

BU/MIT Student Innovations Law Clinic
765 Commonwealth Avenue
Boston, Massachusetts 02215
(617) 353-3131

Before the

Federal Trade Commission

In the Matter of

**Comments on Trade Regulation Rules Concerning Commercial Surveillance and Data
Security**

Re: Commercial Surveillance ANPR, R111004

FTC-2022-0053-1201

Comments of

Alex Berke, Dana Calacci, and Robert Mahari

Darcy Black
Jared Lockhart
Allison Works
Clinical Students
Christopher Conley
Counsel for Commenters

December 1, 2022

As interdisciplinary researchers, we use data science to surface unknown consumer harms to better empower policy makers and regulators with information. To do so, it is essential that consumers have the right both to access their personal data and to pool and share that data as they choose, including, but not limited to, providing it for research such as ours. Consumer data sharing enables us to share insights with both the public and regulators, including the FTC, about the evolving harms of automated decision making, biometric data collection, discrimination based on protected categories, and many more of the FTC's consumer surveillance concerns. Too often, however, corporate practices impede the ability of consumers to access and share their data, and therefore our ability to contribute to the effort to protect consumers from harm.

We encourage the FTC to ensure that its commercial surveillance and data security rulemaking facilitates and empowers consumers to share their data, with their informed consent, with researchers and consumer advocacy organizations, and prohibits corporate practices designed to prevent this. Doing so will permit consumers to leverage their own data to help surface hidden harms, furthering the FTC's mission and the overall aims of the rulemaking.

I. Overview

The emergence of powerful consumer surveillance technologies, in conjunction with incentives for data-driven business models, has fundamentally altered the landscape of online data collection. Consumers need to use mobile devices, ecommerce, and social media to operate in the modern world and therefore no longer have a true choice to avoid mass data collection.

Platforms' and corporations' unbridled ability to collect and monetize consumer data has led to a power asymmetry in which consumers have little information about the data being collected and how the data are used, and how those uses harm consumers. Our research seeks to address this by collecting aggregate datasets from consumers to surface harms that otherwise would remain hidden.

Corporations hinder this research by hampering efforts to aggregate consumer data. Corporations argue that data they collect from consumers is proprietary and rightfully under their control as part of the exchange for using their services. These justifications are used to deny consumers access to personal data and shield commercial surveillance practices from greater public scrutiny. (Question 83). While state and federal privacy laws have made progress in providing consumers with transparency and empowering choice, these protections are still inadequate. There are some jurisdictional data access mechanisms such as the CPRA's data access requests. However, these are often onerous due to restricted third-party requests, expensive processes, time consuming/effort intensive avenues, and limitations to jurisdictions with data access laws. As a result, consumers and the federal agencies tasked with their protection are unable to understand how consumer data are used and are therefore not able to quantify and address the harm resulting from this asymmetry. (Question 6)

FTC rulemaking that facilitates consumers pooling and sharing their data with appropriate limitations to prevent misuse would help researchers and advocacy groups surface consumer harms to benefit the public. These harms include anti-competitive practices, systemic market issues, privacy infringements, and algorithmic bias. A large pool of aggregate data in

conjunction with large-scale audit tools can help surface previously unquantifiable harms that are due to how data are used in aggregate.

II. Researchers Play an Essential Role in Surfacing Consumer Harms

A. Many Harms Arising from Commercial Surveillance Are Unknown to Consumers

Currently, consumers cannot effectively prevent platforms from either collecting their data directly or using their data to draw inferences about them unless they are willing to refrain from participating in the digital ecosystem altogether. This is neither fair nor reasonable and exacerbates the power imbalances between platforms and users. Moreover, many inferences and use cases drawn from consumer data are currently unknown to consumers.

For example, there are various hidden consumer harms associated with the use of AI tools. The use of algorithms to generate inferences can subject consumers to invisible profiling, which may be biased and discriminatory.¹ In some cases, inferring specific information is the goal, such as Target’s application of a model to identify whether a customer was likely pregnant based on their shopping history.² At other times inferences may be implicit and even unintentional, such as Amazon using machine learning to build a recruitment tool that displayed a bias against women.³ Additional harmful uses of algorithmic tools, that are currently unknown, may be brought to light with the help of researcher access to the datasets these algorithmic tools are employed on.

The FTC has examined the use of artificial intelligence and acknowledged its potential to cause harms due to inaccuracy, bias and discrimination, including through its invasive use in commercial surveillance.⁴ The FTC has also acknowledged a lack of transparency in AI applications, “and frustration with the opacity of the ‘black box’ [that] can lead consumers to feel powerless and distrustful.”⁵ These types of harms are particularly difficult for individual consumers to recognize and evaluate.

¹ Nicol Turner Lee, Paul Resnick & Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS INST. (May 22, 2019), <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.

² Kashmir Hill, *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012), <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>.

³ See Lee, Resnick & Barton, *supra* note 1; Isobel Asher Hamilton, *Why It’s Totally Unsurprising that Amazon’s Recruitment AI Was Biased Against Women*, BUS. INSIDER (Oct. 13, 2018), <https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandra-wachter-2018-10>; Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES (Feb. 16, 2012), <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

⁴ FED. TRADE COMM’N, COMBATTING ONLINE HARMS THROUGH INNOVATION 6 (2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Combating%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf.

⁵ Rebecca Kelly Slaughter, *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*, 23 YALE J.L. & TECH. 1, 48 (2021).

B. Researchers with Access to Consumer Data Can Surface These Hidden Harms

If consumers, with fully informed consent, could make their data available to researchers at scale, it would facilitate public-facing research that informs and thereby empowers the FTC to intervene in abusive data practices, while providing transparency and preserving autonomy for consumers. Data-driven research is a valuable tool to understand and thereby evaluate the impact of otherwise-invisible practices—but only if the data needed for such research are available.

Our research with Google’s proposed FLoC paradigm illustrates how researcher access to aggregate data can surface consumer harms.⁶ Google proposed FLoC as a system that would operate within browsers to enable interest-based advertising without enabling individualized user tracking. However, our research found that 95% of users could be individually identified and tracked within a few weeks of use. We were able to perform this research only by leveraging a large database of user web browsing histories to which we had privileged access to as researchers.

Researcher access to data at scale can also help to surface specific harms to individuals that cannot be seen at the individual data level. For example, we developed a calculator for gig workers on the Shipt platform to track and share aggregate data about wages.⁷ By doing this, we found that a new payment model increased the average pay-per-order but cut wages on average. At an individual level, it is hard for workers to prove these harms. In the aggregate, it became clear that there were systematic problems with Shipt’s payment model that harmed workers. Like these harms to gig workers, consumer harms are easier to detect if individuals can pool their data with others.

Empowering consumers to share data with researchers at scale would help address power asymmetries. Ultimately, consumers, operating collectively, should have meaningful control over the data they co-generate with platforms, including (but not limited to) the practical ability to aggregate it for research and other collective uses. Providing public interest researchers and consumer advocacy groups with access to large data sets ensures that consumers and their protectors—including the FTC itself—can recognize and respond to harmful uses of personal data.

III. Corporations Create Obstacles to Data Access for Researchers

A. Corporations Improperly Claim Exclusive Rights to Consumer Data

Corporate practices that conceal the use of data keep consumers and their advocates in the dark and increase corporations’ abilities to use data in ways that harm consumers. These businesses may argue that they own the data because users consented to share it in exchange for

⁶ Alex Berke & Dana Calacci, *Privacy Limitations of Interest-based Advertising on The Web: A Post-mortem Empirical Analysis of Google’s FLoC*, 2022 ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY (Nov. 7-11, 2022), <https://doi.org/10.1145/3548606.3560626>.

⁷ Dana Calacci & Alex Pentland, *Bargaining with the Black-Box: Designing and Deploying Worker-Centric Tools to Audit Algorithmic Management*, PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 6, CSCW2, Article 428, <https://doi.org/10.1145/3570601> (2022).

a business's services. This argument denies the interests that consumers have in their own data and too often deprives consumers of beneficial uses of that data.

When corporations block data access to the extent of having a data monopoly, it allows them to engage in similar harmful acts as traditional monopolies.⁸ For example, existing ridesharing platforms benefit immensely from a data feedback loop within individual platforms that attracts riders and drivers that creates a barrier to entry for new ridesharing companies. The fact that data access alone stymies competition raises the question of why the corporation, not the user, controls the user's data.

Some jurisdictions have recognized the dubiousness of corporations claiming exclusive rights to consumer data. Recently, the city of Seattle ordered Lyft to share data from its ridesharing services via public records ordinances.⁹ One of the reasons highlighted by an assistant city attorney is that taxpayers pay for the roads on which ridesharing applications operate. Likewise, consumer data held by private companies is dependent on public use and public infrastructure. The public's interest in understanding its own data to avoid harms from corporate actors is paramount and researchers can make mass information palatable by providing context and analysis. However, that analysis is only possible if the public can share its data with researchers, which is unlikely to occur as long as consumer data is proprietary to corporations.

B. Corporations Make It Difficult for Researchers to Audit the Data They Collect

Currently, access to data is difficult for consumers and researchers alike. When we audited Shipt, our goal was to help gig workers understand the impact of Shipt's data-driven payment model despite not having easy access to their own data. When we created the Shipt Calculator, we had to consciously avoid obstacles to data collection, including potential legal risks—which led us to eschew data subject access requests under state law that lacked any guarantee that workers would not face reprisals.

Corporations may also refuse to provide enough information for researchers to adequately assess privacy risks. For example, when Google proposed FLoC, they did open the proposal to public comment—but they did not provide the tools necessary to properly inspect the proposal and evaluate its impact on consumers. Only by leveraging third party comScore data purchased by Harvard's Data Privacy Lab were we able to identify FLoC's inadequate protection of identity and privacy. But researchers do not always have access to third-party data, or the money or connections required to access it. If we are limited to what a corporation voluntarily makes available, we are unlikely to be able to fully assess the risks to consumers of novel technologies or practices.

This is consistent with the FTC's own finding that, although researchers have *some*

⁸ Robert Mahari, Sandro Claudio Lera & Alex Pentland, *Time for a New Antitrust Era: Refocusing Antitrust Law to Invigorate Competition in the 21st Century*, 1 STANFORD COMPUTATIONAL ANTITRUST 53 (2021), <https://ssrn.com/abstract=3943548>.

⁹ Dana Calacci, Alex Berke, Kent Larson & Alex (Sandy) Pentland, *The Tradeoff Between the Utility and Risk of Location Data and Implications for Public Good*, OXFORD & LONDON SCHOOL OF ECONOMICS CONNECTED LIFE CONFERENCE (2019), <https://arxiv.org/abs/1905.09350>.

access to this data, their access is limited in problematic ways. For example, in the 2022 report *Combatting Online Harms Through Innovation*, the Commission explained: “To be clear, it is not that no platforms provide any access to researchers. The issue is that they generally do not provide nearly enough, access is often conditioned on non-disclosure agreements.”¹⁰

IV. Conclusion

Consumers cannot avoid mass data collection, as they need to transact with platforms that collect their data in order to operate in society. This has led to a power asymmetry between major corporate players and American consumers.

The FTC should engage in rulemaking that facilitates consumers’ ability to consent to share their data with researchers. Data sharing benefits consumers and the public at large by increasing corporate transparency and reducing power asymmetries. This data portability focused solution gives more agency to consumers and enables a less intrusive audit mechanism for corporations.

Respectfully Submitted,

Alex Berke, Dana Calacci, and Robert Mahari

MIT PhD Students and Data Researchers

¹⁰ COMBATTING ONLINE HARMS THROUGH INNOVATION, *supra* note 4, at 54.