INSATIABLE:
The Tech Industry’s Quest for All Our Data

A REPORT FROM
FREE PRESS ACTION
# TABLE OF CONTENTS

- Executive Summary
- Introduction: The Exploitation of Our Data
- When Privacy Is Not the Policy: Privacy Policies as Data-Collection Policies
- Case Studies: The Race to Replace Twitter
  - Threads
  - Mastodon
  - Bluesky
- Zooming Out: The Broader Data-Collection Ecosystem
- The Need for Policy Reforms
- Conclusion
- Methodology & Acknowledgements
- Endnotes

October 2023
EXECUTIVE SUMMARY

The digital economy is built on the extraction and exploitation of our personal data for profit. This rampant harvesting of our data has facilitated discrimination, disclosure of sensitive information and other abuses. As digital services become increasingly integrated into critical public and private infrastructures, these largely unregulated data-collection practices come with a growing cost.

By analyzing how privacy policies actually function as data-collection policies, this report examines the need to move beyond industry self-regulation. We discuss the privacy policies of three key platforms in the race to replace Twitter (X): Meta’s Threads, Bluesky and Mastodon. These three case studies highlight varied approaches to data collection and retention by platforms seeking to offer similar user experiences. Free Press Action finds that the persistent, under-regulated and profit-driven harvesting of our personal data undermines the potential for greater user control on emerging platforms.

This report describes how often privacy policies fail to detail exactly what companies do with our personal data. In practice, companies use privacy policies not to protect users, but to expand their collection of personal data — or use already-collected information in new ways. Even if every user diligently reads through and regularly reviews each platform’s privacy policies for updates, we would still not have a full picture of how our data are collected and used by platforms, let alone by the myriad other companies whose websites and mobile apps we interact with on a regular basis.

In late 2022, it seemed like we might finally get comprehensive federal privacy regulations in the United States: The Federal Trade Commission was moving forward with a proceeding on commercial surveillance and the House of Representatives passed the American Data Privacy and Protection Act out of the Energy and Commerce Committee on a broad bipartisan basis. Yet at the time of this report’s publication, FTC and congressional efforts to enact comprehensive new privacy protections have not yet come to fruition, while the unchecked collection of our personal data continues.

Our analysis illustrates the urgent need for regulatory and lawmakers action to safeguard our digital privacy and combat corporate hunger for our data.
I. INTRODUCTION: THE EXPLOITATION OF OUR DATA

Using a smartphone. Opening a web browser or mobile app. Browsing an online store. Walking into a brick-and-mortar shop. These are all activities that generate personal data about our identities, locations, preferences and more. And these are all data that a slew of tech, retail, insurance and marketing companies are eager to monetize and use to make judgments about who we are and how and whether they provide goods and services to us.

The rampant harvesting of our personal data has led to discrimination,¹ disclosure of sensitive information,² and other abuses.³ Companies have used it to exclude specific users from the same information and opportunities as others.⁴

Banks, insurance companies, potential employers, government agencies and other institutions with power over our daily lives both collect data directly and also purchase slices of our online profiles⁵ to make judgments about whether we qualify for a loan, a job, housing, patient health care, or affordable insurance coverage.

These discriminatory practices disproportionately harm⁶ highly surveilled communities that have less access to essential services. For example, data can reveal users’ reproductive health-care decisions. The criminalization of reproductive health care in our post-Roe v. Wade landscape already has a disparate impact on women and other birthing people; extensive data collection is just another tool to further restrict autonomy. In states where abortion care is criminalized, law enforcement or even private individuals can use personal data to target people seeking or providing abortion care.

Digital platforms are just a handful of the many kinds of companies hungry to gather up information about us for their own profit. Search engines, retailers, web browsers, mobile apps and others all also exploit our personal data for profit. But platform companies’ data-collection practices are notably far-reaching for one simple reason: A significant majority⁷ of people in the United States use at least one social-media platform — such as Facebook, TikTok, Twitter (now “X”), and YouTube. What’s more, a majority of many platforms’ users visit these sites daily.⁸ When we use these services, we agree to terms of service and privacy policies predicated on gathering troves of information about us. And even when we’re not using these platforms — even if we’ve never used these platforms — many of these companies are still collecting⁹ data on us.
It may feel innocuous when we surrender our data for access to online services. In fact, it’s often designed to feel that way. But as digital services become increasingly integrated into critical public and private infrastructures, these largely unregulated data-collection practices come with a growing cost.

II. WHEN PRIVACY IS NOT THE POLICY: PRIVACY POLICIES AS DATA-COLLECTION POLICIES

What personal information are companies gathering about us, and how are they using it? In theory, answers to these questions can be found in privacy policies — lengthy, often vaguely worded documents that most people do not read. These privacy policies would more accurately be described as data-collection policies. (In fact, Meta only recently rebranded its “data-policy” as a “privacy policy”). Rather than articulating meaningful constraints on companies’ collection, storage, use and even sale of our personal data, privacy policies describe the scope of what companies collect about us and the open-endedness of their ability to store and use that data.

When companies want to collect more data or use it in new ways, they simply update these policies — as we’ve seen in 2023 with the flurry of tech companies updating their terms of service and/or privacy policies to give themselves explicit permission to train generative AI models on users’ data.

As a result, it’s become a Sisyphean task to stay up to date on the data-collection policies of every company or service we interact with. This is especially daunting given that users have no power to negotiate the terms of these policies, even when we have concerns. Even in states where comprehensive data-privacy laws have been enacted (as of October 2023, there are 12 on the books), the burden largely falls on users to opt out of specific data-collection practices. More often than not, we are incentivized to simply scroll down to “accept” the terms and proceed with our online activity.

It’s not that people don’t care about who is collecting our information and how they’re using it — we do. According to a Pew Research study from 2019, most U.S. adults have concerns about how companies use their data. Most similarly lack confidence that companies’ privacy policies adequately protect users’ personal data. Yet the majority of people also feel as if users have little-to-no control over any of this.

We shouldn’t have to track down every voluminous and confusing data-collection and -use policy, try to make sense of it, and then opt out of every unfair, invasive practice across every service we use — particularly when we have few meaningful avenues to mitigate these companies’ persistent collection of our personal data. And yet in the absence of comprehensive federal privacy laws and regulations, the reality is just that. So what data are we really giving away when we log into social-media platforms? And do all of these companies have the same privacy policies?
III. CASE STUDIES: THE RACE TO REPLACE TWITTER

When Elon Musk acquired Twitter in October 2022, the company already had a troubling privacy track record. In May 2022, the Federal Trade Commission and Department of Justice fined the company $150 million for deceptively collecting personal data to use for targeted advertising, violating a prior FTC order.

Musk’s acquisition of the platform and gutting of its workforce created new vulnerabilities for users’ personal data. Among the people Musk immediately fired was Twitter’s head of legal, public policy, and trust and safety. Within a week, Musk had reduced the company’s workforce by almost half, heavily impacting trust and safety teams, with no clear transition measures in place.

As Twitter users debated migrating to new platforms, data privacy and security were key concerns.

Below, Free Press Action illustrates the online privacy landscape by breaking down the privacy policies and data-collection practices of leading social-media platforms in the race to replace Twitter.

From the befuddling legalese to the spontaneous updating of corporate privacy policies, users are too often left holding the bag, with no insights into how much data is being collected about them — or what companies are doing once they gather this information. Questions abound, with no consistent answers for any given platform for any period of time.

We begin with the latest major platform on the scene: Threads. Threads occupies a peculiar position in the social-media landscape: It’s the newest entrant in the race to replace Twitter, yet its parent company — Meta — is one of the most powerful corporate actors. We then discuss how Threads’ policies compare to those of Bluesky and Mastodon, which are similarly trying to lure users away from Twitter with appeals about interoperability, decentralization, and greater user control. Finally, we examine how these platforms’ privacy approaches compare to the rebranded Twitter, as well as to other companies in the platform ecosystem.

For these case studies, we refer both to the companies’ privacy policies as well as the companies’ disclosures to the Apple App Store regarding the categories of data their mobile applications collect.

Free Press Action finds that the largely unregulated harvesting of our personal data for profit overshadows the potential for greater user control over our data on some emerging platforms, like Bluesky and Mastodon. Our analysis shows the urgent need for lawmaking and regulatory action to safeguard our digital privacy and mitigate corporate hunger for our data.
 Threads: Meta’s newest vehicle for data collection

In July 2023, Meta’s Threads became the latest entrant in the race to replace a faltering Twitter. It joined Facebook, Instagram, and WhatsApp in Meta’s suite of platforms. Meta has long led efforts to chip away at users’ expectations of online privacy. While its CEO and co-founder Mark Zuckerberg has been apologizing for data-privacy violations and promising to do better for nearly two decades, monetizing widespread on- and off-platform collection of internet users’ data is central to Meta’s business model. As a result, Meta is facing historic regulatory enforcement actions in both the United States and the EU as regulators monitor ongoing privacy violations at the company’s flagship products:

- The FTC has proposed changes to a privacy order that the agency issued against Meta in 2020 — paired with a historic $5-billion penalty — alleging that the company violated the existing order. Among the proposed changes: Meta would be “prohibited from releasing new or modified products, services, or features without written confirmation [from an independent assessor] that its privacy program is in full compliance with the order’s requirements.” The 2020 privacy order was the result of a settlement over charges that Meta violated a separate FTC order from 2012 by “deceiving users about their ability to control the privacy of their personal information.”
- Meta was fined €1.2 billion ($1.26 billion) for violating EU data-protection rules. Ireland’s Data Protection Commission found that Meta failed, once again, to comply with an existing privacy order — in this case, a 2020 ruling by the European Court of Justice. Both of these enforcement actions were announced in May 2023. Less than two months later, Meta launched Threads.

Meta introduced Threads as “a new app, built by the Instagram team” — a team already responsible for managing a product with reportedly 2-billion monthly active users, according to Meta’s own numbers. By directly linking this new platform to Instagram (users can’t delete their Threads accounts unless they’re willing to sacrifice their Instagram accounts too), it’s clear that Meta hopes to capitalize on Instagram’s substantial global user base — a plan that, in the app’s early weeks, seemed successful.

Meta also announced plans to make Threads interoperable with other social networks using the decentralized social-networking protocol ActivityPub.

Meta executives’ early statements to the press focused on the new platform’s rapid accumulation of users, making headlines with a record-breaking 100-million sign-ups in its first week. Threads launched just months after Meta announced its latest round of mass layoffs as well as a hiring freeze, raising additional concerns about the company’s capacity to implement policies concerning the governance, moderation, and security of an entirely new platform.
According to Reuters, these layoffs directly impacted Meta’s privacy and integrity teams. Free Press and two dozen civil-rights groups wrote to Meta requesting details on how Threads would be moderated, how its content moderation and privacy policies would be distinct from those of other Meta products, and how the platform planned to enforce policies with transparency. We received a lackluster response from Meta over a month later than requested that failed to address or concretely answer any of our questions.

How do Threads’ privacy promises stack up on paper?

As is the case with Instagram, Threads’ policies direct users to the policies Meta developed for Facebook, including its privacy policy. There is one Threads-specific “Supplemental Privacy Policy” that relates primarily to Meta’s plans for the platform’s interoperability. For now, Threads’ data-collection practices appear to take the same approach as other Meta platforms, including allowances for collection and use of:

- **Sensitive data** like race, sexual orientation, pregnancy status and religion, as well as health and fitness data;
- **Geolocation data**, even if you’ve disabled location services;
- **Off-platform data**, including websites you visit and mobile apps you use; and
- **Information from third parties** “using or integrating” Meta products for marketing, even from users who do not have accounts on any Meta platform.

Because Threads, in its early phase of growth, is tied so closely to Instagram users, Meta likely already has a substantial amount of data on the first wave of Threads adopters. But Threads has the potential to expand Meta’s data-collection practices in two key ways. The first is simply collecting more data points on existing Instagram users. (Though it hasn’t announced plans to do so, Meta could also eventually decouple Threads from Instagram, allowing people to sign up for Threads independently and attracting new users who may not already have accounts on other Meta platforms.)

Second, Meta’s plans to make Threads interoperable with other platforms poses unique privacy concerns for those who interact with Threads content and users through accounts on other platforms that use the ActivityPub protocol, like Mastodon. As detailed in the Threads’ “Supplemental Privacy Policy,” Meta will collect data from people using other platforms if they interact with Threads users or content.

Extending its data collection off-platform is nothing new for Meta. According to research by The Markup, Meta Pixel — a data-tracking and analytics tool — is present on “more than 30 percent of popular websites.” This tool collects and sends users’ data directly back to Meta whether or not people visiting those websites have an account with Facebook or any other Meta platform. The Markup also notes that such data includes sensitive financial and health information.
OTHER COMPETITORS IN THE RACE TO REPLACE TWITTER: MASTODON AND BLUESKY

Mastodon

Mastodon entered the scene in 2016 — far earlier than competitors like Threads or Bluesky — but was similarly envisioned as a competitor to Twitter. Mastodon’s founder sought to create an alternative to centralized, for-profit platforms while still facilitating a Twitter-like microblogging experience. Mastodon saw a massive influx of users following Musk’s acquisition of Twitter in late 2022. It is a functionally and aesthetically similar platform but has a distinct technical architecture: It’s a network of independently run “instances,” or servers, that can each set their own membership, privacy, and content-moderation policies and practices.

Mastodon’s mobile application and original mastodon.social server — the federated social network’s largest instance as well as its default server for new sign-ups — take a notably privacy-protective approach to users’ data, with some clear limits on data collection and retention. However, Mastodon’s federated model means that individual servers might have privacy policies and practices that are less secure and/or privacy-conscious. Nonetheless, Mastodon’s privacy policy sets a critical baseline:

- Mastodon does not collect sensitive data such as information about users’ gender, race or ethnicity; sexual orientation; pregnancy status; or religion.
- Mastodon does not collect geolocation data, according to its privacy policy and app-store disclosures.
- Mastodon does not collect off-platform data beyond the name of your browser application and the IP address of your device.
- Mastodon uses cookies by default but respects “Do Not Track” signals from web browsers.
- Mastodon’s direct-messaging feature is not end-to-end encrypted. However, when you send a direct message on Mastodon, the platform reminds you of this potential privacy concern and warns you not to share sensitive information.

Bluesky

Twitter co-founder Jack Dorsey’s newest venture, Bluesky, has yet to officially “launch”: It’s still in invite-only beta despite reportedly passing the one-million-user milestone in mid-September 2023. Although Bluesky aspires to a decentralized governance model similar to Mastodon, the only server available to users as of October 2023 is Bluesky Social, which the Bluesky team operates and governs.

Bluesky’s privacy disclosures to the Apple App Store suggest that it’s taken a relatively minimalist approach to user data collection that focuses on basic platform functionality. But the terms of Bluesky’s privacy policy are less definitive. In several places, the policy employs ambiguous language that is common within privacy policies and makes it difficult for users to know with certainty how the company is collecting and using their personal data.
For example:

- Bluesky’s App Store disclosures suggest that the app does not collect sensitive data such as information about users’ gender, race or ethnicity; sexual orientation; pregnancy status; or religion. But the platform’s privacy policy says it may use “personal information to create de-identified and/or aggregated information, such as demographic information,” with no additional clarity.
- According to Bluesky’s privacy policy, it does not collect “precise location-based information” without consent. It’s unclear whether Bluesky collects more general geolocation data.
- Bluesky collects off-platform data. “Such as pages that you visit before, during and after” using Bluesky. It also specifically states that it does not respect “Do Not Track” signals from web browsers.
- Bluesky does not have a direct-messaging feature at the time of this report’s publication.

Because the platform is still in beta, it remains to be seen how Bluesky’s approach to data privacy might change between now and an official launch as more features are added, as its business model evolves,

Department of Justice court filings indicate that mass layoffs and resignations as well as specific directives from Musk may have violated both the company’s own policies and its settlement agreement with the FTC concerning Twitter’s misuse of users’ personal data. At the time of this report’s publication, the FTC’s investigation remains ongoing.

Like other legacy social-media platforms, Twitter has long collected significant amounts of information on its users. While Twitter’s privacy policy and app-store disclosures do not indicate that the platform collects sensitive information like users’ race, gender or sexual orientation, it does state that the companies’ “ad and business partners” share information such as users’ “demographic or interest data.” There is no additional detail about what kinds of personal information “demographic or interest data” might encompass. In a separate document, Twitter clarifies that it infers information “such as interests, age, and gender” about user accounts.

Meanwhile, updates to the site’s privacy policy have broadened the existing scope of the platform’s data-collection and use practices, and indicate the company’s interest in collecting even more sensitive data from users in the future. Besides explicitly declaring that the company may use the information it collects to train its “machine learning or artificial intelligence models,” the updated privacy policy references the potential collection and use of users’ biometric data. This is a particularly sensitive category of personal data that encompasses facial recognition, fingerprint mapping, and retina scans.

IV. ZOOMING OUT: THE BROADER DATA-COLLECTION ECOSYSTEM

Back at Twitter, federal regulators have raised new red flags about the company’s data-privacy and security practices in the wake of Musk’s acquisition.
The new privacy policy went into effect on Sept. 29, 2023.

Twitter is not the only company that has modified its privacy policy and/or terms of service to specifically clarify that it will collect user data to train its machine learning and/or artificial intelligence models in 2023. Zoom received significant backlash after making a similar update to its terms of service, forcing the company to further update its policies to clarify that it will not use audio, video, or chat customer content to train [Zoom’s] artificial intelligence models without [users’] consent.” When Meta launched its new suite of generative AI features on Sept. 27, 2023, it simultaneously disclosed that Facebook and Instagram users’ public text and photo posts were used to train the underlying generative AI models. Twitter is also not the only company to update its privacy policy to allow for the collection of biometric data: TikTok made a similar change in 2021, allowing itself to collect “biometric identifiers...such as faceprints and voiceprints, from your User Content.” Companies’ ability to unilaterally decide that they will collect more information from users — or use already-collected information in new ways — reflects a broader landscape of corporate data abuses and over-collection practices.

As illustrated in the above case studies, privacy policies also often fail to detail exactly what information is collected even when describing data-collection practices. These policies frequently employ open-ended language like “such as” or “including” that leaves the door open for collecting data beyond what is directly named.

Even if every user diligently reads through and regularly reviews each platform’s privacy policies for updates, we would still not have a full picture of how our data are collected and used by platforms, let alone by the myriad other companies whose websites and mobile apps we interact with on a regular basis.

Similarly, when privacy policies reference third-party access, they rarely disclose who those third parties are—even while referring users to those companies’ own privacy policies for information about how they will use personal data shared with them. Bluesky’s disclosure of its analytics partners is an exception to the far more common practice (on display elsewhere in Bluesky’s privacy policy) of omitting the names of such vendors.

When companies violate their own voluntary commitments to limit their collection or retention of our personal data, users have little insight or recourse unless companies disclose the violation — or researchers and journalists do. For example, one week after the Supreme Court overturned Roe v. Wade in June 2022, Google announced that it would automatically delete location-history data if Google systems detected that a user visited a “particularly personal” medical facility, including abortion clinics and domestic-violence shelters. Less than a year later, reporters and advocates began raising the alarm that Google was failing to follow through on that commitment.

Improving privacy policies and practices one company at a time cannot solve the broader problem.
V. THE NEED FOR POLICY REFORMS

There is an urgent need for regulation to provide comprehensive data-privacy protections; to hold tech giants, data brokers and all who facilitate the data marketplace accountable; and to protect users’ privacy and civil rights. Only with strong and enforceable laws and regulations can we safeguard our privacy and ensure responsible platform governance in the digital age. These regulations must prioritize:

1. **Data minimization**: narrowing the permissible scope for the collection, retention, use and sale of personal data, allowing only what is necessary and proportionate to provide or maintain the specific product or service that a user requests;

2. **Transparency**: enabling us to see whether companies are complying with their own policies (and, in the future, regulatory requirements);

3. **Digital civil rights**: protecting our rights so our personal data aren’t used to discriminate against or disadvantage us on the basis of protected characteristics;

4. **Data control**: giving people easy and clear choices on how their data may be collected and used, as well as the ability to delete previously collected data; and

5. **Private right of action**: letting people go to court when their civil rights and privacy rights are violated.

The United States has no comprehensive federal law restricting companies from continuing the extractive practices by which they gather consumer data, retain it without proper security mechanisms, and sell troves of it to government agencies, data brokers and other actors without our knowledge or meaningful consent. Existing data-privacy laws like the Health Insurance Portability and Accountability Act (HIPAA) are critical but narrow — affecting only certain types of data processed by specific covered entities. In the absence of comprehensive federal action, some states have entered the fray with a variety of proposals. However, as Free Press Action’s Amanda Beckham writes: “Data protection is the right of every individual, and that level of protection shouldn’t depend on the state someone lives in.”

In 2022, it finally seemed like comprehensive federal privacy regulations might be on the horizon. In the House of Representatives, the American Data Privacy and Protection Act (ADPPA) passed out of committee by an overwhelming bipartisan margin, but has since faced political obstacles. ADPPA includes data minimization and transparency requirements and also consumer rights to access, correct and delete data held by a covered entity. ADPPA would prohibit discriminatory collection, use or transfer of personal data and create a strong framework empowering federal regulatory agencies like the FTC to enforce these digital civil rights. The bill would also give consumers the right to object before companies covered by ADPPA transfer personal data to a third party. At the time of this report’s publication, ADPPA had not been reintroduced in the 118th Congress.
VI. CONCLUSION

Companies have been allowed to self-regulate their data-collection practices for more than two decades now — and they’ve failed to protect user privacy and civil rights. In fact, many of their business models are built on exploiting our demographic and behavioral data for profit. New platforms like Bluesky and Mastodon may help create space for less data-extractive alternatives to gain traction. Yet Meta’s efforts to introduce Threads into the same ecosystem underscore the reach and influence of giant incumbents’ data-collection practices even as they, too, begin to pay lip service to the principles of decentralization, interoperability and user control.

Comprehensive federal data-privacy rules are urgently needed to prevent the harms that flow from the unmitigated harvesting of our personal data. The case studies examined in this report represent only one slice of a data economy that is inadequately unregulated. Without congressional and regulatory action, the data mining of our interests, identities, behaviors and social networks will remain a lucrative business model — and continue unchecked.

Civil-rights and consumer-advocacy groups, including Free Press Action, have advocated that these rules prioritize data minimization and security while also defining blatantly discriminatory data practices as unfair and therefore unlawful acts. At the time of this report’s publication, the FTC has not progressed beyond the public-comment period of its rulemaking process.
VII. METHODOLOGY & ACKNOWLEDGEMENTS

To develop this research report, Free Press Action reviewed the privacy policies and Apple App Store disclosures for Bluesky, Mastodon, Threads and Twitter (X). The report relies on the disclosures and policies that were publicly available and in effect through September 2023.

Free Press Action wrote and researched this report. The report was led by Jenna Ruddock with significant research support and editing from Florín Nájera-Uresti and Nora Benavidez, both of Free Press Action. The development and production of this report could not have happened without support from Free Press Action colleagues Craig Aaron, Jessica J. González, Amanda Beckham, Timothy Karr and Matt Wood. This report was edited by Amy Kroin and designed by Imani Oakley, Sara Pritt and Dutch Cosmian.
ENDNOTES


8 Id.


16 See, e.g., Ari Ezra Waldman, Industry Unbound, p. 52, 2021; Julie Cohen, Between Truth and Power, p. 56, 2019


20 See Brooke Auxier et al., supra note 14.

21 Id.


35 See Meta supra note 33.

36 See Peters and Porter, supra note 23.


41 Id.

42 Id.


53 See Lee, supra note 49.


59 See Hays, supra note 18.


