hiring practices that currently exist, while simultaneously obfuscating scrutiny of harm.[129] Facial recognition software performs differently on people of varying races and genders, and is especially likely to misidentify the gender of women with dark skin.[130] The federal government itself has performed studies that found facial recognition technologies to be less reliable for Asian and African-American faces relative to Caucasian faces.[131] This is in part because of the effects of biased, inaccurate, or skewed data, which cause significant harm in many contexts and may be especially pernicious in so-called "predictive policing."[132]

Failures of predictive systems in the justice system are particularly perilous. COMPAS, a recidivism risk prediction tool, is used across the nation in pretrial, parole, and sentencing decisions. Researchers at Dartmouth, however, showed that the capabilities of the tool were grossly overstated, and that the performance of the tool was indistinguishable from predictions made by individuals with little to no criminal justice expertise.[133] Most troublingly, analysis by ProPublica "found that black defendants were far more likely than white defendants to be

[128] Heidi Ledford, *Millions of Black People Affected by Racial Bias in Health-Care Algorithms*, Nature (Oct. 26, 2019), https://www.nature.com/articles/d41586-019-03228-6.

[129] Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 Wm. & Mary L. Rev. 857 (2017), https://scholarship.law.wm.edu/wmlr/vol58/iss3/4.

[130] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 Proceedings of Machine Learning Research 1 (2018).

[131] Chad Boutin, *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*, National Institute of Standards and Technology, Department of Commerce (Dec. 19, 2019), https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software.

[132] Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. Rev. 192 (2019).

[133] Julia Dressel & Hany Farid, *The accuracy, fairness, and limits of predicting recidivism*, Science Advances 4(1), 2018.

incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk."[134] Put simply, countless individuals were required to stay in jail for no other reason than the color of their skin.

In another use of predictive systems by the justice system, a man in Chicago was targeted for surveillance by the police because an algorithm used by the police department predicted that he was likely to be involved in a shooting. The surveillance attracted attention from his neighbors, who decided he had formed a relationship with the Chicago police–and shot him multiple times in the leg.[135]

One set of researchers, compiling harms perpetrated at least in part via algorithmic decision-making, found an astounding pile of injustices:[136] state governments falsely flagging unemployment benefit fraud,[137] landlords using tenant screenings that pushed out a family on the basis of a false arrest record,[138] health administrators withdrawing benefits on corrupted formulae and outdated information,[139] and test monitoring services canceling the visas of suspected cheaters.[140]

[134] Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, How We Analyzed the COMPAS Recidivism Algorithm, ProPublica (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[135] Matt Stroud, *Heat Listed*, The Verge (May 24, 2021), https://www.theverge.com/c/22444020/chicago-pd-predictive-policing-heat-list.

[136] Raji, et al., *supra* note 112.

[137] Robert Charette, *Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold*, 18 IEEE Spectrum 3 (2018).

[138] Lauren Kirchner and Matthew Goldstein, *Access Denied: Faulty Automated Background Checks Freeze Out Renters*, The Markup (May 28, 2020), https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters.

[139] Colin Lecher, *What happens when an algorithm cuts your health care*, The Verge (Mar. 21, 2018); Jay Stanley, *Pitfalls of Artificial Intelligence Decisionmaking Highlighted in Idaho ACLU Case*, ACLU Blogs, https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-Aclu-case.

[140] United Kingdom National Audit Office, Investigation into the Response to Cheating in English Language Tests, (May 24, 2019), https://www.nao.org.uk/press-release/investigation-into-the-response-to-cheating-in-english-language-tests/.

Inaccurate Perceptions of Artificial Intelligence Functionality

Advanced technologies tend to be opaque to the ordinary person. As of 2017, few people in the United Kingdom understood—or were even aware of—the phrase "machine learning."[141] In the United States in 2022, researchers found that Americans' technical understanding of artificial intelligence was still "patchy."[142]

For the average consumer, then, exposure to artificial intelligence comes from popular media, whose fictions "paint[] a distorted image of the present potential and functionality of the technology."[143] Narratives around new technologies tend to be "sensationalist" and "misinformed"—but they are also powerful influences on how the technologies are perceived and governed.[144] A 2022 study demonstrated "a significant relationship between people's beliefs about AI in entertainment media and their beliefs about AI in reality."[145] News media's similarly exaggerated stories of successful AI, as well as out-of-control AI, contributes to the problem.[146] One computer science professor is concerned about an "AI misinformation epidemic."[147]

All of this has caused significant confusion among policymakers and the public about the powers and limits of artificial intelligence and machine learning systems. For instance, a majority of the American public believes that high-level machine intelligence—that is, machines

[141] The Royal Society, Machine Learning: The Power and Promise of Computers that Learn by Example (2017), www.royalsociety.org/machine-learning.

[142] Karim Nader, Paul Toprac, Suzanne Scott & Samuel Baker, *Public Understanding of Artificial Intelligence Through Entertainment Media*, AI & Society (2022).

[143] Isabella Hermann, *Artificial Intelligence in Fiction: Between Narratives and Metaphors*, AI & Society (2021).

[144] Stuart Roberts, *From Homer to HAL: 3000 Years of AI Narratives*, University of Cambridge (Dec. 11, 2019), https://www.cam.ac.uk/ainarratives.

[145] Nader, et al., *supra* note 142, at 10.

[146] Oscar Schwartz, *'The Discourse is Unhinged': How the Media Gets AI Alarmingly Wrong*, The Guardian (July 25, 2018), https://www.theguardian.com/technology/2018/jul/25/ai-artificial-intelligence-social-media-bots-wrong; Sissi Cao, How Tech Coverage Has Turned Meaningful AI News into 'Sensationalized Crap,' The Observer (Aug. 1, 2018), https://observer.com/2018/08/media-tech-coverage-dangerous-artificial-intelligence-scientists-say/.

[147] Schwartz*, supra* note 146 (quoting Zachary Lipton, Assistant Professor of Machine Learning and Operations Research at Carnegie Mellon University).

that are able to perform "almost all tasks that are economically relevant today better than the median human"—will be achieved within ten years.[148] This is exceedingly optimistic; experts predict that artificial general intelligence (AGI) is about 50 years away.[149] This problem is only exacerbated by irresponsible claims from prominent members in the field. In 2016, Geoffrey Hinton, a pioneer in deep learning, confidently proclaimed that artificial intelligence would outperform radiologists within five years, a prediction that has failed to materialize.[150]

Media narratives and popular expectations are affecting business calculations. As the Royal Society observed:

> *The prevalence of narratives focused on utopian extremes can create expectations that the technology is not (yet) able to fulfill. This in turn can contribute to a hype bubble, with developers and communicators potentially feeding into the bubble through over-promising."*[151]

Researchers have found that companies developing machine learning systems, and their investors, do in fact feel incentivized and even pressured to participate in the "AI hype cycle."[152]

These misaligned incentives foster irresponsible and sometimes even willfully deceptive advertising. For instance, the Epic Sepsis Model, which purported to predict the onset of sepsis, was found by external validators to perform "substantially worse" than the company's reports

[148] Baobao Zhang & Allan Dafoe, *Artificial Intelligence: American Attitudes and Trends*, Center for the Governance of AI, Future of Humanity Institute, University of Oxford (2019).

[149] Arvind Narayanan, "How to Recognize AI Snake Oil," Princeton University (2019).

[150] Gary Marcus, *Deep Learning is Hitting a Wall*, Nautilus (Mar. 10, 2022), https://nautil.us/deep-learning-is-hitting-a-wall-238440/.

[151] The Royal Society, Portrayals and Perceptions of AI and Why they Matter 14 (2018), https://royalsociety.org/-/media/policy/projects/ai-narratives/AI-narratives-workshop-findings.pdf.

[152] Amy A. Winecoff & Elizabeth Anne Watkins, *Artificial Concepts of Artificial Intelligence: Institutional Compliance and Resistance in AI Startups*, Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (2022).

had suggested.[153] Similarly, Amazon touted the performance of its facial recognition product "Rekognition" to police departments. A test by the ACLU, however, found that the product incorrectly identified 28 members of Congress as individuals who had previously been arrested for a crime.[154] A timeline of events reconstructed by the ACLU showed that Amazon had misrepresented the confidence thresholds that police should utilize when employing the system.[155]

Consumers don't know that these harms are perpetrated—or even that they are possible—because of a widespread misperception of algorithmic capability, which (as noted above) has been perpetuated by companies, policymakers, and media entities alike. Regulators and entrepreneurs like to speak of building "trustworthy" AI. But AI is not intrinsically trustworthy simply because it is AI, and it is no more to be trusted than are other, less advanced technologies.[156]

Moreover, consumers of AI-powered goods and services often are unaware that their expectations are unmet. Purchasers of products billed as capable of clearing clogged drains will know within hours of purchase whether the product performed as expected. By contrast, individuals who buy "off-the-shelf" AI-enabled products—like a landlord purchasing software that purports to evaluate the reliability of a prospective renter, or an HR representative using a model that claims to estimate job candidates' suitability—are unable to test and verify these

[153] Andrew Wong, Erkin Otles & John P. Donnelly, *External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients*, 181 Jama Internal Medicine 1065 (2021).

[154] Jacob Snow. 2018. Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots. https://www.aclu.org/blog/privacy-technology/ surveillance-technologies/amazons-face-recognition-falsely-matched-28.

[155] ACLU. 2018. ACLU Comment on New Amazon Statement Responding to Face Recognition Technology Test. https://www.aclu.org/press-releases/aclucomment-new-amazon-statement-responding-face-recognition-technologytest.

[156] Mark Ryan, *In AI We Trust: Ethics, Artificial Intelligence, and Reliability*, 26 Science and Engineering Ethics 2749 (2020).

tools' accuracy. Only when researchers undertake an expensive project of external validation can these models be properly assessed as in the case of the aforementioned Epic Sepsis Model and Amazon "Rekognition" product.[157]

The Time for FTC Action to Address Harms from AI Is Now

Whether arising from algorithmic decision-making products that are objectively dysfunctional or from the underperformance of such products relative to their perceived or advertised capabilities, the harms of AI systems are only growing more prevalent. Consumers, already confused by media narratives, have no way of knowing if or when they are duped or harmed. FTC intervention in the harms that emanate from "everyday AI"—harms that will proliferate as AI becomes increasingly ubiquitous—is therefore a necessity. The Commission in this proceeding must undertake swift and comprehensive regulatory action to protect consumers from the harms that flow from the false promises made about algorithmic decision-making, machine learning systems, and their capabilities. Consumers cannot protect themselves from what they do not understand and cannot test effectively. We know that regulation, such as established evaluation processes and rigorous accident reporting, can ease burdens on consumers and reduce harm.[158]

The range of concerns and instances of harm highlighted above make clear that harms of AI systems are the result of prevalent and deceptive practices. Situations like these demand FTC action.

[157] Andrew Wong, Erkin Otles & John P. Donnelly, *External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients*, 181 Jama Internal Medicine 1065 (2021).

[158] *See* Stan Benjamens, Pranavsingh Dhunnoo & Bertalan Meskó, *The State of Artificial Intelligence-based FDA-approved Medical Devices and Algorithms: An Online Database*, 3 NPJ Digital Medicine 1 (2020); Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J. Calvert & Alastair K. Denniston, *Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension*, 2020 BMJ 370.

V. REMEDIES FOR COMMERCIAL SURVEILLANCE

Commercial surveillance and the steady erosion of consumer privacy is driven by the mass data collection business model.[159] As the Commission works to identify and prohibit the pervasive unfair and deceptive practices in this arena, it must simultaneously arm itself with remedies that cut to the heart of this harmful activity. Equitable remedies, such as data deletion and algorithmic disgorgement, allow the Commission to directly address mass data collection and have already started to become an important part of its toolkit in enforcement actions. The Commission's future trade regulation rules should specifically identify data deletion and algorithmic disgorgement as additional remedies available to it for enforcing rules on commercial surveillance. The Commission has broad authority to seek equitable relief when enforcing its trade regulation rules, and there is no limitation on identifying these potential remedies in advance as a component of its trade regulation rules.[160]

Equitable Remedies are Necessary to Combat Commercial Surveillance

For the Commission's trade regulation rules on commercial surveillance to be successful, they must be backed up with remedies and civil penalties that deter bad actors and deliver justice for those harmed by commercial surveillance.

Mass data collection has become so profitable, and the data gleaned from it so fundamental and valuable to the attention exploitation business model, that the Commission will need to go beyond simple monetary fines to properly deter violators. It is a challenge to parse the long-term value of large data sets, or the value to existing data sets of a new collection of inputs,

[159] *See generally* Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight For A Human Future at the New Frontier of Power* (2019).

[160] 15 U.S.C. § 45(l) ("district courts are empowered to grant mandatory injunctions and such other and further equitable relief as they deem appropriate in the enforcement of such final orders of the Commission"); s*ee also* Telebrands Corp. v. FTC, 457 F.3d 354, 357 n.5 (4th Cir. 2006) (holding that the FTC may seek injunctions containing provisions that are broader than the conduct that is declared unlawful).

and so it is unlikely that fines or even traditional profit disgorgement would be sufficient to capture the real value of a specific violation and properly align incentives against data collection.

Equitable remedies are also necessary tools for directly rectifying the privacy harms stemming from bulk data collection. Equitable remedies can reach beyond calculated deterrence through economic sanction, towards genuine restorative justice that begins to rectify the harms of pervasive commercial surveillance. By ensuring that unlawfully obtained data is deleted, and that any algorithms and models developed using that data are destroyed, the Commission can return some sense of privacy to those affected by unlawful commercial surveillance. As such, data deletion and algorithmic disgorgement are specific examples of equitable relief that could have powerful deterrent effects as well as a restorative component.

Trade Regulation Rules Should Include Equitable Remedies

As part of the ANPR, the Commission has inquired about the possibility of enumerating "specific forms of relief or damages that are not explicit in the FTC Act but that are within the Commission's authority."[161] The Commission should do so and should specifically identify data deletion and algorithmic disgorgement as equitable remedies within its authority that it may pursue in enforcing new rules on commercial surveillance and data security. The Commission has already recognized the challenges posed by conventional remedies in enforcement and embraced the use of alternative remedies in its enforcement actions. The challenges posed by pervasive commercial surveillance and mass data collection require the Commission to continue to think creatively about mechanisms for redress and enforcement. Equitable remedies that target the misbegotten gains from violations—specifically, the data sets, data models, and algorithmic products developed from them—are necessary to secure justice for consumers and realign

[161] Advanced Notice of Proposed Rulemaking, Trade Regulation Rule on Commercial Surveillance and Data Security, 87 Fed. Reg. 51273, 51285 (Aug. 22, 2022) ("ANPR").

commercial incentives towards more responsible data privacy practices. Data deletion and algorithmic disgorgement serve different, but complementary, roles in enforcement and have already been deployed by the Commission in a variety of enforcement cases.

Data Deletion

Data deletion is generally a straight-forward and common-sense response to the unlawful collection, acquisition, or retention of data. Simply put, it is the requirement that the violator be required to delete the unlawfully obtained data. This simple form of remedy is the obvious response to harmful practices related to mass data collection: go after the data. Data deletion as a right or remedy is also enshrined within statutory privacy statutory schemes, including the European Union's GDPR.[162] This means that mechanisms and best practices for data deletion implementation and enforcement, while not fully developed and subject to some shortcomings, are likely to continue to grow and mature. Additionally, data deletion conceived as a broad category of equitable remedy can also be deployed in other forms, such as to require proactive and ongoing data minimization or to efficiently reach parties not immediately before the Commission.

Data deletion has been deployed alongside algorithmic disgorgement, discussed further below, but also as a separate requirement in appropriate circumstances, including with some uses that reach beyond the most direct implementations. For example, in an enforcement action against CafePress for failure to safeguard data, the Commission alleged, in part, that CafePress retained sensitive information longer than necessary and without proper safeguards.[163] As a result, in its final Order, the Commission required CafePress to implement "[p]olicies and

[162] General Data Protection Regulation, *supra* note 60 at Article 17.

[163] In the Matter of Residual Pumpkin Entity, LLC formerly d/b/a/ CafePress, et. al.. Docket C-4768, Complaint, https://www.ftc.gov/system/files/ftc_gov/pdf/CafePress-Complaint_0.pdf

procedures to minimize data collection, storage, and retention, including data deletion or retention policies and procedures."[164] In this example, data deletion is included as a component of a broader mandated information security program, demonstrating the flexibility of this particular remedy. Rather than just identifying a specific set of data be deleted, as might be the case where a swathe of data has been collected unlawfully, data deletion can also be used to enforce proactive data minimization policies to guard against the data hoarding that has become worryingly commonplace in the data-driven economy.

Another recent example of the use of data deletion is in the enforcement action taken against Flo which was alleged to have shared health information with outside data analytics providers, despite promises of data privacy.[165] Critically, in this instance the problem facing the Commission was that the data was shared, rather than being improperly obtained or retained. As a result, the data deletion requirement imposed in the final Order required Flo to "instruct any Third Party that has received Health Information from [Flo Health, Inc.] belonging to any Covered App User to destroy such information."[166] This use of data deletion relief provides efficiency in enforcement, reaching the various third parties that benefitted from Flo's misleading data practices without requiring a multitude of enforcement actions to claw back the data from each party.

Any future rules on commercial surveillance and data privacy should specifically identify data deletion as a potential remedy and be written with the flexibility of equitable remedies in mind to ensure that data deletion can be used creatively.

[164] In the Matter of Residual Pumpkin Entity, LLC formerly d/b/a/ CafePress, et. al.. Docket C-4768, Decision and Order at 4 (Jun. 23, 2022), https://www.ftc.gov/system/files/ftc_gov/pdf/192%203209%20-%20CafePress%20combined%20package%20without%20signatures.pdf.

[165] Flo Health Complaint, *supra* note 30.

[166] In the Matter of Flo Health, Inc., Docket C-4747, Decision and Order at 4 (Jun. 22, 2021), https://www.ftc.gov/system/files/documents/cases/192_3133_flo_health_decision_and_order.pdf

However, the effectiveness of data deletion as a remedy should not be overstated and it cannot be the only mechanism for enforcement. Effective data deletion requirements will require strong, clear rules to protect data privacy and limit data collection to ensure they can reach the various parties involved in the complex data ecosystem. Additionally, the specifics of implementing and enforcing data deletion orders can be a challenge. When targeting specific data sets, that data may have been split, anonymized, or combined with other data sets, making it difficult to identify and erase.[167] Even erring on the side of over-inclusive enforcement (e.g. erasing any data set where any component of illicit data may have been integrated), there remains a challenge of ensuring compliance when data is so easily copied, transferred, and disguised.

Finally, it must be recognized that data deletion, considered alone, has important limits. In the context of machine learning systems, the problem of the "algorithmic shadow" means that data deletion alone cannot ensure that a bad actor is no longer able to benefit from deleted data. The algorithmic shadow results from how machine learning models work:

> *"When you feed a set of specific data to train a machine learning model, that data produces an impact on the model that results from such training. Even if you later delete data from the training data set, the already-trained model still contains a persistent "shadow" of the deleted data."*[168]

[167] *Algorithms All the Way Down,* PROTOCOL, Mar. 17, 2022, https://www.protocol.com/newsletters/protocol-enterprise/ftc-algorithmic-disgorgement-japan-chips ("Data obtained through deceptive means may end up in a data set that is then sliced and diced to form multiple data set "splits," each used for separate purposes throughout the machine-learning model development process for model training, testing and validation, said Anupam Datta, co-founder and chief scientist at TruEra, which provides a platform for explaining and monitoring AI models.")

[168] Tiffany Li, *Algorithmic Destruction*, SMU LAW REVIEW (forthcoming 2022), available at https://ssrn.com/abstract=4066845.

This challenge is not insurmountable: algorithmic disgorgement, which targets algorithms and machine learning models alongside the underlying data, is an established solution to the problem of the algorithmic shadow.[169] But it is essential to highlight this shortcoming to make clear that data deletion alone is an insufficient remedy to redress the harms of mass data collection in the era of machine learning.

Despite the implementation challenges and limitations, data deletion should be a key component of a remedial strategy. Specifically identifying data deletion as an equitable remedy to be used to enforce commercial surveillance rules, and using it in creative ways, will give the Commission an important enforcement tool for protecting consumers. However, data deletion is still only one tool, and should be used alongside traditional civil penalties and other equitable forms or relief, like algorithmic disgorgement, where appropriate.

Algorithmic Disgorgement

The ANPR seeks comment on whether algorithmic disgorgement, described as “a remedy that forbids companies from profiting from unlawful practices related to their use of automated systems” should be enumerated as a potential remedy in future trade regulation rules about commercial surveillance.[170] We encourage the Commission to specifically identify algorithmic disgorgement as an available remedy in future rules governing commercial surveillance and data collection. Algorithmic disgorgement is a potent remedy which creates strong deterrent effects against data exploitation and serves as a critical update to the traditional monetary disgorgement

[169] *Id.* (“While algorithmic disgorgement may resolve some of the failures of data deletion, this remedy and potential right is also not without its own drawbacks, even though it does address the harms of the algorithmic shadow.”)

[170]Trade Regulation on Commercial Surveillance and Data Security, 87 Fed. Reg. 51273, 51285 (ANPR Aug. 22, 2022).

remedy. Additionally, as touched upon above, algorithmic disgorgement solves the issue of the algorithmic shadow that data deletion alone is unable to address.

Like data deletion, a plain-language description of algorithmic disgorgement reveals a common-sense premise: "when companies collect data illegally, they should not be able to profit from either the data or any algorithm developed using it."[171] Algorithmic disgorgement has already been used by the Commission in three cases. It was first used in the case against Cambridge Analytica,[172] then against Everalbum,[173] and most recently against W.W. International.[174] In each of these cases, it was deployed to rectify serious violations of consumer privacy and ensure that violators would be unable to profit from their ill-gotten gains. Analysis of algorithmic disgorgement in the wake of these cases has recognized its powerful deterring effects.[175] Unlike monetary fines, which are often simply regarded as "the cost of doing business,"[176] algorithmic disgorgement acts as a true deterrent to flaunting Commission rules because it forces companies to change how they do business, while ensuring that companies cannot continue to profit from ill-gotten data.

[171] Rebecca K. Slaughter, Janice Kopec, & Mohamad Batal, *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*, YALE JOURNAL OF LAW AND TECHNOLOGY 23: 1-59 (Aug. 2021), https://yjolt.org/sites/default/files/23_yale_j.l._tech._special_issue_1.pdf ("Algorithms and Economic Justice").

[172] In the Matter of Cambridge Analytica, LLC, Docket 9383, Final Order at 4 (Dec. 6, 2019), https://www.ftc.gov/system/files/documents/cases/d09389_comm_final_orderpublic.pdf.

[173] In the Matter of Everalbum, Inc., File No. 192 3172, Agreement Containing Consent Order (Jan. 11, 2021), https://www.ftc.gov/system/files/documents/cases/everalbum_order.pdf.

[174] U.S. v. Kurbo Inc., et. al., No. 3:22-cv-00946-TSH (N. D. Cal. Mar. 3, 2022).

[175] *See e.g.* Jevan Hutson & Ben Winters, *America's Next 'Stop Model!': Algorithmic Disgorgement* (September 20, 2022), available at https://ssrn.com/abstract=4225003 ("Hutson & Winters"); see also Tiffany Li, *Algorithmic Destruction*, SMU Law Review (forthcoming 2022), available at https://ssrn.com/abstract=4066845; Avi Gesser, Paul D. Rubin, & Anna R. Gressel, *Model Destruction – The FTC's Powerful New AI and Privacy Enforcement Tool*, COMPLIANCE & ENFORCEMENT BLOG, Mar. 30, 2022, https://wp.nyu.edu/compliance_enforcement/2022/03/30/model-destruction-the-ftcs-powerful-new-ai-and-privacy-enforcement-tool/ ("Model Destruction").

[176] Kate Kaye, *The FTC's new enforcement weapon spells death for algorithms*, PROTOCOL, Mar. 14, 2022, https://www.protocol.com/policy/ftc-algorithm-destroy-data-privacy.

The deterrent effect of algorithmic disgorgement is what makes specifically identifying it in future trade regulation rules on commercial surveillance critically important. Algorithmic disgorgement should be viewed as a remedy "designed to change the behavior of both people and machines. Not only does it change the decisions people must make in the course of developing artificial intelligence and machine learning technologies, it will invariably change the performance of deployment of those very technologies."[177] Naturally, to have this preventative, prospective, deterring effect, its intended use must be firmly and publicly established; to paraphrase Peter Sellers, the whole point of this "doomsday" remedy is lost if you keep it a secret, you must tell the world you intend to use it![178]

Aside from its deterring effect, algorithmic disgorgement, when used alongside data deletion, has a restorative justice component. As previously explained, in cases where a person's data is used in machine learning models, data deletion alone cannot fully eliminate the algorithmic shadow of that privacy harm. However, when data deletion is used in conjunction with algorithmic disgorgement, the remedies together more fully restore privacy to the victims of privacy harms. While not compensatory for the non-economic harms of privacy violations, this model of privacy restoration through algorithmic disgorgement should be a key goal of any future trade regulation rules on commercial surveillance.

Finally, as with data deletion, it must be noted that algorithmic disgorgement is not a panacea or silver bullet. There are challenges regarding implementation, compliance verification, and balancing the need for specificity with broad effect in Commission orders. In some circumstances data deletion alone, monetary fines, or other equitable relief may prove to be the

[177] Hutson & Winters at 11.

[178] DR. STRANGELOVE OR: HOW I LEARNED TO STOP WORRYING AND LOVE THE BOMB (Columbia Pictures 1964), scene available at https://www.youtube.com/watch?v=2yfXgu37iyI.

more proper remedy in a given case. Nevertheless, the Commission should affirmatively embrace algorithmic disgorgement as a potential remedy. Many of the challenges facing both data deletion and algorithmic disgorgement are attributable to the fact that these are new remedies, and as they become more established best practices will develop to address the current suite of challenges. The Commission should stay on the course it has already been charting and continue to embrace equitable remedies to deter bad actors and secure justice for those harmed by commercial surveillance.

The Commission Should Identify and Implement Equitable Remedies in Regulation

The Commission has inquired about if there is "a limit to the Commission's authority to implement remedies by regulation."[179] The Commission is limited only by the bounds of permitted relief for regulatory violations when it comes to its implementation of specific remedies. Data deletion and algorithmic disgorgement are equitable remedies with strong ties to traditional forms of equitable relief like profit disgorgement and are remedies well within the Commission's authority. Therefore, the Commission can identify and implement specific equitable relief of this kind in connection with its regulations concerning commercial privacy.

The Commission's rulemaking authority comes from Section 6(g) and Section 18 of the FTC Act. The former authorizes the Commission "to make rules and regulations for the purpose of carrying out the provisions of this subchapter" concerning unfair competition[180] and the latter provides the Commission specific authority for issuing trade regulation rules with respect to unfair and deceptive acts or practices.[181] Once established, violations of trade regulation rules are

[179] ANPR, *supra* note 170 at 51285.
[180] 15 U.S.C. § 46(g).
[181] 15 U.S.C. § 57a.

punishable through civil penalties as authorized under Section 5(m)[182] and Section 19.[183] Section 5(m) authorizes monetary penalties on a per violation basis, while Section 19 authorizes the Commission to seek any relief "necessary to redress injury to consumers or other persons, partnerships, and corporations resulting from the rule violation or the unfair or deceptive act or practice."[184] The broad language of Section 19 provides the mechanism for the Commission to obtain equitable remedies like data deletion and algorithmic disgorgement, as well as more traditional remedies like restitution and profit disgorgement. Having a broad array of equitable remedies ensures that injuries will be redressed and the conduct that created the injury will not occur again. While the award of any penalty is contingent on the judgment of the federal courts, since those remedies are within the Commission's authority to seek in connection with violations of its regulations, the Commission should therefore adopt rules that identify specific equitable remedies appropriate to enforcing its commercial surveillance rules.

VI. RULEMAKING RECOMMENDATIONS

As the FTC considers the contours of a commercial surveillance, much can be learned from the decades-long privacy discourse that the FTC has engaged in. Last year, 24 civil society organizations sent a letter[185] to the FTC outlining not only the harms that arise from the lack of privacy rules, but ideas for crafting rules.

First and foremost, the FTC could use its rulemaking power to promulgate rules based on traditional principles of fair data collection and use. This can include creating specific use limitations, like banning targeted advertising, but it can also include data minimization

[182] 15 U.S.C. § 45(m).

[183] 15 U.S.C. § 57b.

[184] 15 U.S.C. § 57b(b).

[185] Letter from Access Now et al. to the FTC Commissioners (2021), https://www.lawyerscommittee.org/wp-content/uploads/2021/08/FTC-civil-rights-and-privacy-letter-Final-1.pdf.

requirements,[186] retention and deletion rules, as well minimum-security standards. These ideas are foundational parts of the Fair Information Practice Principles and have been used by the FTC since the 2000s to assess companies and their relationship with personal data.[187]

The FTC, as the preeminent consumer protection organization within the federal government, could also draw from that well of experience. Requiring that companies who collect consumer data make that data available for access, correction, and deletion allow consumers some control of how their data is used. But providing those rights would not be enough. There must be rules in place that outline how companies must respond to those requests, as well as how an appeals process would work for requests that are denied. But the FTC should go farther. Prohibitions against requiring pre-dispute arbitration and mandatory class action waivers would ensure that consumers have the opportunity to vindicate their rights in court. Furthermore, the FTC should draw on their recent work on dark patterns[188] to create rules that would prohibit companies from manipulating their users into consenting away their data.

Last, but certainly not least, the FTC could draw upon decades of civil rights litigation and scholarship to create rules. This should include prohibitions against using protected characteristics to serve ads or make eligibility determinations in such categories as housing, employment, education, credit, or insurance—which could be modeled on protections that already exist in current law. The FTC should also require assessments for algorithmic processes

[186] *How the FTC Can Mandate Data Minimization Through a Section 5 Unfairness Rulemaking*, Electronic Privacy Information Center (Jan. 2022), https://epic.org/documents/how-the-ftc-can-mandate-data-minimization-through-a-section-5-unfairness-rulemaking/.

[187] Federal Trade Commission, Privacy Online: Fair Information Practices in the Electronic Marketplace (May. 2000), https://www.ftc.gov/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission.

[188] Federal Trade Commission, Bringing Dark Patterns to Light Staff Report (Sep. 2022). https://www.ftc.gov/news-events/news/press-releases/2022/09/ftc-report-shows-rise-sophisticated-dark-patterns-designed-trick-trap-consumers

to test for bias. These assessments would be performed before deployment to ensure that they do not directly or indirectly exclude persons from economic opportunities on the basis of a protected class or characteristic. And those assessments would be regularly performed to ensure compliance. Requiring algorithmic assessments, rather than prohibiting certain types of data from being used to train algorithms, ensures that the focus is on disparate or discriminatory outcomes, which is where the harm occurs.

The recently introduced American Data Privacy and Protection Act (ADPPA) uses all three approaches effectively, and we would recommend using ADPPA as a template for this rulemaking.

The Commission's action here is necessary, but only the first step of many. We again thank you for your work in this vital area and encourage your thoughtful consideration of the comments you will receive. We look forward to reviewing the proposed rule, and request to testify if hearings are held.

For more information, please contact Sara Collins at sara@publicknowledge.org.