The Bitter Pill: Tracking and Remarketing on EU Pharmacy Websites

Zahra Moti, Kimberley Frings, Christine Utz, Frederik Zuiderveen Borgesius, and Gunes Acar

Radboud University

Abstract. We investigate online tracking and remarketing practices on 50 pharmacy websites in five European countries, focusing on information shared with third parties. By manually shopping for pregnancy tests and automatically analyzing the HTTP traffic data captured in HAR files, we find that users' personal data and shopping activities are routinely collected by third parties. Many pharmacy websites share product names, email addresses and phone numbers with third parties even when consent was declined. Investigating novel forms of online tracking, we find several cases of server-side tagging and CNAME-based tracking, which can be used to circumvent tracking protections offered by adblockers and modern browsers. Monitoring the advertisements targeted to our shopping profiles on several news websites and large online platform apps, we find re-targeted advertisements of the pregnancy tests we had shopped for. We further find that while declining consent reduces third-party data sharing, it does not eliminate it, and deceptive designs often discourage users from opting out. Through GDPR data access requests we reveal that companies vary in the completeness of the personal data they disclose, with none providing a full list. Overall, our study reveals widespread potential legal violations and adoption of evasive tracking technologies on websites that handle users' most sensitive personal data.

Keywords: Privacy, online tracking, pharmacy, online advertising

1 Introduction

Over the past few years, the online pharmacy sector has grown significantly, driven by the convenience of home delivery, price comparisons, and customer reviews. Online pharmacies offer convenience but pose risks due to extensive data sharing with third parties for ads and analytics. The online pharmacy section of Walgreens has also previously been shown to leak prescription information to *session replay* companies [1]. A recent investigation by The Markup showed that third-party data collection was taking place on 49 out of 50 US telehealth websites, in certain cases for targeted advertising purposes [22]. The US Federal Trade Commission investigated ad-related data sharing by GoodRx, BetterHelp, and Cerebral, resulting in multimillion-dollar settlements and bans on sharing data with advertisers like Facebook [23–25]. While these investigations showed

the risks for the US users, it is unclear whether European online pharmacy users are protected by stricter privacy laws. Our study conducts an empirical investigation to answer this question, considering novel tracking mechanisms and re-targeted advertisements.

In the context of online pharmacies, data collected by third parties may include pages visited, products browsed, purchases made, and even personal information entered during checkout. Such data can be used for relatively innocuous purposes, such as improving user experience. However, users' activities on pharmacy websites could also be used for advertising and marketing. Previous work has shown that many telehealth websites leak personal data to third parties [22], which can reveal intimate details about a user's private life. In regions where reproductive healthcare is contentious, tracking data may even be used in legal prosecution [6, 36].

This paper investigates the prevalence and nature of online tracking and ad retargeting (remarketing) practices on 50 European online pharmacies. Specifically, we focus on pharmacy websites that offer non-prescription medications in the four most populous EU countries —Germany, France, Spain, and Italy —as well as in the Netherlands. We focus on the most popular pharmacy websites in each country, as they attract the majority of users and reveal the tracking practices most consumers are likely to encounter. We examine the prevalence of third-party data collection by simulating a user shopping for pregnancy products. In addition, we study novel tracking mechanisms such as CNAME-based tracking and Server-Side Tagging/Tracking (SST), which bypass tracking protections that rely on blocklists. Further, we attempt to trace how the collected data is used by examining the advertisements we receive on the Web after our pharmacy browsing sessions. We also use GDPR rights to request our data from the large third parties that collect data through pharmacy websites. To evaluate the effectiveness of user controls, we compare tracking and advertising practices in two scenarios: when website visitors accept cookies and when they reject them. Overall, our contributions include the following:

- We compare tracking practices on 50 pharmacy websites across five European countries, including novel methods such as SST and CNAME-based tracking.
- We quantify personal information and product name leaks to tracker domains, showing the extent of leaks even if the user declines consent.
- Through an exploratory study of retargeted ads based on our shopping activity on pharmacy websites, we show that even sensitive products such as pregnancy tests are used for ad retargeting.
- We compare client-side data collection by major platforms with their GDPR data access responses, revealing significant discrepancies.

2 Related Work

Our study builds upon prior work on web tracking in health-related contexts and considers novel tracking techniques.

Tracking on health-related websites. Several papers investigated third-party tracking on health-related websites, with most focusing on the United States. Friedman et al. [28] researched abortion clinic websites, while McCoy et al. [38] focused on websites related to COVID-19. Both studies relied on webXray [54] to log third-party requests and cookies, a scope that likely underestimates harder-to-detect methods such as server-side tracking. Nevertheless, 99% of pages in both studies contained third-party trackers. In 2022, The Markup collaborated with STAT to investigate telehealth websites in the US [22]. They analyzed the presence of third-party trackers and shared data type (e.g., product details or shopping cart items). Among 50 telehealth websites, all but one sent personal details—often hashed or even plaintext email addresses—to major tech companies, most during checkout or questionnaire submission. In 2023, a study of 12 U.S. drugstores [53] found that all shared information about viewed or purchased products with major tracking companies.

Research into the tracking practices of European health-related websites is more sparse. Rauti et al. [49] analyzed the tracking practices on 163 Finnish online pharmacies. They found that 57 (35%) pharmacies leaked both the queried prescription name and identifying personal data. Yu et al. [55] studied 19,483 hospital websites in 152 countries—including 5,936 in Europe—and found tracking scripts on 53.5% of sites worldwide (48.8% in Europe) and tracking cookies on 14.6% (7.5% in Europe). Cookiebot, a Danish company, conducted similar research on EU health and government websites and found that 52% of EU public health service websites contained commercial trackers [8].

Emerging web tracking techniques. As major browsers block third-party trackers and cookies, websites increasingly adopt new techniques to bypass these restrictions. One such method is CNAME-based tracking, which uses DNS aliases to disguise trackers as first-party resources. Dimova et al. [14] presented a large-scale, longitudinal study of this technique, finding increasing adoption, especially on high-traffic sites, and posing serious security risks due to bypassing the Same-Origin Policy. Another emerging technique is Server-Side Tagging (SST), introduced by Google in 2020 [26]. Unlike client-side tracking, SST shifts data collection to a server, hiding tracking activity from the user's browser. In a recent study, Fouad et al. [27] investigated SST at scale. They flagged SST domains by identifying subdomains absent in pre-2020 crawls, confirming they were registered to entities other than the parent domain and that their requests included tracking data previously sent elsewhere.

Our approach. Unlike prior work, such as Rauti et al. [49], we examine tracking after consent is declined, quantify email and phone number leaks, identify CNAME cloaking and server-side tracking using a history-free detector, and link these leaks to retargeted ads seen on the Web and mobile. To identify SST endpoints on websites, we took a different approach from Fouad et al. [27], who compared website behavior before and after SST implementation and found SST on 28 websites. Instead, our analysis relies on fixed URL parameter structures and request initiators, yielding a much higher prevalence of SST. A caveat is that our method focuses on Google Tag Manager's SST implementation due to its

popularity, rather than detecting generic server-side tracking. Finally, we leverage GDPR data access rights to compare data collected on pharmacy websites by large online platforms to data disclosed in response to subject access requests.

3 Methods

We investigate tracking and advertising practices on 50 pharmacy websites across five EU countries. We simulated shopping for pregnancy tests under two consent conditions (accept/reject), using fresh browser profiles, predefined personas, and VPNs. We analyzed HTTP traffic to identify trackers and detect techniques like server-side tagging and CNAME cloaking. To assess advertising, we monitored targeted ads on news websites and mobile apps. Finally, we compared GDPR data access responses with our observed tracking activity.

3.1 Website Selection

When studying online pharmacies, we distinguish between those offering prescription and non-prescription medications. As regulations differ across countries, we focus exclusively on websites selling non-prescription drugs to maintain consistency. We target popular, legitimate pharmacy websites, as users are more likely to visit them. Under Directive 2011/62/EU [20], legitimate pharmacies must register with national authorities and link to an official database. We retrieved registered pharmacies from Germany [7], France [47], Italy [42], Spain [4], and the Netherlands [41]. Popularity rankings were based on Similarweb's DigitalRank [50]. We selected the top ten pharmacies per country to balance breadth and manual feasibility, while ensuring our sample includes the sites most online shoppers for pharmacy products are likely to visit.

3.2 Data Collection

We collected data in two distinct phases, as described below. The first phase focused on tracking and web-based retargeting (Algorithm 1), while the second focused on ads on large online platforms' mobile apps (Algorithm 2).

Algorithm 1: Measurement of Tracking and Targeted Ads. We followed a fixed procedure for each website to capture all relevant HTTP traffic in a reproducible manner. Algorithm 1 presents a high-level overview of our data collection process, outlining the steps we followed to capture HTTP traffic across various consent modes, countries, and pharmacy websites. We started with a fresh browser profile for each website and followed the steps below for each consent mode in every country, across all pharmacy websites in our dataset:

1. Open Developer Tools, enable HTTP logging, and detach the panel to avoid detection influencing tracking behavior [44].
2. Load the homepage and handle the cookie dialog per consent mode.

Algorithm 1 Tracking and Targeted Web Ads Analysis

```
for each consent mode do
    Prepare predefined personal info
    for each country do
        Use VPN to simulate location
        for each pharmacy website do
            Create a fresh profile
            Checkout a product
            Save the HAR file
            Check news websites for ads
        end for
    end for
end for
```

Algorithm 2 Analysis of Data Collection by Large Online Platforms

```
for each consent mode do
    Create a fresh profile
    Log in to platform accounts
    for each country do
        Use VPN to simulate location
        for each pharmacy website do
            Checkout a product
        end for
    end for
    Check platform apps for ads
    Scroll for 2 minutes
    Wait until the next day
end for
```

3. Search for pregnancy tests or browse the menu if no results appear.
4. View the first product page, return to the results, and open the next product.
5. Add the product to the cart, adjusting quantity if required.
6. Proceed through checkout as far as possible without placing the order, using guest checkout and predefined personal data; register if required.
7. Save all HTTP requests and responses as an HTTP Archive (HAR) file.

We leveraged HAR files to identify tracking-related requests using the uBlock Origin Core npm package [33]. We relied on uBlock Origin's default filter lists, including EasyList and EasyPrivacy, among others [32]. We then mapped tracker domains to their respective owner entities using DuckDuckGo's entity map [16].

Targeted Ads on the Web. After visiting each pharmacy website, we visited a set of news websites to observe any targeted or retargeted ads resulting from the prior shopping activity. We used Similarweb [50] to select the top five "content publishing" sites per country and five global sites, as these categories include ad-supported news websites and align with prior ad targeting research [13]. We excluded duplicates and subscription-based, ad-free sites. To analyze ad behavior and disclosures, we followed these steps:

1. Visit the homepage and interact with the cookie banner.
2. Scroll to the bottom of the page or stop after 10 seconds for infinite scrolling.
3. If pregnancy ads appear, click the AdChoices icon for the explanation page.
4. Visit two inner pages (prioritize the most prominent items and avoid health-related pages) and follow steps 4 and 5 above.

We acknowledge that our data collection incurred some cost on the pharmacy websites' advertising budgets by causing ad impressions during the advertisement monitoring. We believe the societal benefits of our investigation outweigh its negligible cost to advertisers.

Algorithm 2: Data Collection by Large Online Platforms. To investigate whether data collected by third parties was used for personalized ads and disclosed to users properly, we created separate Instagram, Microsoft, TikTok, Facebook, Snapchat and Google accounts for each consent mode (accept/reject) on two iPhones. Algorithm 2 outlines the data collection process: For each consent mode, we created a fresh browser profile, logged into the six platform accounts, and searched for pregnancy tests on each pharmacy site, proceeding through checkout as far as possible without payment. After completing daily website visits, we monitored the online platforms' mobile apps three times a day in two-minute scrolling sessions, continuing for up to a week[1]. We captured screenshots of any ads related to health, pharmacies, or pregnancy tests. Finally, we requested and examined data downloads from these platforms (§3.4).

3.3 Measurement Setup

Our experiments were conducted using Chromium browsers running on Ubuntu 24.04.1 LTS. For sites with a cookie banner, we collected data in both "accept" and "reject" modes; if no banner appeared, the same data was used for both. We used two separate computers per mode to minimize the cross-contamination risk between different consent modes. Visiting the same website twice, even after clearing cookies and browser history, could still allow tracking through fingerprinting, potentially influencing ads based on prior visits. Using two computers minimizes the risk of cross-contamination between browsing sessions of different consent modes. We used a predefined persona on each computer during checkout, allowing us to later check if personal data was leaked to third parties. To access the websites from their respective countries of origin, we used Mullvad VPN [43]. This enabled us to better impersonate a local pharmacy shopper, which may be relevant for ad targeting.

3.4 Detecting Tracking Methods and Leaks

CNAME-based Tracking. A potential method to bypass blocklist-based tracking protection is CNAME-based tracking. To evade blocking, the website owner maps a first-party subdomain to the tracker's domain via CNAME records. Due to the increasing popularity of this technique [14], many defenses, such as uBlock Origin and AdGuard—have introduced countermeasures [31, 39]. uBlock Origin, for instance, performs DNS lookups and replays filtering with the resolved CNAME address. We adopt this method, which is enabled by uBlock Origin's `cnameReplayFullURL` option. If a hostname has a CNAME record, we replace it with the resolved domain and rerun tracker detection using the uBO Core npm package [33]. DNS lookups are automated using the dnspython library [15].

[1] We did not monitor ads on Google mobile apps, as our Web-focused measurement (Algorithm 1) targets Google ads on websites. For Microsoft, we used the Bing app; for others, we used their respective mobile apps.

Server Side Tagging. Many websites embed multiple third-party resources, adding performance overhead due to increased page weight. Adblockers and tracking protections now offered by many mainstream browsers block tracking- and advertising-related third-party traffic. Server Side Tagging (SST) was proposed as a way to reduce this overhead of third parties while also bypassing tracking protections [26]. In SST, the end user's browser or mobile app only sends tracking and analytics data to a single server, which then relays it to multiple third parties (a.k.a. tags). SST may make it challenging to identify the third parties collecting data on a website, and hence poses a transparency problem. To detect SST usage, we relied on a simple observation. Despite the change in endpoints, many URL parameters used to send data remain the same. For instance, in both SST and non-SST integrations, Google Analytics uses the parameters `dt`, `dl`, and `sr`, which correspond to page title, page URL, and screen dimensions, respectively. However, instead of manually picking parameters, we automated the parameter detection using our dataset to bootstrap the process. We first identified all requests triggered by Google Tag Manager (GTM) scripts using `initiator` fields, since SST uses GTM under the hood [29]. To detect self-hosted GTM scripts, we used a pattern we extracted from the official GTM scripts. We then took the intersection of URL parameters observed in requests triggered by GTM scripts. This yielded a list of 36 parameters, which we searched for in all requests. Similar to Fouad et al. [27], we then verified whether these requests were indeed SST by retrieving the IP address pointing to the first-party subdomain in the request and checking to which organization this IP address is registered. Then we used the terminal command `whois` to check whether the first-party subdomain organization differs from the website. We also used the request initiators for further confirmation.

Detecting Product Name and Personal Information Leaks. When placing an order, users provide personal information such as name, address, email, and product details, which may be shared with third parties. Identifying when and how different types of data are shared can be challenging, particularly across languages. To enable systematic analysis, we compiled search terms including product names and persona details used during checkout. Personal information or product names can be sent to tracking parties using encodings or cryptographic hashes such as SHA-256 [40]. To detect such transformed leaks, we followed Englehardt et al. [17] to search for permutations of various encodings and hashes (e.g., Base64, SHA-256) in request URLs and POST bodies.

Data Retrieval from Third Parties. We retrieved personal data from major platforms via their account settings or privacy centers. From Google, we exported service-wide activity data. Facebook and Instagram provided lists of companies sharing off-site activity with Meta, including browsing and purchases [35]. TikTok's "Ads and data" section contained advertising-related data. Microsoft's Privacy Dashboard included ad profiles and inferred interests. From Snapchat, we downloaded user data such as purchase history, memories, and other account activity. Note that data requests were made using automated tools provided by the platforms, without contacting any employees.

Table 1: Most common categories of third-party entities found on pharmacy websites. It shows the number of websites where requests to these domains observed, along with the number of distinct request domains and entities per category.

Entity Category	Websites		Request domains		Request entities	
	Accept	Reject	Accept	Reject	Accept	Reject
Advertising	50	49	73	52	58	41
Ad-motivated tracking	50	49	72	50	56	38
Analytics	50	45	45	32	37	27
3rd party analytics marketing	49	45	34	25	33	24
Audience measurement	49	43	27	20	24	17
Ad fraud	39	28	8	7	5	4

3.5 Analysis of Consent Notices

To provide insights into the mechanisms that online pharmacies offer customers to control the processing of their personal data, we manually inspected screenshots taken from each pharmacy's main page for the presence of consent notices and the options they offer. We focus on control mechanisms available on the first layer of the notices, as only a few people are willing to explore deeper layers of consent notices for options to deny consent [46]. Our analysis was guided by the requirements of European data protection authorities that it must be as easy to reject data collection as to consent to it [19]. Thus, we annotated the screenshots of consent notices for the interaction options offered to website visitors on the first layer and their formatting and placement within the banner. One of the authors did the annotations, and edge cases were resolved in joint discussion.

4 Findings

4.1 Third Parties and Tracking

We analyzed HTTP requests and responses from the captured HAR files to identify third parties and various types of data sharing with them. All pharmacy websites embedded at least one third-party domain, regardless of giving or declining consent. The median number of third-party domains per site varied substantially across countries—16 in France and 56 in Italy—with Italian and German sites embedding the most (Figure 1). Rejecting cookies reduces the number of third-party domains across all countries, with Germany seeing the largest drop. Also, we identified a substantial number of pharmacy websites where third parties set cookies with the `SameSite=None` attribute and a lifespan exceeding two months—47 and 33 websites in accept and reject mode, respectively. Analyzing cookie purposes is out of the study scope, but `SameSite=None` cookies enable third parties to track users across domains.

Tracker entities. A large portion of third-party embeds on pharmacy websites were classified as trackers. Figure 1 shows the median number of tracking entities per website, revealing substantial variation across countries: French

websites had the fewest entities (median of nine in accept mode), while Italian websites had the most (median of 43.5). Rejecting cookies generally reduced the number of trackers, except for French sites.

Most prevalent trackers. As shown in Table 2, Google appeared on 96% of websites, followed by Microsoft, Facebook, and PayPal. Another frequently encountered third party is Criteo, which specializes in personalized advertising [12]. Other prevalent trackers such as Awin [5], Outbrain [48], and Taboola [52] were linked to marketing, native ads, and content recommendations. We also identified ID5 [34], a provider of privacy-focused identity solutions designed to replace third-party cookies. The presence of these trackers on pharmacy sites raises privacy concerns, as users may not expect health-related browsing activities to be used for ads or data sharing.

Table 2: Frequent third-party entities on pharmacy websites.

Third-party entity	Accept	Reject
Google	50	48
Microsoft	36	16
Facebook	32	13
Virtual Minds	18	7
Criteo	16	3
ID5	16	2
PayPal	15	15
Trusted Shops	14	12
Awin	13	4
Outbrain	13	2

Prevalence of third-party categories. Table 1 summarizes third-party service categories across pharmacy sites. Domains were categorized based on the Tracker Radar dataset [16]; with some domains belonging to multiple categories. "Advertising" and "Ad motivated tracking" appeared on nearly all websites, with over 70 unique domains and 50 entities focused on tracking.

CNAME-based Tracking. To detect CNAME-based tracking, we replaced request hostnames with their CNAME records and reran detection using uBO Core (§3.4). Focusing on requests that were detected as trackers only after the CNAME replacement, we identified six distinct pharmacy websites that use CNAME-based tracking (Table 3). Registrable domains of all CNAME records appear in the EasyPrivacy list blocklist. Etracker.com describes how site owners can "avoid data loss due to ad blocking" using CNAME records [18]. Similarly, Mapp's help pages [37] explain how to set up first-party tracking by defining a CNAME record pointing to `go-direct.flx1.com`, a domain used by two pharmacies. In another case, despite rejecting consent, our persona's name and email were sent to Spotler (`activate.deonlinedrogist.nl`), which provides email marketing services [51].

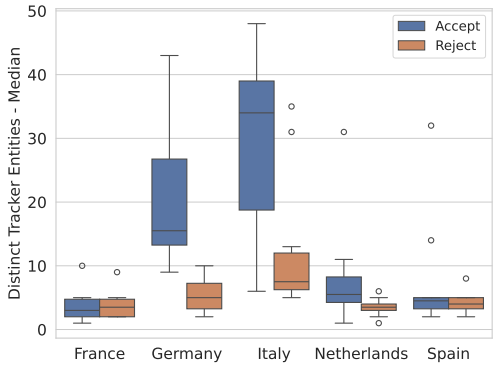

Fig. 1: Median of tracker entities per site by consent mode and country.

Server Side Tagging. Using the URL parameters in GTM traffic (§3.4), we found that 19 of the 50 sites used SST. In all cases, a first-party subdomain

Table 3: Detected CNAME-based tracking domains, showing the original request host, the resolved CNAME and the consent mode(s) in which they were observed.

Loc.	Website	Request host	CNAME	Cons.
DE	medikamente-per-klick.de	e.medikamente-per-klick.de	customer.etracker.com	Both
IT	farmasave.it	ddbm2.paypal.com	ddbm2.paypal.com.[...].datadome.co	Reject
IT	topfarmacia.it	dmp.email.topfarmacia.it	go-direct.flx1.com	Both
IT	docpeter.it	dmp.mapp.docpeter.it	go-direct.flx1.com	Both
IT	1000farmacie.it	the.sciencebehindecommerce.com	tag.device9.com	Accept
NL	deonlinedrogist.nl	activate.deonlinedrogist.nl	ujemkxutgo.relay.squeezely.tech	Both

of the pharmacy website was used. In 14/19 cases, the SST endpoint was used in both accept and reject modes. Four of the five German websites used the SST endpoint only in accept mode, while the only French pharmacy used it only in reject mode. We found that 12 of the 19 SST servers were hosted on Google, easing the setup of SST servers [29]. To determine the hosting details, we used a combination of `Via` and `Server` response headers captured in the HAR files and additional WHOIS information we queried for the server IP addresses. The majority of SST endpoints used the default `/collect` path of Google Analytics, while two Italian pharmacies used a random path starting with `ngt`. Use of a random path could be an additional effort to evade blocking.

4.2 Product Name and Personal Information Leaks

We examined two types of information leakage: product names and identifying personal details. To prevent false positives, we only considered leaks to third-party domains and to 19 SST hostnames identified in §4.1.

Table 4: SST endpoints by country (Loc.) and used consent mode (A: Accept, R: Reject).

Loc.	SST Endpoint	Google Hosted	Cons.
DE	measure.medpex.de/g/collect	True	A
DE	tmsst.aponeo.de/g/collect	True	A
DE	klpoz.shop-apotheke.com/g/collect	True	A/R
DE	sgtm.mycare.de/g/collect	False	A
DE	measure.docmorris.de/g/collect	True	A
IT	otasf.redcare.it/g/collect	True	A/R
IT	gtm.efarma.com/g/collect	False	A/R
IT	sgtm.farmasave.it/ngtwyxyzwjg	False	A/R
IT	sgtm.docpeter.it/ngtmapwbued	False	A/R
ES	datos.farmaciasdirect.es/g/collect	True	A/R
NL	pipeline.drogist.nl/g/collect	True	A/R
NL	metrics.deonlinedrogist.nl/g/collect	True	A/R
NL	sgtm.plein.nl/g/collect	True	A/R
NL	ecom-data.trekpleister.nl/g/collect	True	A/R
NL	ecom-data.kruidvat.nl/g/collect	True	A/R
NL	inc.da.nl/g/collect	False	A/R
NL	v3-pixal-web.etos.nl/g/collect	False	A/R
NL	sst.koopjesdrogisterij.nl/g/collect	False	A/R
FR	care.soin-et-nature.com/g/collect	True	R

Product name leaks. In accept mode, 34 websites leaked the product name, 77% via URLs and 23% via POST request bodies. 28 websites leaked product names even when consent was declined. Google was the top recipient of product name leaks (36 Accept, 23 Reject; Table 5), followed by Microsoft, ByteDance, and Facebook. While leaks to `doubleclick.net` dropped substantially in reject mode (24 to 3), leaks to `google-analytics.com` increased (12 to 17), which may be a fallback domain in reject mode. Product names are still leaked to several third

parties in reject mode. For instance, `efarma.com` (IT) leaked product names to six domains. In contrast, seven of ten German sites avoided such leaks, while leakage patterns in other countries remained largely unchanged (Table 6). On all sites but two, URL encoding is used when leaking the product name to third parties or SST hostnames. On `shop-apotheke.com` (DE) and `redcare.it` (IT) the product name was leaked in Base64 encoded form to `adtriba.com`, a digital marketing company [3].

Personal information leaks. To examine personal data leaks, we focused on email addresses and phone numbers, which uniquely identify users. As with product name analysis, we considered only third-party domains and SST hostnames. In accept mode, emails leaked in 15 cases and phone numbers in three; in reject mode, email leaks slightly dropped to 13, while phone leaks rose to four. SHA-256 was the most common hashing/encoding method observed in email leaks (39 of 164). Overall, hashed email leaks were detected on five distinct sites. Facebook received hashed emails from three websites in accept mode and from two sites in reject mode (`boticas23.com`, `okfarma.es`). Other domains receiving hashed emails include `awin1.com`, `zenaps.com`, `dynamicyield.com`, `pinterest.com` and `tiktok.com`. Notably, `awin1.com` received a salted hash, which prevents linking user identities via hashed emails.

Table 5: Number of sites leaking product names to third-party entities.

Entity	Accept	Reject
Google	36	23
Microsoft	23	7
ByteDance	7	0
Facebook	4	1
Virtual Minds	4	0

4.3 Consent Notices

All but one of the 50 online pharmacies (`pharmaciedesdrakkars.com`, France) displayed a consent notice. Since consent notices often employ deceptive design patterns to steer website visitors towards accepting all cookies and tracking technologies [46], our analysis focused on whether the pharmacies transparently communicated options to decline such data collection. As we will discuss in §5, EU privacy law requires consent to be "freely given," which requires equal prominence for "Accept" and "Reject" options. A 2023 report by the European Data Protection Board found that the majority of surveyed national data protection authorities considered embedding refusal links within text paragraphs invalid unless they are visually highlighted to attract users' attention [19]. We only found five pharmacies (two from each NL and ES and one from IT) that display "Accept" and "Reject" as equally prominent options on the first layer. While 27 additional pharmacies did feature a "Reject" option on the first layer, 21 used color highlighting to point visitors towards the "Accept" option and nine did not place the "Reject" option next to the "Accept" button. 12 pharmacies did not feature any explicit "Reject" option on the first layer, but four of them had an "Accept

Table 6: Number of sites leaking product names per country and consent mode.

Country	Accept	Reject
Germany	10	3
Spain	10	9
Italy	9	10
Netherlands	8	9
France	7	7

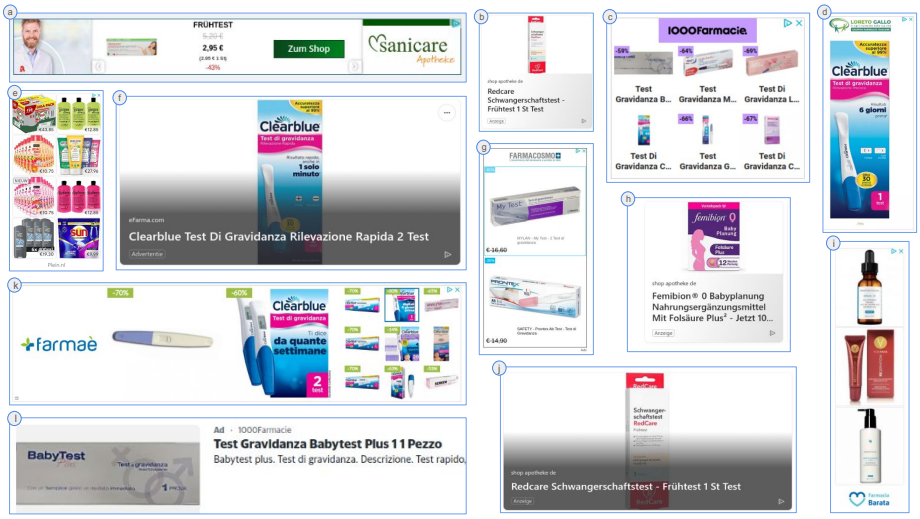

Fig. 2: Examples of targeted ads observed during the experiment.

necessary [cookies]" button instead. Thus, only five out of 50 online pharmacies featured a consent dialog that did not outright violate the requirement for "freely given" consent. A comprehensive legal analysis of whether valid consent was actually obtained would require in-depth assessment on a case-by-case basis.

4.4 Advertisements

Targeted Ads on the Web. To assess the impact of targeted and retargeted ads, we visited news websites after browsing pharmacy sites (§3.2). The results showed notable cross-country differences in targeted and retargeted ads. We observed such ads from 15 of 50 pharmacy websites in the Netherlands, Spain, Germany, and Italy, but none from France. In the Netherlands and Spain, we saw no re-targeted ads but did receive pharmacy-related ads from visited sites (`plein.nl` and `farmaciabarata.es`) via Google Ads (Figure 2 e, i). In some cases, we saw pregnancy-related ads, though not for the exact items searched, suggesting broader behavioral targeting (Figure 2 h from `shop-apotheke.com` in Germany and d from `farmacialoreto.it` in Italy). Retargeted ads appeared on two out of ten German sites and three out of ten Italian sites, matching products we had browsed or added to our cart. These ads were served by Google, Microsoft, Criteo, as well as Taboola (Figure 2 l) and RTB House.

Ads on Large Online Platform Apps. A day after visiting pharmacy websites, our accept-mode Facebook feed showed numerous pregnancy- and baby-related posts and reels, but no ads. TikTok and Instagram displayed ads, yet none for pregnancy products or pharmacies. This absence may be due to fresh, low-credibility profiles and, for Facebook, the off-Facebook-activity setting—found

disabled after the study. In an earlier pilot with an author's long-standing account, pregnancy-related Facebook ads did appear.

4.5 Data Takeout from Third Parties

Under the GDPR, companies that process personal data must honor data access rights. Comparing each platform's Takeout archive with our HAR logs—and the retargeted ads we later observed—shows that none provided a complete record.

Based on our HAR logs, 37 of the 50 sites contacted Google Analytics, but Google Takeout returned records for only 27. The Takeout data included only visited URLs, but omitted other data collected by Google Analytics, such as product names, prices, quantities, and cart actions. In contrast, HAR captures the full Analytics payloads, revealing complete product metadata and user-action events collected by Google. This gap highlights the incompleteness of Google Takeout data compared to the detailed, real-time tracking in its analytics services. TikTok's "Off TikTok Activity" log includes events such as `InitiateCheckout`, `ViewContent`, and `AddToCart`, but provides minimal metadata—for example, checkout entries lack product details. In contrast, our HAR logs show seven sites sending product names and eight sending hashed personal data (email, phone, name) to TikTok, none of which appeared in the returned data. Instagram's Takeout data listed advertisers that used our "activity or information", including okfarma.es and unrelated brands such as Netflix and Paramount. The "ads and topics" folder, which logs viewed and clicked ads, contained no ads from pharmacy websites. Facebook's "Activity Off Meta" Takeout yielded only generic privacy details and no records of pharmacy websites, despite ads related to pharmacies and pregnancy. Post-collection, we learned that off-Facebook activity ads were disabled, which may explain the absence of records. Microsoft's ad dashboard showed new interest labels (e.g., Baby and Children) and served related ads on MSN (Figure 2), but the downloaded profile lacked the underlying data. Snapchat's Takeout contained no data on our pharmacy visits, consistent with our client-side observations.

5 Legal Analysis

Here we provide a brief legal discussion—not an individual compliance assessment—of the tracking practices identified in this paper, focusing on the General Data Protection Regulation [21]. The GDPR generally applies to the tracking practices discussed in this paper because it applies when "personal data" such as cookies and other online identifiers, are used. The GDPR applies to companies ("data controllers") based in the EU, but also to certain non-EU companies, e.g., if the company "monitors" the behavior of people in the EU (Art. 3(2)), as in online tracking. The online pharmacy and the tracking company are jointly responsible for GDPR compliance [10]. GDPR defines "special categories of personal data" that include "data concerning health or [...] a natural person's sex life" (Art. 9(1)). The Court of Justice of the European Union (CJEU) stated that

data concerning health "must be interpreted broadly", so if somebody orders a medical product at an online pharmacy, that fact constitutes data concerning health [9]. The use of sensitive personal data is prohibited, subject to specific exceptions such as for hospitals, which do not apply here.

The only possible legal basis for online tracking and targeted advertising is the Internet user's "explicit consent" (Art. 9(2)) [11]. For consent to be valid, it needs to be a "freely given, specific, informed, and unambiguous indication of the data subject's wishes by which [they], by a statement or by a clear affirmative action, signif[y] agreement to the processing of [their] personal data" (Art. 4(11)). This means that the individual must actively do something, e. g., tick a box or click a button. A company is not allowed to assume consent if someone continues to use a service or fails to opt out. As noted in §4.3, we saw tracking for targeted advertising without the individual's "freely given" consent, a clear violation.

6 Limitations

While our 50-site sample favors depth over breadth, covering the top ten pharmacies per country likely reflects the experience of millions of users [2]. We observed targeted ads from 15 of the 50 pharmacies. The absence of ads from other websites could be due to a lack of advertising campaigns targeting the products we shopped for. Our use of fresh profiles on separate devices minimized the risk of prior browsing history influencing tracking behavior and ad delivery, though residual effects beyond our control cannot be entirely excluded. While our study focuses on pregnancy tests and results may not fully generalize to other sensitive health-related products, the observed tracking and ad targeting demonstrate how even sensitive product purchases are monitored for ad retargeting. Future investigations could also examine whether browsing for such products triggers ads for related categories, for example, baby items, to shed light on the broader profiling strategies employed by advertisers. We used manual checkouts to avoid bot detection and ensure ecological validity. Future work could explore LLM-guided automation [45], though this may trigger bot detection or ad fraud defenses. Google allows users to limit ads about sensitive topics such as "pregnancy and parenting" [30]. Due to scope limitations, we could not evaluate the effect of this opt-in setting. While we searched for various types of encodings and hashes to identify leaked data, custom encodings or obfuscation can bypass our detector. Hence, our ad targeting results should be taken as lower bounds. Our SST endpoint identification method focused on the server-side use of Google Tag Manager, rather than generic server-side tracking. Since our method relies on common URL parameters extracted from the data we collected, it may not generalize to other datasets or more customized uses of SST.

7 Conclusion

Users may expect a high level of privacy when shopping for health-related products online. Our findings show that even shopping for sensitive products such as pregnancy tests on most popular European pharmacy websites is subject to extensive third-party tracking for advertising purposes. Through a lightweight detection method, we identify a sharp increase in the use of server-side tracking, along with continued use of other stealthy techniques such as CNAME cloaking. Tracking often occurs without valid consent as many websites do not use compliant consent dialogs, and some ignore user choices altogether. Moreover, data access requests often yield incomplete information, leaving users in the dark about what online activities are monitored. Our findings raise significant concerns regarding transparency, user rights, and compliance with regulations.

References

1. Acar, G., Englehardt, S., Narayanan, A.: No boundaries: data exfiltration by third parties embedded on web pages. Proceedings on Privacy Enhancing Technologies. p. 220–238 (2020). https://doi.org/10.2478/popets-2020-0070
2. Adamic, L.A., Huberman, B.A.: Zipf's law and the internet. Glottometrics **3**(1), 143–150 (2002)
3. AdTriba GmbH: Future-proof marketing measurement & optimization (Oct 2024), https://www.adtriba.com
4. Agencia Española de Medicamentos Productos Sanitarios: Listado de farmacias que realizan la venta a distancia (2024), https://distafarma.aemps.es/farmacom/faces/inicio.xhtml
5. AWIN Inc.: Join our global affiliate platform (2024), https://www.awin.com/
6. Baker-White, E.: Facebook Gave Nebraska Cops A Teen's DMs., https://www.forbes.com/sites/emilybaker-white/2022/08/08/facebook-abortion-teen-dms
7. Bundesinstitut für Arzneimittel und Medizinprodukte: Versandhandelsregister (Oct 2024), https://versandhandel.dimdi.de/pdfs/vhr-apo.pdf
8. Cookiebot: Ad Tech Surveillance on the Public Sector Web (2019), https://www.cookiebot.com/media/1121/cookiebot-report-2019-medium-size.pdf
9. Court of Justice of the EU: Judgment in Case C-21/23, 4 Oct. 2024, https://curia.europa.eu/juris/liste.jsf?num=C-21/23
10. Court of Justice of the EU: Judgment in Case C-40/17, 29 July. 2019, https://curia.europa.eu/juris/liste.jsf?num=C-40/17
11. Court of Justice of the EU: Judgment in Case C-446/21, 4 Oct. 2024, https://curia.europa.eu/juris/liste.jsf?num=C-446/21
12. Criteo: The Commerce Media Platform for the Open Internet (2024), https://www.criteo.com/
13. Datta, A., Tschanz, M.C., Datta, A.: Automated Experiments on Ad Privacy Settings – A Tale of Opacity, Choice, and Discrimination. Proceedings on Privacy Enhancing Technologies. (1), 92–112 (2015)
14. Dimova, Y., Acar, G., Olejnik, L., Joosen, W., Van Goethem, T.: The CNAME of the Game: Large-scale Analysis of DNS-based Tracking Evasion. Privacy Enhancing Technologies. pp. 394–412 (2021)

15. Dnspython Contributors: dnspython (2024), https://dnspython.readthedocs.io/en/latest
16. DuckDuckGo: Tracker Radar (2024), https://github.com/duckduckgo/tracker-radar/
17. Englehardt, S., Han, J., Narayanan, A.: I never signed up for this! Privacy implications of email tracking. Proceedings on Privacy Enhancing Technologies. pp. 109–126 (2018)
18. eTracker GmbH: Set up your own tracking domain (2025), https://help.etracker.com/en/article/set-up-your-own-tracking-domain
19. European Data Protection Board: Report of the work undertaken by Cookie Banner Taskforce. Tech. rep. (2023), https://www.edpb.europa.eu/our-work-tools/our-documents/other/report-work-undertaken-cookie-banner-taskforce
20. European Parliament and the Council of the EU: Directive 2011/62/EU (2011), http://data.europa.eu/eli/dir/2011/62/oj
21. European Parliament and the Council of the EU: Regulation (EU) 2016/679 (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj
22. Feathers, T., Palmer, K., Fondrie-Teitler, S.: Dozens of Telehealth Startups Sent Sensitive Health Information to Big Tech Companies (2024), https://themarkup.org/pixel-hunt/2022/12/13/out-of-control-dozens-of-telehealth-startups-sent-sensitive-health-information-to-big-tech-companies
23. Federal Trade Commission: Enforcement Action to Bar GoodRx (2023), https://www.ftc.gov/news-events/news/press-releases/2023/02/ftc-enforcement-action-bar-goodrx-sharing-consumers-sensitive-health-info-advertising
24. Federal Trade Commission: FTC Order Prohibits Telehealth Firm Cerebral from Using Sensitive Data for Ads (2024), https://www.ftc.gov/news-events/news/press-releases/2024/04/proposed-ftc-order-will-prohibit-telehealth-firm-cerebral-using-or-disclosing-sensitive-data
25. Federal Trade Commission: FTC to Ban BetterHelp from Revealing Consumers' Data (2024), https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-ban-betterhelp-revealing-consumers-data-including-sensitive-mental-health-information-facebook
26. Fisher, B.: Improve performance and security with Server-Side Tagging (2023), https://blog.google/products/marketingplatform/360/improve-performance-and-security-server-side-tagging
27. Fouad, I., Santos, C., Laperdrix, P.: The Devil is in the Details: Detection, Measurement and Lawfulness of Server-Side Tracking on the Web. Proceedings on Privacy Enhancing Technologies p. 450–465 (2024)
28. Friedman, A.B., Bauer, L., Gonzales, R., McCoy, M.S.: Prevalence of Third-Party Tracking on Abortion Clinic Web Pages. JAMA Internal Medicine **182**(11) (2022)
29. Google: Setting up a new server container (2023), https://developers.google.com/tag-platform/learn/sst-fundamentals/4-sst-setup-container
30. Google: Limit ads about sensitive topics on Google (2024), https://support.google.com/My-Ad-Center-Help/answer/12155260
31. Hill, R.: uBlock Origin works best on Firefox, https://github.com/gorhill/uBlock/wiki/uBlock-Origin-works-best-on-Firefox#cname-uncloaking
32. Hill, R.: uBlock Origin – make-rulesets.js (2023), https://github.com/gorhill/uBlock/blob/491bc87e94a503a17fd11cdee35c1f1b6fea24be/platform/mv3/make-rulesets.js#L1285-L1296
33. Hill, R.: uBlock Origin Core (2024), https://www.npmjs.com/package/@gorhill/ubo-core

34. ID5: ID5 – Future-proofed user identification for Digital Advertising (2024), https://id5.io/
35. Instagram Help Center: Why am I seeing ads from an advertiser on Instagram?, https://help.instagram.com/609473930427331
36. Kaste, M.: Nebraska cops used Facebook messages to investigate an alleged illegal abortion (2022), https://www.npr.org/2022/08/12/1117092169/nebraska-cops-used-facebook-messages-to-investigate-an-alleged-illegal-abortion
37. Mapp: Custom Track Domain (C-Name) (2025), https://docs.mapp.com/docs/custom-track-domain-c-name
38. McCoy, M.S., Libert, T., Buckler, D., Grande, D.T., Friedman, A.B.: Prevalence of Third-Party Tracking on COVID-19–Related Web Pages. Journal of the American Medical Association (JAMA) **324**(14), 1462–1464 (2020)
39. Meshkov, A.: Gotta catch 'em all: how AdGuard scanned the entire web in search of hidden trackers (2024), https://adguard.com/en/blog/cname-tracking.html
40. Meta: Customer File Custom Audiences (2024), https://developers.facebook.com/docs/marketing-api/audiences/guides/custom-audiences/#hash
41. Ministerie van Volksgezondheid, Welzijn en Sport: Aanbiederslijst online medicijnen (2024), https://aanbiedersmedicijnen.nl/aanbieders/aanbiederslijst
42. Ministero della Salute: Soggetto autorizzato al commercio online di medicinali (2024), https://www.salute.gov.it/LogoCommercioElettronico/CercaSitoEComm
43. Mullvad VPN AB: – Free the internet (2024), https://mullvad.net/en
44. Musch, M., Johns, M.: U Can't Debug This: Detecting JavaScript Anti-Debugging Techniques in the Wild. In: Proceedings of the 30th USENIX Security Symposium. pp. 2935–2950 (2021)
45. Müller, M., Žunič, G.: Browser Use: Enable AI to control your browser (2024), https://browser-use.com/
46. Nouwens, M., Liccardi, I., Veale, M., Karger, D., Kagal, L.: Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In: Proceedings of the 2020 CHI. pp. 1–13 (2020)
47. Ordre National des Pharmaciens: Rechercher un site de vente en ligne autorisé à vendre des médicaments – CNOP (2024), https://www.ordre.pharmacien.fr/je-suis/patient-grand-public/rechercher-un-site-de-vente-en-ligne-autorise-a-vendre-des-medicaments?vl-region=&vl-departement=&vl-commune=&vl-site=&vl-pharmacy=&vl-incumbent=
48. Outbrain Inc.: Drive Better Business Results (2024), https://www.outbrain.com/
49. Rauti, S., Carlsson, R., Mickelsson, S., Mäkilä, T., Heino, T., Pirjatanniemi, E., Leppänen, V.: Analyzing third-party data leaks on online pharmacy websites. Health and Technology **14**, 375–392 (2022)
50. Similarweb LTD: Unlock Digital Growth (2024), https://www.similarweb.com/
51. Spotler: Email marketing automation with Spotler software (2024), https://spotler.com/solutions/use-cases/email-marketing-automation
52. Taboola: Restricted Content, Products, Services (2025), https://taboola.com/
53. Tahir, D., Fondrie-Teitler, S.: Need to Get Plan B or an HIV Test Online? Facebook May Know About It (2023), https://themarkup.org/pixel-hunt/2023/06/30/need-to-get-plan-b-or-an-hiv-test-online-facebook-may-know-about-it
54. webXray: webXray Privacy Search Engine (2024), https://webxray.ai/
55. Yu, X., Samarasinghe, N., Mannan, M., Youssef, A.: Got Sick and Tracked: Privacy Analysis of Hospital Websites. In: IEEE Euro S&P Workshops. pp. 278–286 (2022)