An Empirical Inquiry into Surveillance Capitalism: Web Tracking

Nils Bonfils
nils.bonfils@mail.utoronto.ca
University of Toronto
Toronto, Ontario, Canada

Abstract

The modern web is increasingly characterized by the pervasiveness of Surveillance Capitalism. This investigation employs an empirical approach to examine this phenomenon through the web tracking practices of major tech companies — specifically Google, Apple, Facebook, Amazon, and Microsoft (GAFAM) — and their relation to financial performance indicators. Using longitudinal data from WhoTracks.Me spanning from 2017 to 2025 and publicly accessible SEC filings, this paper analyzes patterns and trends in web tracking data to establish empirical evidence of Surveillance Capitalism's extraction mechanisms. Our findings reveal Google's omnipresent position on the web, a three-tier stratification among GAFAM companies in the surveillance space, and evidence suggesting an evolution of tracking techniques to evade detection. The investigation further discusses the social and environmental costs of web tracking and how alternative technologies, such as the Gemini protocol, offer pathways to challenge the extractive logic of this new economic order. By closely examining surveillance activities, this research contributes to an ongoing effort to better understand the current state and future trajectory of Surveillance Capitalism.

Keywords

Surveillance Capitalism, Web Tracking, Privacy, Empirical Study

ACM Reference Format:
Nils Bonfils. 2025. An Empirical Inquiry into Surveillance Capitalism: Web Tracking. In *Proceedings of 11th Workshop on Computing Within Limits (LIMITS '25)*. 10 pages. https://doi.org/10.48550/arXiv.2508.07454

1 Introduction

Contemporary society is increasingly characterized by the pervasiveness of Surveillance Capitalism, a new economic order where human experience is claimed as free raw material for hidden commercial practices of extraction, prediction, and sales. While fundamentally reshaping digital interactions, this phenomenon remains conceptually elusive and difficult to grasp in concrete terms. This elusiveness calls for empirical investigation to understand its actual mechanisms and impact.

Surveillance Capitalism originated at Google, when the corporation discovered that users' seemingly inconsequential digital traces, also called "data exhaust", could be algorithmically processed and analyzed to predict future behavior and monetized through targeted advertising [19]. The internet, particularly the World Wide Web, provided the essential infrastructure for this new economic logic to flourish. The web's client-server architecture, coupled with lagging regulatory frameworks, created optimal conditions for surveillance operations to take root without restraints.

The growing trends towards the use of web applications over more traditional "native" applications has accelerated this transformation. With individuals increasingly migrating daily activities to web applications, the web has evolved from a document-sharing medium to an application platform. This consolidation of activity on the web provides an ideal environment for surveillance operations, enabling tech companies to monitor user behavior across various services through web tracking. This surveillance power is further concentrated as these same companies provide access to the web platform through their browsers, with Google, Apple, and Microsoft's web browsers representing approximately 90% of all browser usage.

Web tracking refers to the collection of data about users' online activities, including pages visited, clicks made, time spent on sites, and numerous other behavioral signals. These collection mechanisms range from simple cookies to sophisticated fingerprinting techniques capable of identifying users across multiple devices and sessions. While some tracking serves legitimate purposes like authentication and site functionality, the vast majority enables the surveillance apparatus that powers contemporary Surveillance Capitalism. Both direct outcomes (targeted advertising) and indirect outcomes (product improvement) of this surveillance enable the creation of "behavioral prediction products" that have tangible monetary value.

Web tracking data offers researchers a valuable opportunity to investigate the materialization of Surveillance Capitalism. By analyzing patterns and evolution in tracking technologies, we can begin to map the actual extent and mechanisms of surveillance on the web, moving beyond theoretical frameworks to empirical evidence. This methodological approach allows for critical assessment of surveillance practices across digital ecosystems and their relationship to corporate financial performance.

This investigation will seek to establish empirical evidence of Surveillance Capitalism by analyzing web tracking practices of major tech companies — specifically Google, Apple, Facebook, Amazon, and Microsoft (GAFAM) — and their relation to financial performance indicators. Three primary research questions will be addressed:

- **RQ1**: Within GAFAM, which entities demonstrate the highest engagement in surveillance practices?
- **RQ2**: Do web tracking metrics correlate with the entity's advertising revenue?
- **RQ3**: Can web tracking provide quantitative evidence of Surveillance Capitalism activity?

2 Background

This work builds on the theoretical framework of Surveillance Capitalism established by Zuboff [19, 20]. Most contributions based on this framework are predominantly qualitative in nature [9, 15].

LIMITS '25, Online
2025. https://doi.org/10.48550/arXiv.2508.07454

Few empirical studies directly reference Surveillance Capitalism [1].

The dearth in empirical studies can be attributed to the active concealment of activities and practices by corporations partaking in Surveillance Capitalism. Entities that surveil also actively conceal their data collection practices to maintain this asymmetrical access to information that is essential for Surveillance Capitalism to function. In Zuboff's seminal book, she defines Surveillance Capitalism as an economic order, an economic logic, a mutation of capitalism, a foundational framework, a threat, the origin of a new power, a movement and an expropriation of human rights. [20] Despite this rich definition, such a concept remains difficult to operationalize, making it challenging to identify empirically. Additional work is needed on defining Surveillance Capitalism, but more importantly on identifying ways to evaluate and measure its different aspects. However, empirical researchers investigating this topic implicitly place themselves in opposition to the dominant entities of this new economic order.

Nonetheless, there has been a wealth of empirical research examining web trackers, ad blockers, and privacy enhancing technologies outside of the Surveillance Capitalism theoretical framework. Previous research on web tracking has focused on specific countries [2], continents [5], or regulatory landscapes [6]. Some studies adopt a more global perspective [17]. However, most of those studies only provide static snapshots of the web tracking landscape. Few studies conduct longitudinal analyses [10, 11] and fewer systematically link trackers to their parent corporate entities. This investigation uses data from a long-standing open source database that allows investigation of web-tracking over multiple years [7].

This study presents a significant opportunity for interdisciplinary research that bridges empirical research of privacy-enhancing technologies with the theoretical framework of Surveillance Capitalism.

3 Approach

This investigation employs an empirical approach to examine the role of GAFAM companies in the web tracking space. Two primary data sources were used:

(1) WhoTracks.Me: an open source database of web trackers.
(2) GAFAM's publicly accessible U.S. Securities and Exchange Commission (SEC) filings.

Both datasets were integrated to examine the interplay between tracking practices and financial performance.

3.1 WhoTracks.Me Dataset

WhoTracks.Me is a website and database that document web tracking activities [7]. It originated within the Cliqz company in 2017 after the acquisition of the Ghostery web browser extension[1]. While Cliqz, a company focused on creating both a privacy-oriented web browser and search engine, discontinued operations, the work on Ghostery extension and associated database continues to receive monthly updates.

Data collection occurs through "real" web browsing by Ghostery users who have consented to share their tracking data during their browsing. This dataset is self-described as "the largest and longest measurement of online tracking" with data dating back to 2017. This gives a unique opportunity to study the evolution of web tracking over time.[2]

WhoTracks.Me data is a publicly accessible AWS S3 bucket. As of April 2025, there are approximately 29 gigabytes of tracker data available. The data is structured hierarchically. At the top, directories represent each month of data collected. The monthly directory is further divided by region (e.g. "Global", "US", "EU"). Each region corresponds to the place the data was collected from which enables comparative analysis across regulatory environments. Given the transnational nature of Surveillance Capitalism, this research will focus on the "Global" region.

Within the regions' folders, there are five comma-separated value (CSV) files. Each file aggregates data around a specific entity in the web tracking ecosystem. This investigation will focus on the "companies.csv" dataset.[3] This dataset aggregates web tracking data about the top companies and Table 1 describes the relevant variables within that dataset.

Karaj et al. describes how they operationalized the capture and aggregation of trackers' data [7]. There are three important concepts necessary to understand what this data represents.

First, a **page load** is defined by:

- Creation in the main web request of a tab when entering a URL in a web browser's URL bar.
- Ending when the tab is closed or another main web request is observed for the same tab, usually occurring when a link is clicked and the browser loads a new page.

Second, **third-party requests** are the building blocks of a tracker [18]. During a page load, subsequent requests to a URL on a different domain than the current loaded page are counted as a third-party request.

Third, **aggregation** is possible with Ghostery's trackerdb[4], a manually curated database mapping domain names to the companies they operate under. Using trackerdb's information, an aggregation of data by companies is possible by linking third-party requests to specific domains and their parent company.

3.2 SEC Filings

Publicly traded companies in the United States have an obligation to comply with the Securities and Exchange Commission (SEC). This requires them to submit standardized forms and reports which are then published through the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. As this study focuses on GAFAM, all publicly traded companies' financial data is publicly available. Quarterly and annual earnings reports, published as part of the 10-Q and 10-K forms, provide us with insight into the companies' finances and their evolutions.

Fundamental financial metrics were systematically extracted from these quarterly reports. Our analysis is based on four core financial metrics, from which two additional measures were derived, as shown in Table 2.

[1] https://www.ghostery.com/ghostery-ad-blocker

[2] https://github.com/whotracksme/whotracks.me

[3] For an explanation of the other datasets available, the reader is encouraged to consult the database's documentation directly: https://github.com/whotracksme/whotracks.me/blob/master/whotracksme/data/Readme.md#datasets

[4] https://github.com/ghostery/trackerdb

Table 1: Description of the relevant variables in the "companies.csv" dataset

Variable	Description	Possible values
reach	Proportional presence across all page loads (i.e. if a tracker is present on 50 out of 1000 page loads, the reach would be 0.05).	Floating point between 0 and 1
site_reach	Presence across unique first party sites. e.g. if a tracker is present on 10 sites, and there are 100 different sites in the database, the site reach will be 0.1. **Important note**: In February 2019, this measure was redefined to the number of sites in the top 10,000 which have this tracker on more than 1% of page loads. To stay consistent with the previous definition of that measurement, that value is divided by 10,000.	Floating point between 0 and 1
trackers	Average number of trackers present on the sites that uses at least one of this company's trackers.	Positive floating point
content_length	Average *Content-Length* HTTP headers received by third-party requests to trackers' domains owned by this company during a page load. It is meant to be an approximate measure of the bandwidth usage of trackers. Expressed in kilobytes (KB). **Important note**: The distribution of this variable can have a fat tail due to audio or visual content sometimes served by third-party tracker requests.	Positive floating point
requests	Average number of third-party requests made to this company's tracker per page load.	Positive floating point
requests_tracking	Average number of third-party requests that contains potentially identifying information (cookie or query string) made to this company's tracker per page load.	Positive floating point

Table 2: The financial metrics collected and derived from 10-Q and 10-K SEC filings

Financial Metrics	Description
Gross Revenue	The total amount of money earned in a quarter (in millions).
Advertising Services Revenue	The amount of money earned by the company through its advertising-related services (in millions).
Total Expenses	The total amount of money spent during a particular quarter (in millions).
Sales and Marketing Expenses	The amount of money spent in sales and marketing (in millions).
Share of Advertising Revenue	The proportion of the total revenue coming from advertising services (in percent).
Share of Marketing Expenses	The proportion of the total expenses dedicated to sales and marketing (in percent).

3.3 Visual Analysis

To grasp the GAFAM's role within the space of web tracking and their role in Surveillance Capitalism, visual analysis techniques were used to highlight patterns and trends. The analysis examined the progression of the tracking and financial data from May 2017 to March 2025.

This study used Python and Jupyter notebooks for data aggregation and visualization processing. Two additional Python libraries for data manipulation and representation were used: Pandas and Matplotlib. All code used in this study will be released in the public domain to facilitate replication and extension of this work (*see* Appendix A).

4 Findings

Our analysis reveals significant patterns regarding GAFAM tracking practices. To contextualize the overall web tracking landscape, several non-GAFAM companies demonstrate substantial reach or site reach when averaged across the full temporal range (see Table 3). Two of them, Twitter/X and Kaspersky Lab, are not publicly traded but have an estimated quarterly gross revenue of approximately 0.25B USD. Cloudflare and ComScore are publicly traded, and their latest quarterly gross revenues (Q4 2024) are 0.46B USD and 0.90B USD. These figures are eclipsed by the earnings of the GAFAM, with Facebook having the lowest revenue among GAFAM at 48.38B USD for Q4 2024. 46.78B USD of the 48.38B USD (96.69%) came from advertising alone. The five companies examined in this study were chosen because these companies' yearly revenues rival the GDP of medium-sized nations, such as Finland, Greece, and Portugal[5] as shown in Table 4.

Table 3: Top 7 companies based on average reach and site reach over the full time period

Rank	Company	Reach	Company	Site Reach
1	**Google**	**77.86%**	**Google**	**98.02%**
2	**Facebook**	**21.01%**	Kaspersky Lab	61.81%
3	**Amazon**	**17.46%**	**Facebook**	**54.37%**
4	Cloudflare	7.72%	**Amazon**	**50.33%**
5	**Microsoft**	**7.41%**	Cloudflare	32.37%
6	Twitter/X	7.00%	**Microsoft**	**29.69%**
7	ComScore	6.70%	Twitter/X	29.49%

[5] https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=GR-FI-PT

Two significant observations emerge from Table 3: Google dominates the rankings both in terms of tracker reach and site reach, but Apple is notably absent. Figure 1 demonstrates that Apple is the GAFAM company that tracks the least on the web.

Table 4: 2023 Annual Revenue/GDP (in billions)

Company/Country	Revenue/GDP
Amazon	**574.78 USD**
Apple	**383.28 USD**
Google	**307.39 USD**
Finland	295.53 USD
Portugal	289.11 USD
Greece	243.50 USD
Microsoft	**211.91 USD**
Facebook	**134.90 USD**

It is important to clarify that, in Figure 1, the spike in site reach shown in February 2019 is due to a change in how site reach is measured.[6] As expected from Table 3, Google dominates the realm of web tracking, with trackers present on almost 100% of the top 10,000 websites visited by the users of the Ghostery extension.

The reach, site reach, and average number of trackers data suggest a three-tier stratification among GAFAM in the web tracking space:

(1) Google is leading by a large margin.
(2) Facebook, Microsoft, and Amazon seem to be competing for the remaining surveillance opportunities.
(3) Apple appears strategically absent, possibly recognizing asymmetric competitive conditions and focusing their surveillance activities in alternative spaces.

Another relevant tracker metric is the Content-Length average of third-party tracking requests. As shown in Figure 1, this metric presents findings that do not exhibit particular trends. These patterns suggest that this data should be considered cautiously. As specified in the documentation of the dataset's variables, the Content-Length serves as an approximation. Empirical conclusions cannot be drawn on the accurate amount of bandwidth used by GAFAM's user tracking. However, presenting this data is important to showcase that trackers incur a cost borne by users [4].

Three surprising patterns emerge from the data:

First, Google's and Amazon's trackers seem to consume as much as 10MB of data on average when loading a page that contains their trackers. While caching would mitigate this to some extent, 10MB is very large for web content. The distribution of this data is supposed to have a fat-tail towards high Content-Length because some of the tracker domains also serve audio and video content [7]. Averaging data with such a distribution will inflate the average, thus explaining this surprisingly large Content-Length.

Second, there is a sharp drop in Google trackers' Content-Length in June 2023. Upon closer inspection of the dataset, it became apparent that it was primarily due to the "Youtube" tracker which corroborates the assumptions for the first pattern. This fall could be explained by an update to their tracking software that would stop reflecting the Content-Length of audio and visual content. Interestingly, in June 2023, Google's Privacy Sandbox initiative announced a change in their Topics API that is explicitly described as reducing data. Google present the Topics API as "designed to enable websites to serve relevant ads in a privacy-preserving manner, without resorting to covert tracking techniques, like browser fingerprinting. Topics utilizes several techniques to preserve user privacy, **including reducing data** [emphasis added], [...]."[7]

Third, starting in May 2024, all the trackers' Content-Length trends toward zero. It seems improbable that all the GAFAM companies suddenly decided to reduce the bandwidth usage of their trackers at the same time. Rather, this pattern suggests a more fundamental shift in the way web tracking operates, showcasing the web tracking space as ever evolving, which further adds to its opacity. This shift is made obvious by Figure 2 which highlights a decrease in the average number of requests and an even clearer trend in requests detected as containing identifying information. May 2024, denoted by the black dotted line in Figure 2, corresponds to Google's Privacy Sandbox announcement at Google I/O 2024 about third-party cookie deprecation in Google Chrome[8] further corroborating our assumption of an evolving web tracking landscape.

There is a high likelihood that the second and third patterns are related as they correspond to two announcements concerning the same products made by the same entity within Google (Privacy Sandbox)[9]. Those two successive patterns bring to light another clue hinting at Google being the unequivocal leader of web tracking: Google first announced a change in the Chrome web browser API that is immediately implemented on their services (notably Youtube). That change is then advertised and promoted through an announcement at Google I/O a year later, which forced all the other GAFAM to follow suit and gradually adopt that change. One could even argue that Google is the architect of the web tracking landscape.

Figure 3 focuses on the second-tier companies of the web tracking space, namely Facebook, Microsoft, and Amazon. Here, Facebook demonstrates a steady declining reach and site reach while simultaneously increasing the number of trackers and Content-Length size. This potentially indicates intensification strategies to maximize data extraction despite a declining presence. Microsoft exhibits slow but gradual increases in reach and site reach, with a particularly sharp increase in site reach in December 2023. This likely reflects Microsoft's strategic positioning in the artificial intelligence space and their partnership with OpenAI, requiring expanded data acquisition for model training.

Recognizing that Google is the dominant player in surveillance, Figure 4 presents the company's advertising revenue growth overlaid on their tracker reach. A plausible interpretation of the inverse correlation suggests an increased efficiency in advertising revenue extraction despite the reduced user reach. The rise in reach until 2020, where the trend reverts, may indicate a "critical mass" of data has been acquired which enabled optimization of advertising revenue generation without a corresponding increase in tracking

[6] Refer to Table 1 for an explanation about the change in methodology

[7] https://privacysandbox.google.com/blog/topics-enhancements#update_june_15_2023 - Internet Archive link

[8] https://privacysandbox.google.com/blog/google-io-2024 - Internet Archive link

[9] https://privacysandbox.com - Internet Archive link

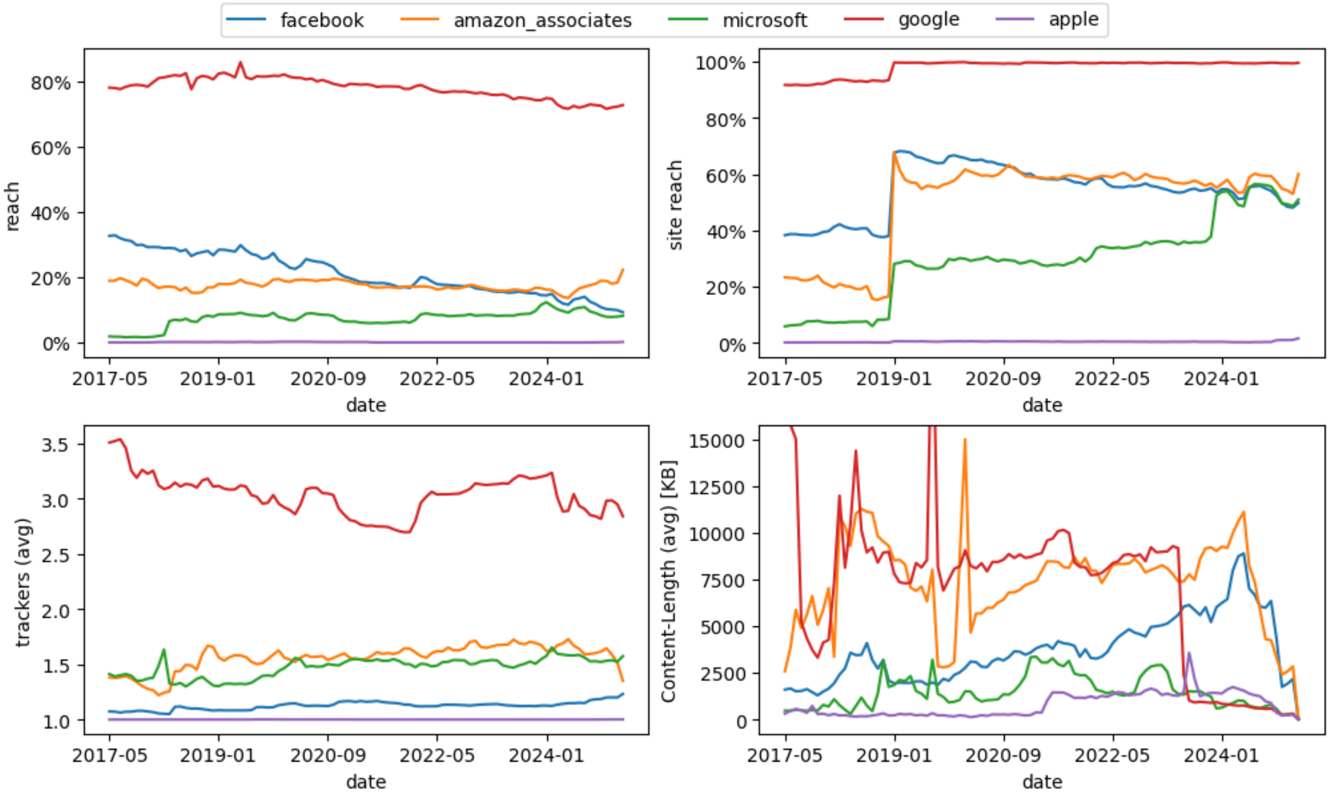

Figure 1: Historical evolution of GAFAM's web tracking

reach. Furthermore, looking at the three companies[10] showcased in Figure 4, the graphs do not appear to show any correlation between tracking reach and ad revenue. This suggests that changes in tracking reach, as measured in the WhoTracks.Me dataset, do not have a direct impact on the companies' advertisement revenue. It is also worth noting that the impact of the Covid-19 pandemic can only be seen in the financial data of Google.

5 Discussion

Web tracking represents an integral part of Surveillance Capitalism's new economic order. As Zuboff argues, 'Big Other' — her term for the expression of power produced by the uncontested global architecture of computer mediation essential to Surveillance Capitalism — is constituted by mechanisms of extraction, commodification, and control. Web tracking is a materialization of this first step of extraction. With this paper's findings, it becomes clear how "High tech firms, led by Google, perceived new profit opportunities in these facts. Google understood that were it to capture more of these data, store them, and analyze them, they could substantially affect the value of advertising. As Google's capabilities in this arena developed and attracted historic levels of profit, it produced successively ambitious practices [...]." [19]. The difficulty in accurately capturing tracking activities is the foundation upon which the power imbalance of Surveillance Capitalism is built. In essence, web tracking functions as a one-way mirror where GAFAM learn about users and their habits, while providing little to no transparency regarding what kind of information they extract, where they store it, and how they leverage it for profit.

5.1 Apple & Google

In Figure 1, Apple displays the least amount of web tracking activity. This seems consistent with Apple's publicly held stance of defending the privacy rights of their customers. However, the recent class action lawsuit settlement regarding Apple's Siri eavesdropping on their user raises questions about that stance.[11] Even more so, in the age of Surveillance Capitalism, it seems implausible for a corporation of Apple's scale not to leverage its customers' metadata for profit generation. A plausible hypothesis, warranting further investigation, suggests that Apple conducts surveillance through alternative channels rather than the web. An example of such an alternative channel is the App Store. Each time an Apple device searches for, comments on, rates, or installs an app, it sends requests to Apple's servers, leaving behind a significant amount of valuable digital traces.

[10] Both Microsoft and Apple were omitted because extracting advertisement revenue from their SEC filings presented significant challenges discussed in the Limitations section.

[11] https://www.cbc.ca/news/business/apple-siri-privacy-settlement-1.7422363 - Internet Archive link

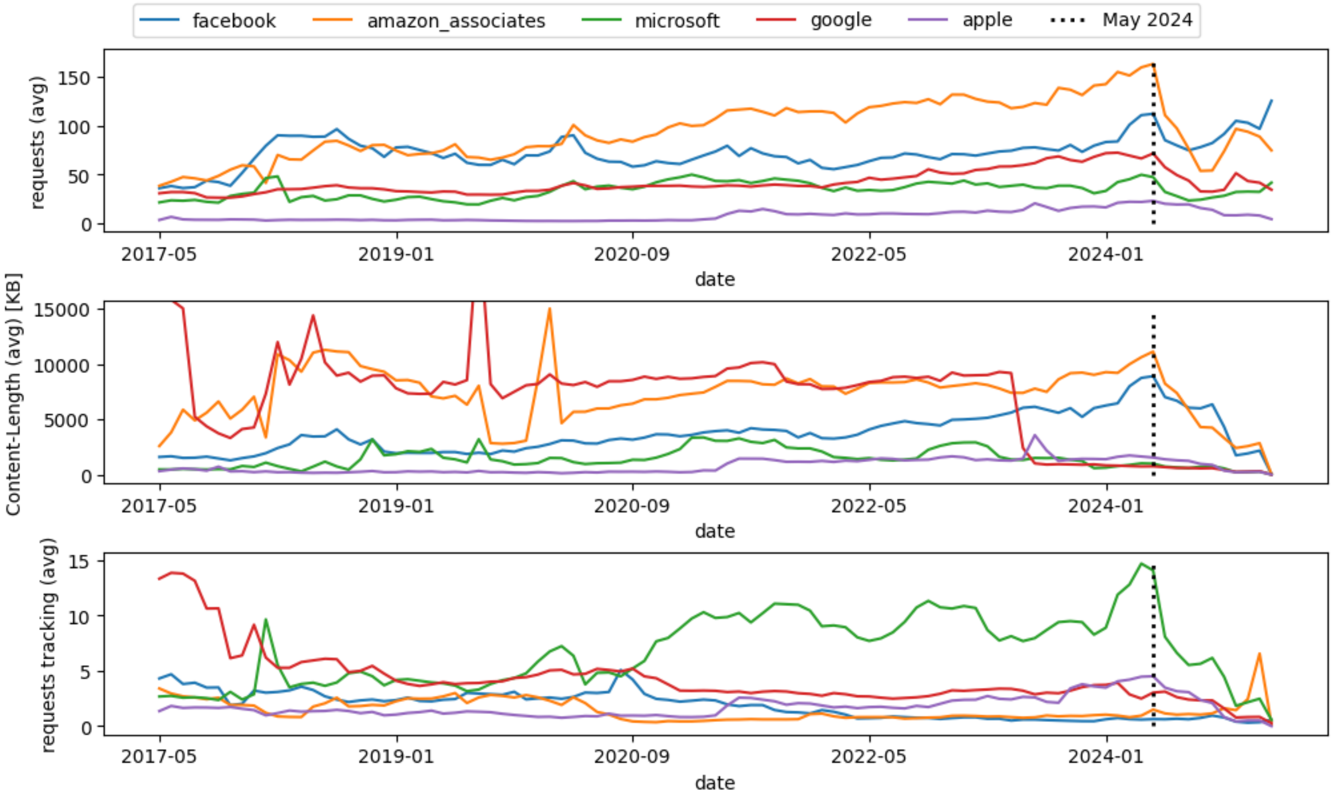

Figure 2: Historical evolution of third-party trackers requests

This may explain Apple's resistance to the integration of progressive web-apps (PWAs), a feature of web browsers introduced by Google[12], in their mobile operating systems. It is a strategy potentially designed to maintain users within the confines of the Apple's closed ecosystem rather than letting them venture into Google's surveillance territory. This understanding could also explain the controversial stance of Apple advocating for their customer's privacy as a right. Apple's threat model is obscure, they never disclosed to their users against who they would protect their privacy and against who they would not.

Though Apple is the least active in the web tracking space, Google is the key player. As Zuboff explains, Surveillance Capitalism originated at Google, making their dominant position unsurprising. An ex-Googler that worked on the Google Chrome team claimed that "the web is what browser vendors ship, you know, that's just the reality".[13] This type of testimony, along with documented empirical evidence of significant web tracking by Google and its latest attempt at overtaking the web tracking landscape with its Privacy Sandbox initiative[14], reveals a fundamental corporate strategy: transforming the open web into a product to generate profit from.

5.2 Cost of Surveillance Capitalism

To quote Zuboff [19], for Google (and other GAFAM companies), "What matters is quantity not quality... Google is 'formally indifferent' to what its users say or do, as long as they say it and do it in ways that Google can capture and convert into data." Such an extractive approach is particularly concerning since Surveillance Capitalism's carries both social and environmental costs. GAFAM's 'formal indifference' enables their obsession to accumulate data for commodification and profit without concern for the damages to humans, society, or the environment.

Web trackers enable companies to measure and optimize various metrics (e.g., views, clicks, user engagement, conversion rate, impressions, etc.). This optimization, if left unchecked, can have dramatic social consequences. 'Formally indifferent' companies can and will trigger behaviors to generate data for capture and commodification. An example is the dissemination of increasingly politically divisive and controversial content online [13, 16]. The economic value of behavioral data derives from the ability to influence consumer behavior. With this power in the hands of 'formally indifferent' corporations, it has been used to increase consumerism without concern for the environmental burden it has placed on our already destabilized ecosystems and limited resources.

[12] PWAs are a feature of web browsers that allow users to install a web-app/website very similarly to a native app. This has the side effect of bypassing the various app-stores altogether, effectively promoting even further the web from a space to share document, to an application platform.
[13] https://www.localfirst.fm/2 - Internet Archive link
[14] https://www.theverge.com/2021/3/16/22333848/google-antitrust-lawsuit-texas-complaint-chrome-privacy - Internet Archive link

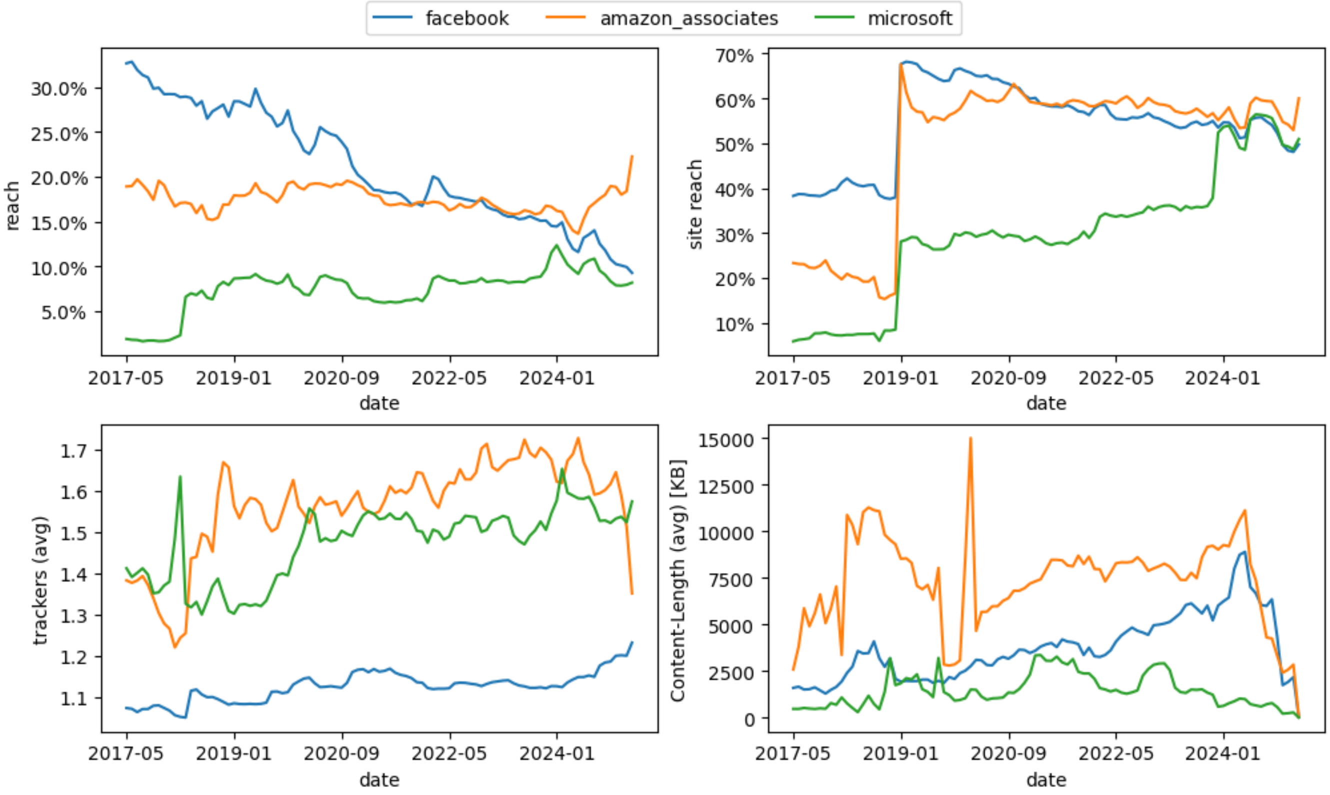

Figure 3: Historical evolution of "FAM's" web tracking

This investigation revealed web tracking's non-zero bandwidth cost [4]. Given more precise and reliable measurements, this bandwidth could be linked to CO2 emissions [3, 14]. Beyond transmitted tracking data, the surveillance apparatus enables targeted advertising and encourages sophisticated online advertising campaigns against web users. Far from innocuous, ads incur bandwidth and additional power consumption costs [8].

5.3 An Alternative to the Surveilled Web

Though Surveillance Capitalism is omnipresent, various online communities[15] and movements[16] have taken an active stance against Surveillance Capitalism. As much as those communities deserve attention, a more radical alternative needs to be highlighted, one that leaves behind most modern web technologies — JavaScript, CSS, Cookies, and other elements essential to the digital surveillance economy — while preserving the web's most essential function: sharing and browsing hyperlinked documents. There exists alternative, non-extractive pathways that can serve as blueprints to share and engage with content on the internet.

Gemini[17] is described as "a new internet technology supporting an electronic library of interconnected text documents." The Gemini protocol is similar in function to HTTP but with a limited subset of functionalities centered around document sharing. Together with the Gemtext format — a simple markup language for Gemini similar in function to HTML for the web — they form the foundation of a small but active alternative network. Gemini's lightweight nature and limited features foster an ecosystem of free and open-source software promoting simplicity, transparency, sharing, and learning. Importantly, Gemini's radical departure from the web makes its content inaccessible from mainstream browsers unless translated and mirrored to web technologies (possible due to Gemtext's simplicity). This pseudo-isolation combined with technological simplicity has enabled a subculture and community to part ways with the corporate web and its extractive and harmful surveillance practices. The narrow feature-set provided by Gemini limits corporate interests and creates a "[...] lightweight online space where documents are just documents, in the interests of every reader's privacy, attention and bandwidth."[17]

6 Limitations

Several limitations are present in this paper. Most of the data used in this investigation was sourced from WhoTracks.Me. Ghostery is a privacy-enhancing extension that cannot freely collect information on its users. This feature prevents systemic control for sample bias. In fact, evidence of such biases can be seen in the popularity ranking of websites within the data. For March 2025, fiverr.com was ranked as the most popular website globally, placing google.com

[15] https://smolweb.org - Internet Archive link
[16] https://indieweb.org - Internet Archive link
[17] https://geminiprotocol.net - Internet Archive link

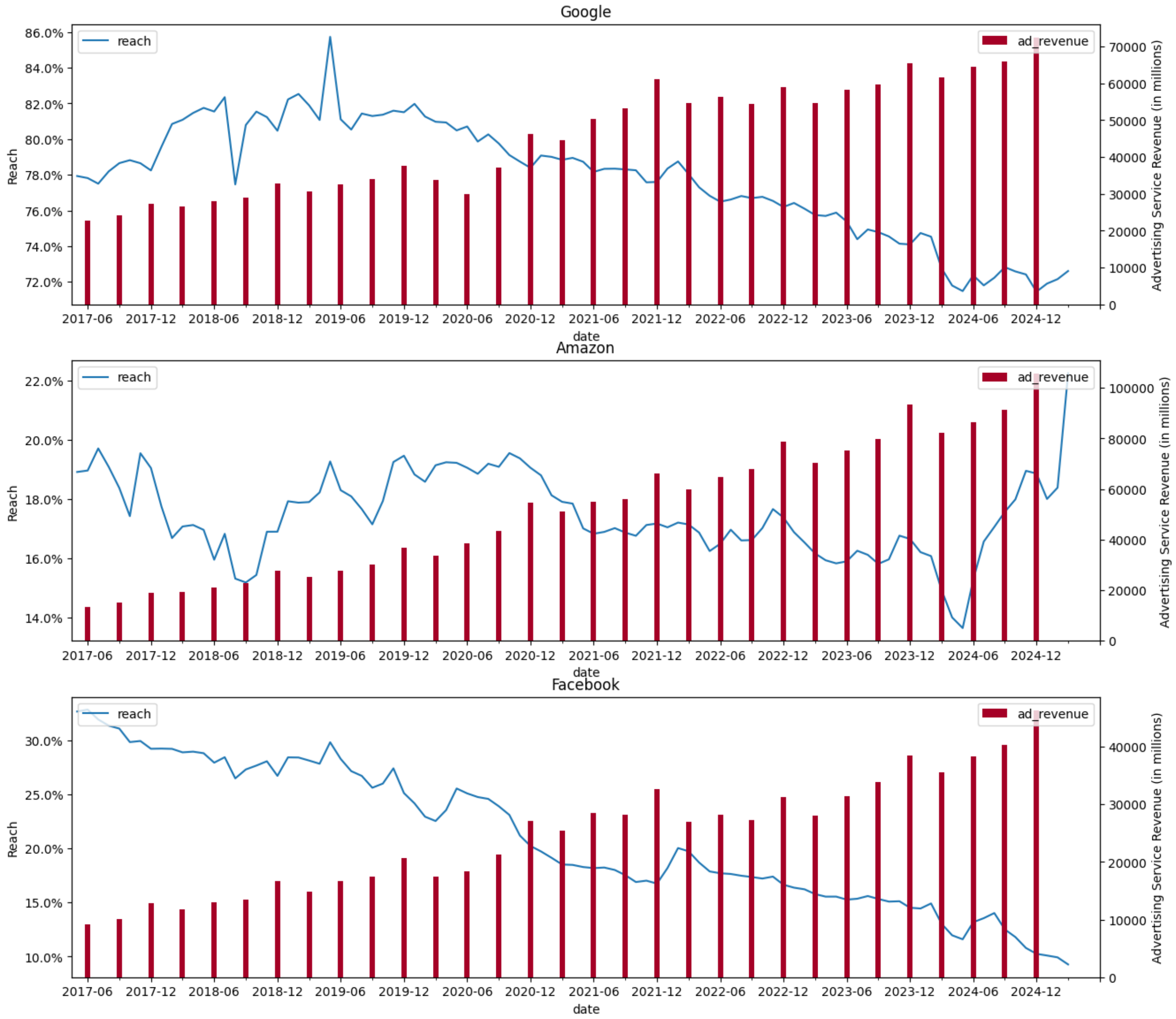

Figure 4: Google, Amazon and Facebook's quarterly advertising revenue overlayed on top of historical reach[*]

[*] Note: The y-axes are all on different scales as we mostly interested by the trends

second. For December 2023, the fifth most popular website was loot.tv. These anomalies nonetheless do not invalidate the whole dataset as the rest of the website rankings are relatively consistent with rankings coming from established sources such as Semrush[18] or Similarweb[19]. There is also no reason to believe that these irregularities disrupt longitudinal trends.

Furthermore, as demonstrated in the findings, the WhoTracks.Me data cannot be relied upon to accurately measure the amount of data consumed by web tracking. The data instead suggests that web tracking relies on data (such as scripts, cookies, images, etc.) transmitted via HTTP requests and this data consumes bandwidth. The cost of web tracking is non-zero and warrants further study with more reliable metrics to establish a proper lower bound on the data consumption of web tracking.

A second limitation relates to the dataset's exclusive account of page loads containing trackers. If a website does not track its users, it will not appear in the dataset. This forecloses the analysis to the possibility of discovering potential alternatives or "ways-out" of the Surveillance Capitalism apparatus. Though, looking at the big picture, all top 50 most frequented websites track their users. This suggests minimal impact on the findings of this investigation.

[18] https://www.semrush.com/website/top/

[19] https://www.similarweb.com/top-websites/

There are inherent limitations to using financial data from the quarterly and annual earnings reports. Unfortunately, access to comprehensive financial data typically requires the use of proprietary or subscription-based services. Performing manual collection of the financial data from 10-Q and 10-K filings limited our analysis to a subset of the financial metrics. 10-Q and 10-K filings do not have a standardized format across companies or even across years within the same company's filings, making specific data, such as ad revenue, challenging to extract. For instance, Microsoft changes the definition of its advertisement category. This is reflected in the name of the line item that has changed over the years from "Advertising" to "Search advertising" to "Search and news advertising", permitting inconsistent reporting of ad revenue. In the case of Apple, the reporting of ad revenue is hidden within a broader "Services" category which reports Advertising, AppleCare, Cloud Services, Digital Content, and Payment Services revenues under a single number.

Lastly, Surveillance Capitalism is a large and systemic phenomenon characterized by complex system interactions. Although web tracking provides valuable insights into a concrete mechanism of surveillance and monetization, these findings are, however, limited only to one category of data extraction. Surveillance Capitalism encompasses many additional aspects: data storage and processing, prediction services derived from that data, and infrastructure designs enabling various extraction and manipulation practices, such as dark patterns and digital rights management. The combination of web tracking and SEC filing data can only provide some clues about a small part of this broader socioeconomic phenomenon. It does not capture how the tracking data is then processed and used, nor does it provide insights into potential harms, such as undermining democratic processes or diverting attention away from the environmental consequences of unbridled neoliberalism.

7 Conclusion

This empirical investigation provides valuable insights into how Surveillance Capitalism is operationalized by the world's most powerful technology companies. Our longitudinal analysis of tracking data from May 2017 to March 2025 reveals patterns that confirm and extend our theoretical understanding of the extent of surveillance on the web.

Google was identified as the dominant surveillance entity on the web, with tracking reach largely exceeding the other GAFAM companies. The analysis of our results surfaced a three-tiered stratification among the GAFAM companies, with Google at the top, Facebook, Amazon, and Microsoft competing in a second tier, and Apple strategically absent. This stratification hints at how surveillance practices reflect and reinforce capital market power dynamics.

The relationship between tracking reach and advertising revenue, particularly in Google's case, suggests that Surveillance Capitalism may be entering a new phase characterized by more efficient and aggressive data exploitation rather than merely increasing the scale of collection. This evolution has roots in reality, as Google's trackers are already present on almost 100% of the top websites. Further expansion is unfeasible forcing Google to innovate ways to extract behavioral predictive data out of a "data exhaust" that already reached its maximum.

This study confirms that web tracking data can serve as concrete empirical evidence of Surveillance Capitalism. By quantifying surveillance activities, through web tracking, and connecting them to financial metrics, we move beyond qualitative frameworks to measurable phenomena.

Several promising directions for future research emerge from this investigation. Expanding the historical scope beyond 2017 could provide an understanding of the continued evolution of Surveillance Capitalism. The approach suggested by Lerner et al. to leverage the Internet Archive's Wayback Machine to identify historical trackers offers a promising approach [10]. Apple's apparent lack of presence in web tracking warrants closer examination of how Surveillance Capitalism operates within closed ecosystems. Recognizing how surveillance can manifest under seemingly privacy-respecting systems can reveal different forms of data extraction and behavioral surplus generation. Developing more sophisticated methods for extracting and analyzing financial data in relation to surveillance activities — potentially through automated extraction and analysis of SEC filings — can also surface new insights into the economic mechanisms of Surveillance Capitalism. In particular, ad revenue may be related to increasingly manipulative practices to engage users with politically divisive content. This phenomenon deserves greater attention, especially with political polarization increasing within Western democracies [12]. Finally, comparative studies examining impacts of jurisdictional differences, particularly between the EU and US, on web tracking and financial metrics could illuminate how regulatory environments shape surveillance practices and enable or limit surveillance-enabled economic growth.

As web technologies continue to evolve with tracking practices, ongoing empirical monitoring of surveillance practices remains essential to understand both the current and future trajectory of Surveillance Capitalism. This study contributes to that effort by demonstrating the value of web tracking data as a window into the actual mechanisms through which our collective online experience is captured and commodified.

Acknowledgments

Christoph for inspiring this investigation and for his guidance. Sayaka for her unwavering support and advice.

References

[1] Tel Amiel, Filipe Saraiva, Leonardo Ribeiro Da Cruz, and Priscila Gonsales. 2023. Mapping Surveillance Capitalism in South American Higher Education. *Revista Latinoamericana de Tecnología Educativa - RELATEC* 22, 1 (Jan. 2023), 221–239. doi:10.17398/1695-288X.22.1.221

[2] John Bailey, Mikael Laakso, and Linus Nyman. 2019. Look Who's Tracking. *Informaatiotutkimus* 38, 3-4 (Dec. 2019). doi:10.23978/inf.87841

[3] Marion Ficher, Françoise Berthoud, Anne-Laure Ligozat, Patrick Sigonneau, Maxime Wisslé, and Badis Tebbani. 2021. Assessing the carbon footprint of the data transmission on a backbone network. In *2021 24th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. 105–109. doi:10.1109/ICIN51074.2021.9385551 ISSN: 2472-8144.

[4] Molly Hanson, Patrick Lawler, and Sam Macbeth. 2018. The Tracker Tax: the impact of third-party trackers on website speed in the United States. (2018).

[5] Rasmus Helles, Stine Lomborg, and Signe Sophus Lai. 2020. Infrastructures of tracking: Mapping the ecology of third-party services across top sites in the EU. *New Media & Society* 22, 11 (Nov. 2020), 1957–1975. doi:10.1177/1461444820932868

[6] Garrett A. Johnson, Scott K. Shriver, and Samuel G. Goldberg. 2023. Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR. *Management Science* 69, 10 (Oct. 2023), 5695–5721. doi:10.1287/mnsc.2023.4709

[7] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M. Pujol. 2019. WhoTracks .Me: Shedding light on the opaque world of online tracking. doi:10.48550/arXiv.

1804.08959 arXiv:1804.08959 [cs].

[8] Khan Awais Khan, Mohammad Tariq Iqbal, and Mohsin Jamil. 2024. Impact of Ad Blockers on Computer Power Consumption while Web Browsing: A Comparative Analysis. *European Journal of Electrical Engineering and Computer Science* 8, 5 (Oct. 2024), 18–24. doi:10.24018/ejece.2024.8.5.650 Number: 5.

[9] Marvin Landwehr, Alan Borning, and Volker Wulf. 2023. Problems with surveillance capitalism and possible alternatives for IT infrastructure. *Information, Communication & Society* 26, 1 (Jan. 2023), 70–85. doi:10.1080/1369118X.2021.2014548

[10] Ada Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner

[11] Karlo Lukic, Klaus M. Miller, and Bernd Skiera. 2024. The Impact of the General Data Protection Regulation (GDPR) on Online Tracking. doi:10.2139/ssrn.4399388

[12] Jennifer McCoy, Tahmina Rahman, and Murat Somer. 2018. Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist* 62, 1 (Jan. 2018), 16–42. doi:10.1177/0002764218759576 Publisher: SAGE Publications Inc.

[13] Killian L. McLoughlin, William J. Brady, Aden Goolsbee, Ben Kaiser, Kate Klonick, and M. J. Crockett. 2024. Misinformation exploits outrage to spread online. *Science* 386, 6725 (Nov. 2024), 991–996. doi:10.1126/science.adl2829 Publisher: American Association for the Advancement of Science.

[14] Joanna Moulierac, Guillaume Urvoy-Keller, Marco Dinuzzi, and Zhejiayu Ma. 2023. *What is the carbon footprint of one hour of video streaming?* Technical Report. Université Côte d'Azur. https://hal.science/hal-04069500

[15] Michael Power. 2022. Theorizing the Economy of Traces: From Audit Society to Surveillance Capitalism. *Organization Theory* (July 2022). doi:10.1177/26317877211052296 Publisher: SAGE PublicationsSage UK: London, England.

[16] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 140–149. doi:10.1145/3287560.3287580

[17] Nayanamana Samarasinghe and Mohammad Mannan. 2019. Towards a global perspective on web tracking. *Computers & Security* 87 (Nov. 2019), 101569. doi:10.1016/j.cose.2019.101569

[18] Zhonghao Yu, Sam Macbeth, Konark Modi, and Josep M. Pujol. 2016. Tracking the Trackers. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 121–132. doi:10.1145/2872427.2883028

[19] Shoshana Zuboff. 2015. Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology* 30, 1 (March 2015), 75–89. doi:10.1057/jit.2015.5 Publisher: SAGE Publications Ltd.

[20] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* PublicAffairs. Google-Books-ID: lRqrDQAAQBAJ.

A Source Code

Table 5: Software Tools

Name	Version	Website
Python	3.12.4	https://www.python.org
Jupyter Lab	4.3.4	https://jupyter.org
Pandas	2.2.3	https://pandas.pydata.org
Matplotlib	3.10.1	https://matplotlib.org

The code is published on the author's own website[20] in the form of a Fossil repository in addition to an archive of the code in the Internet Archive[21]. GitHub has been considered but will be avoided as it currently belongs to Microsoft, one of the key actors of Surveillance Capitalism. Table 5 lists the software tools along with their version that were used for the empirical analysis part of this study.

[20] https://fsl.blazebone.com/empirical_inquiry_into_sc_limits_2025_source_code

[21] https://archive.org/details/empirical_inquiry_into_sc_limits_2025_analysis