Prompt like a Butterfly, Sting like a Tracker: A Privacy Analysis of Web and Mobile Conversational AI Agents

Guilherme Oliveira
IMDEA Networks

Miguel Sanchez
IMDEA Networks

Juan Manuel De Santa Olalla Gómez
IMDEA Networks

Roi S. Serna
IMDEA Networks/UC3M

Tautvydas Jackevicius
IMDEA Networks

Jorge Garcia-Herrero
Independent

Aniketh Girish
IMDEA Networks

Guillermo Suarez-Tangil
IMDEA Networks

Narseo Vallina-Rodriguez
IMDEA Networks

Abstract

As prominent conversational AI providers like OpenAI adopt advertising-based business models, traditional web and mobile tracking practices are expanding into conversational AI services [15]. However, despite their growing adoption, the tracking, data-sharing, and monetization practices of conversational AI services remain largely opaque and have received comparatively limited scrutiny from researchers, regulators, and the public.

In this paper, we present a systematic privacy analysis of the web and mobile deployments of nine prominent conversational AI services. Using a combination of static and dynamic analysis, we study the presence of third-party Advertising and Tracking Services (ATSes), characterize their data flows, and evaluate how consent choices, subscription tiers, and access-control mechanisms influence conversation exposure to third parties. We uncover privacy risks unique to conversational AI platforms: multiple providers disclose sensitive conversation-derived artifacts—including titles, prompts, and screenshots—to third parties, often alongside persistent user identifiers that enable user attribution. We also find that some providers publicly expose conversation permalinks without access controls, allowing trackers to read the entire conversation.

Our findings reveal how traditional tracking technologies are increasingly intertwined with AI-mediated interactions, creating new pathways through which sensitive user and conversational information can be collected, inferred, and disseminated. To assess the broader implications of these practices, we analyze them in the context of the GDPR and ePrivacy Directive. We conducted a responsible disclosure process involving affected providers and competent European Data Protection Authorities. Our results demonstrate that conversational AI services introduce a novel privacy attack surface in which provider-generated conversational artifacts become subject to tracking and public exposure, highlighting the need for stronger safeguards governing AI-mediated interactions.

Keywords

Conversational AI, LLMs, Privacy, Mobile, Web, Trackers

Proceedings on Privacy Enhancing Technologies YYYY(X), 1–18

https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Recent advances in Large Language Models (LLMs) have enabled the emergence of conversational AI services such as ChatGPT, Gemini, and Claude, capable of supporting persistent interactions, multimodal processing, and autonomous task execution. As adoption of these services grows for personal and professional activities [43], service providers are exploring new business models to monetize their growing user bases.

Advertising is emerging as one such model, potentially extending into conversational AI the tracking and attribution infrastructures traditionally associated with web and mobile platforms. For example, Reuters reported that OpenAI partnered with Criteo to conduct an advertising pilot for ChatGPT free-tier users in the United States in early 2026 [52].

However, the integration of these tracking technologies raises distinct privacy concerns. Unlike traditional web and mobile applications, conversational AI services routinely process highly sensitive prompts, contextual information, behavioral patterns, uploaded documents, and persistent interaction histories that may reveal intimate aspects of users' lives and professional activities. The disclosure of such information to third-party tracking services, particularly without meaningful transparency or consent, may therefore expose users and organizations to significant privacy risks.

Prior work by Jazlan et al. has examined the integration of third-party tracking in web-based conversational AI services [32], primarily focusing on identifying trackers and characterizing their data collection practices. However, the unique interaction models of conversational AI services introduce new privacy risks across their web and mobile clients: these services generate conversation-derived artifacts—including conversation identifiers, URLs, titles, previews, prompts, responses, and interaction metadata—that may be disclosed to third parties or exposed through publicly accessible resources. Moreover, how these exposures are shaped by by consent choices, privacy settings, subscription tiers, and access-control mechanisms remains largely unexplored. To address this gap, we investigate three research questions:

- **RQ1:** To what extent do conversational AI services integrate third-party tracking, analytics, advertising, and attribution infrastructures across their web and mobile clients?
- **RQ2:** What conversation-derived artifacts and user information are exposed by conversational AI services, either to third-party entities or through publicly accessible resources, and what privacy risks emerge from their disclosure?

- **RQ3:** How do cookie consent choices, subscription tiers, privacy settings, and access-control mechanisms shape the disclosure and accessibility of information in conversational AI services?

To answer these questions, we conduct a systematic privacy analysis of nine prominent conversational AI services, covering the web clients of all nine providers and the Android clients of the eight that offer an Android mobile app. We combine static and dynamic analysis to evaluate their privacy practices across consent choices, subscription tiers, and access-control configurations. Specifically, we make the following contributions:

(1) Across the evaluated services, we identify 44 third-party organizations and observe that every evaluated AI service integrates at least one third-party advertising or tracking service. We further uncover substantial differences between web and Android clients and identify third-party services that are activated only after users explicitly accept non-essential cookies, demonstrating that consent decisions directly influence the tracking surface of conversational AI platforms (§5).
(2) We uncover novel privacy risks specific to conversational AI services. Unlike traditional tracking systems that primarily observe browsing activity, conversational AI platforms generate artifacts that directly encode user interactions. We show that 6/9 web and 3/8 Android clients disclose conversation URLs, titles, prompts, and screenshots to third-party services, often alongside persistent user identifiers. We further demonstrate that the privacy implications of these disclosures are strongly shaped by consent choices and sharing functionality, with several providers exposing entire conversations through publicly accessible permalinks lacking access controls. These findings reveal new channels through which trackers and external actors can gain access to users' entire conversations (§6).
(3) We conduct a legal analysis of observed practices under EU data-protection law, assessing the compatibility of tracker activation, consent mechanisms, conversation-artifact disclosures, identity-linkage practices, and publicly accessible conversational resources with the GDPR and ePrivacy Directive (§7).

Our findings show that integrating traditional tracking infrastructures into conversational AI services creates novel pathways for exposing sensitive user information, including conversation-derived artifacts and publicly accessible conversations. More broadly, our results challenge the perception of conversational AI services as confidential exchanges between users and AI providers. Instead, they are increasingly integrated into the broader online tracking ecosystem, raising important technical and regulatory challenges for AI-mediated services.

Responsible Disclosure. We followed a responsible disclosure process for all identified issues, notifying affected providers and the competent Data Protection Authorities (DPAs) as described in the Ethical Considerations section.

2 Background

This section provides background on conversational AI services and their growing integration with tracking technologies (§2.1), and on tracking mechanisms commonly deployed across web and mobile platforms (§2.2).

2.1 Conversational Agents

Conversational AI services are LLM-based systems that interact with users through natural language interfaces, typically accessible through web and native mobile clients. The public release of ChatGPT by OpenAI in November 2022 marked a major inflection point in the AI industry, rapidly reaching hundreds of millions of users and triggering an industry race to deploy conversational AI platforms across consumer and enterprise ecosystems [42]. Since then, other providers have released competing services, including Perplexity AI, Anthropic's Claude, Google's Gemini, Microsoft's Copilot, xAI's Grok and DeepSeek.

Despite differences in architecture and deployment models, all these services share several common characteristics. Most conversational AI services maintain persistent user accounts and interaction histories, and integrate external services such as search engines, analytics platforms, telemetry frameworks, advertising infrastructures, and cloud-hosted APIs. Modern AI agents also increasingly support multimodal capabilities, including image analysis, voice interaction, document analysis, browsing assistance, and autonomous task execution.

The rapid development and adoption of these services amplify the economic incentives to introduce data-driven monetization models. Recent industry developments and press releases suggest that conversational AI services are beginning to integrate into the existing advertising and tracking ecosystem rather than replacing it. For example, Criteo reported that 40% of surveyed U.S. consumers already use AI agents for product discovery and shopping assistance [15], while industry actors increasingly describe agentic AI as the next opportunity for targeted advertising and *"agentic commerce"* [2]. Similarly, major tech companies are developing infrastructure that enables AI agents to interact directly with commercial platforms and external digital services, such as Google's Universal Commerce Protocol (UCP), which facilitates AI-driven commerce and interoperable agent ecosystems [4].

2.2 Mobile and Web Tracking

Modern web and mobile services are deeply intertwined with third-party advertising and tracking services that enable user profiling, personalization, attribution, and targeted advertising at scale [48]. On the web, trackers rely on techniques such as cookies, tracking pixels, browser fingerprinting and cookie syncing to collect browser metadata, interaction events, device characteristics, network information, and account-linked identifiers, enabling persistent cross-site identification and behavioral profiling [1, 17, 29, 34]. Mobile applications similarly integrate third-party SDKs [24, 48] that collect device and behavioral data, frequently relying on platform-supported identifiers such as the Android Advertising ID (AAID) and Apple's Identifier for Advertisers (IDFA) [28, 53], and hashed email addresses (HEMs)[1] to support cross-device tracking [62, 64].

Both browsers and mobile operating systems provide mechanisms to limit such tracking. Web browsers increasingly restrict third-party cookies, fingerprinting, and other cross-site tracking

[1]Hashed email addresses (HEMs) are pseudonymous identifiers derived from users' email addresses that can enable identity matching across services without transmitting the email address in plaintext [25].

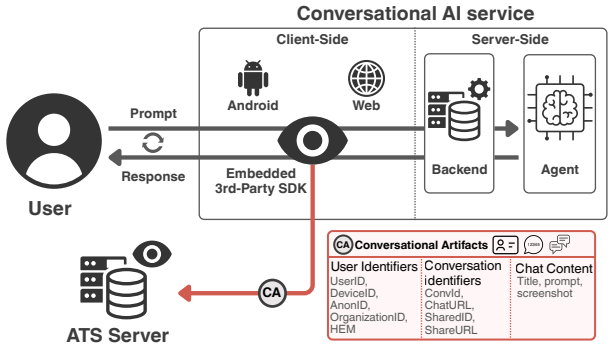


Figure 1: Privacy risks.

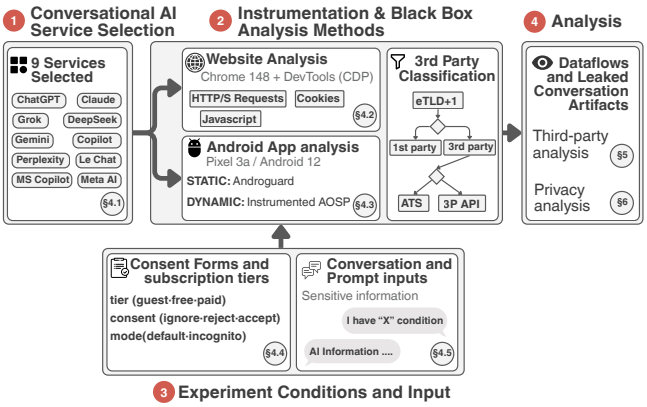


Figure 2: Methodology overview.

techniques, while content blockers and privacy-enhancing extensions can prevent requests to known tracking domains. Mobile platforms instead rely on application sandboxing, runtime permissions, and restrictions on access to advertising identifiers and other sensitive resources. However, these protections do not eliminate third-party data collection: embedded SDKs execute within the host application's security context and may access data and permissions available to the application [28]. These differences in tracking mechanisms and platform protections motivate our separate analysis of the web and mobile clients of conversational AI services, as explained in §4.

3 Privacy Risks in Conversational AI Services

Integrating traditional advertising and tracking technologies into conversational AI services introduces unique privacy risks. Unlike conventional web and mobile services, these services routinely process highly sensitive and contextual information, including free-form conversations, behavioral interactions, uploaded documents, interaction histories, and persistent user profiles. Consequently, embedded third-party tracking technologies can access not only behavioral and device information, but also artifacts that encode the content and context of users' interactions with the AI providers.

We consider three primary entities, shown in Figure 1: the user, the conversational AI service (the first party), and third-party ATSes such as analytics providers, advertising networks, crash reporting services, and anti-fraud products. The conversational AI service comprises both client- and server-side components. Third-party code and libraries embedded in web or mobile clients may collect and transmit information directly from the clients to their cloud infrastructure, while server-side components may disclose information to third parties outside the boundaries of user devices.

Within this setting, new privacy risks emerge when conversation-derived artifacts are disclosed to third parties. These artifacts may include prompts containing information directly provided by users, conversation titles that summarize the content of an interaction, screenshots capturing conversational context, or persistent conversation URLs (permalinks). Unlike tracking methods on conventional web and mobile services, conversation artifacts can directly reveal the content, context, and potentially sensitive nature of users' interactions with the AI services, in addition to the exposure of model responses to potential competitors.

The risks are amplified when such information is disclosed alongside persistent user or device identifiers like email hashes, allowing third parties to associate sensitive conversational data with individual users and potentially link it across sessions or services. Moreover, conversation URLs may provide access to additional content when the underlying resources lack adequate access controls, potentially exposing the entire conversation to third parties.

These risks challenge the perception of conversational AI as a confidential interaction between a user and an AI provider. When combined with conventional tracking infrastructures, the rich and persistent artifacts generated by conversational AI create new channels through which sensitive information can be disclosed, linked to individual users, or made accessible to third parties.

4 Methodology

Figure 2 provides an overview of our research methodology to answer our three research questions. Following this workflow, we first describe the selection of representative conversational AI services (§4.1), followed by our instrumentation and black-box analysis methods for their web (§4.2) and Android clients (§4.3), how we assess the effects of consent choices and subscription tiers (§4.4), and the conversation inputs used to trigger systematic and reproducible behaviors on the conversational AI services (§4.5). All the experiments were conducted in Spain during May 2026.

4.1 Conversational AI Service Selection

We analyze a set of prominent conversational AI services supporting both web-based and Android-based clients. Rather than maximizing breadth, we conduct an in-depth and systematic analysis on a small set of representative AI services that account for a substantial share of the market. Due to the lack of reliable market share figures per provider, we select those with large user bases using objective popularity proxies, including Tranco rankings [35] for web services and cumulative Google Play installation counts for mobile.

Table 1 summarizes the nine services we include in our study together with their providers, web and mobile implementations, and the popularity indicators that guided our selection. All these

Table 1: Web and mobile conversational AI services analyzed in this study ordered by Tranco rank domain (May 2026).

Service	Provider	Web Domain	Android Package	Tranco Rank	Play Store Installs
ChatGPT	OpenAI	chatgpt.com	com.openai.chatgpt	48	>1B
Claude	Anthropic	claude.ai	com.anthropic.claude	617	>10M
Grok	xAI	grok.com	ai.x.grok	956	>100M
DeepSeek	DeepSeek	chat.deepseek.com	com.deepseek.chat	1,196	>50M
Perplexity	Perplexity AI	perplexity.ai	ai.perplexity.app.android	1,249	>100M
Gemini	Google	gemini.google.com	com.google.android.apps.bard	6,045	>1B
MS Copilot	Microsoft	copilot.com	com.microsoft.copilot	11,011	>50M
Mistral (Le Chat)	Mistral AI	chat.mistral.ai	ai.mistral.chat	12,039	>1M
Meta AI	Meta	meta.ai	com.facebook.stella	13,701	>50M

services remain in the top-14K services for Tranco[2] in May 2026. Three of them are in the Top-1K: ChatGPT (top-48), Claude (top-617), and Grok (top-956). On the mobile side, all their mobile apps have at least 1M cumulative installs, and two of them (Gemini and ChatGPT) have more than 1B cumulative installs.

4.2 Website Analysis

We analyze the presence of trackers and their data collection practices on web-based conversational AI services using Google Chrome (v148.0.7778.167) Developer Tools (CDP), and store the resulting HAR traces for subsequent analysis. This setup enables observation of HTTP(S) requests, JavaScript execution, browser storage access, cookies, tracking pixels, and other client-side tracking mechanisms that are generated during user interactions.

We then inspect the communication to endpoints including request parameters, request bodies, cookies, browser storage entries, and protocol metadata to identify the transmission of (i) user identifiers (e.g., account IDs and email addresses); and (ii) conversation-specific metadata, conversation content, and identifiers (e.g., chat identifiers and sharing links). We also search for transformed representations of such data, including Base64 encodings and common hashing algorithms used to generate hashed email addresses (SHA-256, SHA-1, and MD5). We inspect browser storage mechanisms and monitor stable IDs across sessions to identify persistent identifiers and tracking artifacts. Finally, all observed disclosures are mapped to their corresponding recipient domains and correlated with the experimental configuration that triggers them.

All sessions are conducted manually by a researcher and cover authentication, onboarding, and pre-defined conversational exchanges, as further developed in §4.5. Experiments are repeated across subscription tiers, platform configurations, and privacy conditions, as detailed in §4.4, to evaluate the impact of consent scenarios (acceptance or rejection of non-essential cookies) and subscription tiers (guest, free, and premium accounts). Pilot experiments show highly deterministic tracking behavior for every configuration, and therefore each configuration is analyzed once.

Third-party Domain Classification. Labeling of tracking domains was performed manually by a co-author with over a decade of research experience, applying conservative criteria to avoid over-reporting. We distinguish between first- and third-party domains using publicly available blocklists and tracker intelligence sources, including `uBlock Origin` [59] and `whotracks.me` [9] with the support of DNS lookups and Certificate Transparency logs [7]. Because some conversational AI providers also operate advertising, analytics, and cloud infrastructures (e.g., Google, Microsoft and Meta), considering corporate ownership alone may obscure tracking relationships. We therefore classify provider-owned advertising, analytics, and telemetry endpoints as third-party Advertising and Tracking Services (ATSes). These services may facilitate data sharing with other ad-tech actors through Real-Time Bidding requests, or even operate under separate legal entities, as in the case of xAI and X Corp. This approach provides a consistent comparison of tracking practices across all services.

4.3 Android App Analysis

We analyze conversational AI Android apps through a combination of static and dynamic techniques to maximize behavior coverage:

Static Analysis. We decompile every app's APK using Androguard [16]. We parse `AndroidManifest.xml` to enumerate declared sensitive permissions relevant to tracking (*e.g.,* `AD_ID`, `READ_PHONE_STATE`, `ACCESS_FINE_LOCATION`). We identify embedded third-party SDKs by extracting package namespaces and mapping the app's package name (e.g., `com.company.app`) to its corresponding eTLD+1 (`company.com`) following prior work practices [24, 28, 64]. Packages and contacted domains whose ownership does not match the app's eTLD+1 are treated as third-party components [48, 53]. We then manually match these packages and domains using public SDK documentation and prior work mappings [28, 45], and the third-party classification method described in §4.2.

Dynamic Analysis. We execute each app on an instrumented Google Pixel 3a using an Android 12 build that transparently monitors runtime access to permission-protected APIs, file I/O operations, and all outbound network traffic; equivalent coverage is achievable using mitmproxy [10] and Frida [46]. We observe reads and writes to TLS sockets at the system level, enabling traffic inspection without certificate injection and without disrupting TLS handshakes, including in certificate-pinned apps [44, 47]. The instrumentation traces access to sensitive resources including device IDs (AAID, Android ID, IMEI, GSF ID, Boot ID), hardware IDs (WiFi MAC address), network scan data (WiFi SSIDs, BSSIDs), and account-linked IDs (*e.g.,* email address). [3] To complement static

[2] https://tranco-list.eu/list/3Q25L/1000000

[3] Each device is provisioned with pseudonymous IDs (email address, phone number) to register test accounts on each platform. Because ID values are known per device,

SDK detection, we instrument the Android Runtime to log classes loaded at runtime by tracking the `FindClass` method of the class linker, enabling identification of obfuscated SDK components that static analysis may miss. Captured traffic is automatically decoded for common encodings (gzip, Base64) and SDK-specific obfuscation methods, and parsed to extract field names, values, and destination endpoints using the third-party endpoint classification method described in §4.2. Sessions are screen-recorded to support post-hoc verification of data flows.

4.4 Consent Forms and Subscription Tiers

We conduct controlled experiments across different consent choices and subscription tiers, allowing us to assess how these factors shape the observed privacy risks. Our preliminary experiments revealed differences between web and mobile consent mechanisms. While web services generally present cookie consent banners, most Android apps do not expose an equivalent consent mechanism at launch. We therefore evaluate the consent conditions and privacy settings supported by each client.

For web clients, we perform separate runs in which we explicitly accept or reject non-essential cookies through the consent mechanisms offered by each service. These interactions are performed manually rather than through automated means to accurately capture service-specific consent flows and privacy controls. Each experiment starts from a fresh browser profile, with all local state removed to prevent previously stored identifiers, consent choices, authentication tokens, or telemetry artifacts from influencing subsequent measurements.

We additionally conduct controlled interactions across guest, free, and paid subscription tiers when supported by each service. For authenticated tiers, we maintain separate accounts and execute equivalent interaction scenarios across configurations. Meta AI and DeepSeek do not offer premium tiers, and neither provides a guest tier. For guest and paid configurations, we accept all cookies to capture flows to third-party organizations and isolate differences associated with authentication and subscription modes.

4.5 Conversation and Prompt Inputs

To capture dynamic evidence of the dissemination of conversation-derived artifacts to third-party services alongside user identifiers, we manually interact with every service on each platform following a predefined and reproducible interaction protocol. For each experimental mode described in §4.4, we conduct a chat session comprising several prompts designed to trigger a broader range of behaviors. When supported by the service, we additionally generate and open a chat-sharing link in a separate browser session to assess whether overly permissive access controls expose shared conversations to third parties.

While exhaustive coverage of all interaction paths is unfeasible, the prompts were iteratively refined during a preliminary measurement campaign to capture representative behaviors. Rather than aiming for absolute completeness, we use rich, health-related content to emulate realistic users following a persona-based auditing approach (§9). Specifically, the prompts emulate a user consulting the service about a medical condition, introducing sensitive context to evaluate platform behavior under realistic and privacy-sensitive scenarios. We discuss the limitations of this targeted approach in §8.1; nevertheless, it provides a consistent baseline for comparing behaviors across conversational AI services.

we can automatically search captured traffic for direct occurrences and their common hash transformations (MD5, SHA-1, SHA-256).

Table 2: Most frequent third-party organizations across the nine conversational AI services tested. We report their presence separately for web and mobile clients, as well as across both client types (∩) and either client type (∪). Entries are sorted by the number of services in which the organization is present on both web and mobile clients. Organizations are broken down by product when possible. Legend: ◐ web only, ◑ mobile only, ● both web and mobile, and empty if not contacted.

	# Services				Per service								
Organization / product	**Web**	**Mobile**	∩	∪	ChatGPT	Claude	Copilot	DeepSeek	Gemini	Grok	Meta AI	Mistral	Perplexity
Google	**8**	**8**	**7**	**9**	●	●	●	●	◐	●	◑	●	●
Firebase	0	8	0	8	◑	◑	◑	◑		◑	◑	◑	◑
Search	8	2	2	8	◐	◐	◐	◐	◐	◐		●	●
Ads	7	0	0	7	◐	◐	◐	◐	◐	◐			◐
Tag Manager	7	0	0	7	◐	◐	◐	◐	◐	◐			◐
Accounts	5	0	0	5	◐	◐	◐					◐	◐
Sentry	**2**	**4**	**2**	**4**	◑	◑	●					●	
Datadog	**3**	**2**	**2**	**3**	●	●							◐
Intercom	**2**	**1**	**0**	**3**	◑	◐						◐	
Meta	**3**	**1**	**1**	**3**		◐				◐	●		

5 Third-party Service Analysis

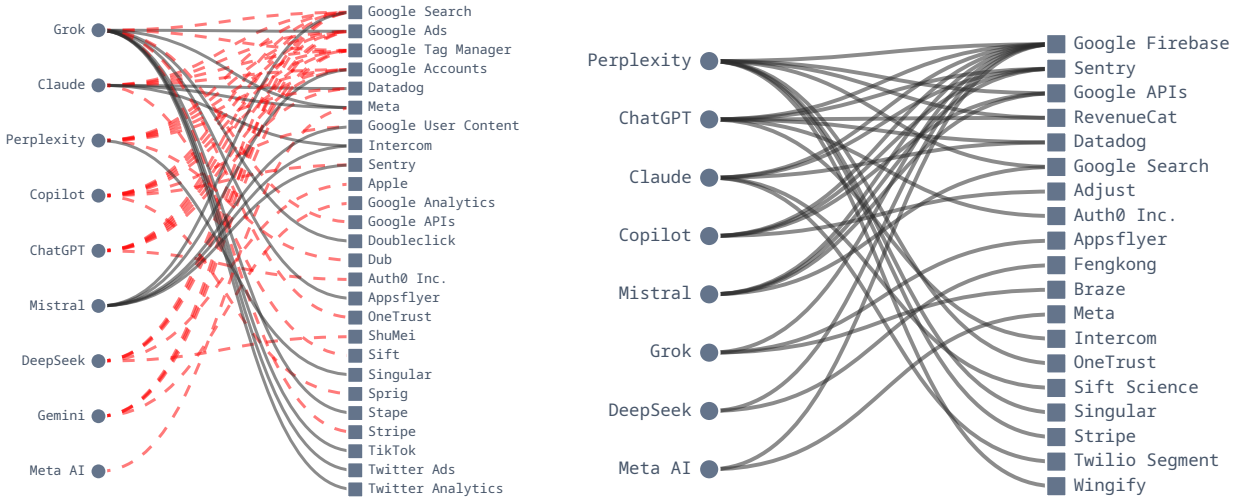
We study the structural integration of third-party services across the web and mobile clients of conversational AI services (§5.1) and how consent choices and subscription tiers influence their presence (§5.2). We then characterize the data disseminated to third-party organizations in §6.

5.1 Web vs. Mobile Tracking

Every conversational AI service contacts at least one third-party organization categorized as an ATS. Across our measurements, we observe 124 distinct third-party domains, which we attribute to 44 organizations, of which 34 are ATSes. Figure 3 captures the ATS ecosystem observed across services and client types.

However, third-party integration differs substantially between web and mobile clients. While 11 ATSes appear on both client types, 15 are observed exclusively on the web, including Google Tag Manager, TikTok, and consent-management platforms such as OneTrust. In contrast, 8 ATSes are found exclusively on mobile clients, including Braze.

Google products are the most pervasive across client types, as reported in Table 2, followed by Sentry, Meta, Datadog, and Intercom. The presence of these organizations reflects the broad reliance of conversational AI services on third-party products and services for functions including error monitoring and observability, customer support and engagement, but also advertising and analytics. Google



(a) Web client. **(b) Android client.**

Figure 3: Bipartite graph representing the connections to ATSes for web (left) and Android (right) clients. Legend: —— Interaction occurring when the user accepts non-essential cookies or ToS (Android only); - - - Interaction occurring both before and after cookie rejection (web).

has a particularly broad footprint, with products spanning advertising and analytics (e.g., Google Ads and Google Tag Manager), search, authentication (e.g., `accounts.google.com`), platform APIs (e.g., subdomains under `googleapis.com`), and mobile-specific telemetry services such as Firebase.

Our analysis also reveals the presence of lesser-known Chinese tracking services associated with DeepSeek, including Fengkong Cloud (mobile, `fp-it.fengkongcloud.com`) and ShuMei (web, `fp-it-acc.portal101.cn`). These services are rarely documented in the academic literature and are absent from popular tracker databases such as `WhoTracks.Me`. However, public reports associate them with device-fingerprinting and risk-scoring technologies [41].

In-app Browsing. 71% of all distinct endpoints contacted by mobile clients originate from WebViews (in-app browsing) rather than from native bytecode.[4] Reliance on WebViews varies substantially across apps, ranging from 98% of endpoints in Grok to 91% in Perplexity and 71% in ChatGPT. In contrast, most traffic from Copilot, Mistral, and Meta originates from native code. While WebViews facilitate the integration of web content into native apps and ease development, they also enable web-based tracking methods within mobile applications [64]. In Grok, for example, we observe content from Google Ads, Google Tag Manager, TikTok Analytics, X/Twitter Analytics, and Meta Pixel loaded within WebViews.

5.2 Impact of Consent and Subscription Tiers

Consent decisions influence the presence of trackers in web-based clients. In contrast, mobile apps require users to accept the platform's Terms of Service (ToS) and privacy policy as a prerequisite for use. We therefore evaluate the three consent scenarios described in §4.4 only on web clients: (i) ignoring the consent banner (*ignore*), (ii) rejecting non-essential cookies (*reject all*), and (iii) accepting all cookies (*accept all*).

Our results show that consent mechanisms vary substantially across providers. Some services allow users to continue interacting without making an explicit choice (e.g., Perplexity, Claude, and Grok), whereas others require a consent choice before accessing the platform (e.g., Gemini and Meta AI). Mistral requires users to fully accept its ToS and Privacy Policy before interacting with the service, as shown in Figure 4. Therefore, all the connections are labeled as *accept all*. Appendix C provides examples of observed consent forms.

Even when users reject non-essential cookies (i.e., the *reject all* scenario), we still observe connections to third parties. For example, Perplexity, DeepSeek, Gemini, Copilot, ChatGPT, and Claude connect to Google Ads under the *reject all* configuration, as shown by the dashed lines in Figure 3.

Accepting all non-essential cookies activates additional third parties in Claude, Perplexity, and Grok. These correspond primarily to well-known ATSes including Meta, TikTok, Twitter Ads, DoubleClick, and AppsFlyer. These results show that some providers conditionally activate additional advertising and tracking infrastructures following user consent, as represented by the solid lines in Figure 3. Ignoring the consent banner does not result in connections to third-party domains beyond those observed when explicitly rejecting non-essential cookies.

Subscription tiers have a comparatively limited impact on the set of third parties contacted. Free and premium accounts exhibit nearly identical tracking infrastructures across services. One exception is Claude's mobile client, where we observe Intercom and Sentry in the free tier but not in the premium tier. However, such differences may arise from dynamically activated code paths that vary across experimental runs.

[4]We attribute each outbound flow to the Android UID that generated it and classify it as originating from the assistant's native package, Custom Tabs, or a WebView process.

Table 3: Conversation artifact leakage among providers and ATSes across web and Android clients. The dissemination of shared URLs, prompts, and screenshots resulting from conversation-sharing actions is studied in §6.4.

	# of Providers		# of ATSes	
Artifact Leaked	**Web**	**Android**	**Web**	**Android**
Conversation URL	5	0	9	0
Conversation ID	2	2	1	1
User Prompt	1	1	2	1
Conversation Title	3	0	9	0
Conversation Screenshot	1	0	1	0
Shared Conversation URL	5	0	9	0
Shared Conversation ID	0	1	0	1

Conversation ID and Shared Conversation ID counts cover only cases where the identifier leaked without the full URL.

Provisioned Tracking Surface on the Web. Observed network traffic captures only the subset of third parties activated during a particular dynamic test. However, for web-based clients, the Content-Security-Policy (CSP) response headers emitted by each service (e.g., `script-src`, `connect-src`, `img-src`, and `frame-src`) enumerate other external origins that a page is authorized to contact or load resources from, thereby revealing potential relationships that may remain dormant or untriggered during a test, for example due to regional differences or execution contexts. The most prevalent third-party services included in CSP headers belong to Google Tag Manager (`googletagmanager.com`), Google Analytics (`google-analytics.com`), and Google Ads (`googleadservices.com` and `doubleclick.net`). Yet, as Table 8 shows, CSP policies commonly include other prominent actors in the advertising industry, such as TikTok and Meta.

6 Privacy Analysis

Having characterized the third-party ecosystem in §5, we now study the dissemination of sensitive information from the web and mobile clients to third parties.

6.1 Dissemination of Conversation Artifacts

Unlike traditional web tracking, conversational AI platforms may leak artifacts that encode the subject matter of user conversations. permalinks, prompts, conversation titles and screenshots. Table 3 summarizes the number of third-party leaks per artifact. Overall, 6 web and 3 mobile conversational AI clients leak artifacts to 11 and 2 ATSes, respectively during regular user interactions. This includes services like Google Ads, TikTokinteractions, and Twitter Analytics as detailed in Table 4.

Conversation URLs. Conversational AI services generate persistent conversation URLs, commonly referred to as *permalinks*, that uniquely identify individual conversations. Conversation permalinks are stable URLs used to retrieve and manage conversation history, directly tied to a specific conversation.[5] Overall, we observe that 5 web clients disclose conversation URLs or their associated global conversation identifiers to 9 ATSes. Among these, 60% (3/5) of web clients perform such disclosures by default, whereas the remaining ones only do so after users explicitly accept non-essential cookies. The web clients of ChatGPT and Claude additionally leak the globally unique conversation ID as an independent parameter to Datadog. Although less explicit than a permalink, the conversation ID allows the reconstruction of the public URL. We do not observe the conversation URL disclosure on mobile clients.

Conversation Title. Many providers automatically generate short conversation titles that summarize conversations' content or purpose as exemplified in Table 7 in the Appendix. These titles are AI-generated summaries that concisely encode semantic information about the underlying conversation theme, purpose or intent, hence revealing sensitive interests, intentions, health concerns, financial situations, professional activities, or other personal topics discussed by the end user with the AI system. Across the evaluated services, 33.3% (3/9) of web clients leak conversation titles to 9 third parties, including Meta, TikTok, and Doubleclick. Among these, 88.9% (8/9) of them only occur when users accept non-essential cookies. This disclosure is not observed on mobile versions.

Web vs. Mobile. Many third parties appear simultaneously across both web and mobile assistants (§5), but the information they collect differs substantially. Web trackers execute within the page's JavaScript context and can directly observe and record browser and application state, including chat URLs, page titles, conversation identifiers, and tracking cookies. In contrast, mobile clients collect conversation-related artifacts such as conversation ID, but the user-facing conversation artifacts like the titles common on the web versions are absent.

Impact of Consent and Subscription Tiers. Rejecting non-essential cookies can reduce information dissemination to third parties, as shown in Table 4. For example, rejecting non-essential cookies in Claude prevented the activation of Meta Pixel, Datadog telemetry, and server-side forwarding to eleven advertising platforms. However, across all analyzed free-tier services, third-party trackers still collect data in 44.4% (4/9) of services even when non-essential cookies are rejected. We do not observe clear differences between free and premium account tiers regarding the data collection practices of third parties.

6.2 Linking Conversations to User Identities

The disclosure of conversational artifacts becomes substantially more privacy-invasive when transmitted alongside user or device identifiers, such as hashed email addresses [25] or resettable advertising identifiers such as the Android Advertising ID (AAID) [28, 64], which enable third parties to associate those interactions with individual users or long-term cross-platform behavioral profiles [62]. We therefore analyze the extent to which conversational AI platforms transmit conversational artifacts alongside identifiers commonly used for advertising attribution, audience measurement, analytics, and cross-platform tracking by the industry.

[5] For example, in Grok's conversation URL: https://grok.com/c/0b3cc700-5a98-40f0-8e39-7311df76a700?rid=7b3e3eb2-0930-4490-b0ec-9f52fb3839c4, the path segment and parameter uniquely identify the conversation. However, this URL is by default publicly readable for any actor knowing it, as we further develop in §6.3.

Table 4: Simultaneous dissemination of conversation artifacts and user identifiers (e.g., emails, hashed emails, and cookies) to third parties. Results are shown by platform, interaction type, consent state, and account tier.

			Conversation Artifacts		User Identity		Account Tier		
LLM Service	**Platform**	**Tracker**	**Permalinks**	**Content**	**User PII**	**Cookies / SessionId**	**Guest**	**Free**	**Premium**
ChatGPT		Datadog	ConvId	–	UserId, DeviceId	SessionId	×	×	×
Claude		Datadog	ConvId , ShareId	–	AnonId, UserId	–	–	×	×
Perplexity		Wingify	–	Prompt	–	–	×	–	–
ChatGPT		Datadog	ConvId, ChatUrl	–	UserId, AnonId	SessionId	◐	◐	◐
Claude		Datadog	ConvId	–	UserId, AnonId, OrganizationId	SessionId	–	●	●
Claude		Intercom	ConvId, ChatUrl	–	UserId, Email, OrganizationId, User Hash	intercom-device-id	–	◐	◐
Claude		Datadog	ShareId, ShareUrl	–	AnonId	SessionId	●	–	–
Gemini		Google Analytics	–	Title	–	cid	–	◐	◐
Gemini		Google Search	ShareId, ShareUrl	–	–	–	◐	–	–
Grok		DoubleClick	ConvId, ChatUrl	Title	–	em	–	●	●
Grok		Google Ads	ConvId, ChatUrl	Title	–	em	–	●	●
Grok		Google Search	ConvId, ChatUrl	Title	–	em	–	●	●
Grok		Google Tag Manager	ConvId, ChatUrl	Title	–	_fbp, _ttp, _dcid	–	●	●
Grok		Meta	ConvId, ChatUrl	Title	–	_fbp	–	●	●
Grok		TikTok	ConvId, ChatUrl	Title	–	_ttp	–	●	●
Grok		Twitter Analytics	ConvId, ChatUrl	Title	–	_twpid	–	●	●
Grok		Appsflyer	ShareId, ShareUrl	–	–	–	●	–	–
Grok		DoubleClick	ShareId, ShareUrl	Title	–	–	●	–	–
Grok		Google Search	ShareId, ShareUrl	Title	–	–	●	–	–
Grok		Google Tag Manager	ShareId, ShareUrl	Title	–	_dcid	●	–	–
Grok		Meta	ShareId, ShareUrl	Title, Prompt	–	_fbp	●	–	–
Grok		TikTok	ShareId, ShareUrl	Title, Prompt, Screenshot	–	_ttp	●	–	–
Grok		Twitter Analytics	ShareId, ShareUrl	Title	–	_twpid	●	–	–
Mistral		Intercom	ConvId, ChatUrl	Title	UserId, Email, OrganizationId, User Hash	intercom-device-id, intercom-id	–	●	●
Mistral		Intercom	ShareId, ShareUrl	Title	–	intercom-device-id, intercom-id	●	–	–
Perplexity		Datadog	ConvId, ChatUrl	–	UserId, AnonId, Email	SessionId	◐	◐	◐
Perplexity		Datadog	ShareId, ShareUrl	–	AnonId	–	◐	–	–

Legend: Browser; Android; Disclosure during a normal conversation; Disclosure when accessing a shared conversation; Disclosure also in incognito mode; Disclosure only after accepting non-essential cookies; Disclosure under both accept and reject cookie settings; × No cookie-consent mechanism available. Prompt and Screenshot refer to the most recent interaction in the conversation. On mobile platforms, disclosures occur independently of cookie-consent settings.

Table 5: Number of ATSes collecting conversation artifacts along user and device IDs, including tracking pixels. For web products, we also report the fraction that appear only after accepting non-essential cookies.

Product	Platform	# ATSes	% Accept-All
ChatGPT	Web	1	0%
	Android	1	–
Claude	Web	2	50%
	Android	1	–
Gemini	Web	1	0%
Grok	Web	7	100%
Mistral	Web	1	100%
Perplexity	Web	1	0%

6.2.1 Web Tracking Identifiers. Beyond the disclosure of conversational artifacts, we examine tracking mechanisms through which conversational AI services expose web identifiers that can enable third parties to recognize or link users across interactions and browsing contexts. Overall, 77.3% of observed transmissions of web identifiers to third parties occur only after the user accepts non-essential cookies. The simultaneous collection of multiple identifiers by third-party trackers enables them to bridge otherwise distinct user identities and interactions under a common profile, even across platforms. We observe the following tracking techniques:

- **Web cookies and Tracking Pixels.** Cookies and impression pixels remain the dominant web tracking mechanism. We observe third-party cookies in 4 conversational AI services from 9 third parties, including Meta, TikTok, Twitter Analytics, and Intercom. Of those, 8 distinct cookies (e.g., `_fbp`, `_ttp`, `_twpid`) are transmitted alongside conversational artifacts. In the case of Grok Web, Meta Pixel alone collects the conversation ID, the chat title, the full conversation URL, the share ID, and the share URL, each linked to the user's Meta identity by the synced `_fbp` cookie. Meta documents the use of Pixel events for audience measurement and attribution, potentially aggregating them to users' Meta accounts when one exists [38, 62].
- **Email Hashes.** Although often presented as privacy-preserving representations, hashed email addresses (HEMs) act as stable ID because third parties can independently compute identical hashes for known email addresses and use them for identity resolution, as pointed out by the USA Federal Trade Commission [25, 49]. Our results show that Perplexity transmits user's HEMs to the marketing and analytics company Singular.
- **Service Usernames and Account IDs.** Four providers additionally transmit users' usernames or account IDs on the AI services alongside conversational artifacts. Additionally, three providers share stable hash-based IDs to third parties like Intercom. Although we cannot determine the exact inputs used to generate these values, they remain consistent across sessions for the same account and therefore appear to function as persistent pseudonymous IDs. For example, Mistral and Claude transmit such IDs to Intercom, while Grok sends a similar one to DoubleClick, Google Ads, and Google Search. These IDs may facilitate user identification and linkage across sessions even when direct account IDs are not disclosed.
- **Cookie Syncing.** We observe instances of cookie syncing [1] and server-side tracking [26] only when users accept non-essential cookies. Claude uses Segment Analytics proxied through a first-party domain (a-cdn.anthropic.com), loading a Conversion API configuration that forwards user events server-to-server from its infrastructure to eleven trackers (e.g., Facebook, LinkedIn, TikTok, Reddit, and Google Enhanced Conversions). This forwarding evades ad blockers and carries two shared IDs per event, which could enable data bridging across advertising platforms to

a single session. Similarly, Grok embeds a Google Tag Manager container configured to route events through a server-side GTM instance (sGTM). A custom event transmits the conversation URL and chat topic title server-to-server to Meta Conversions API and TikTok Events API through the sGTM container, invisible to browsers and unblockable by ad blockers. The same payload carries both Meta's `_fbp` and TikTok's `_ttp` cookies, enabling ID bridging between Facebook and TikTok under a single identity.

- **JS Fingerprinting Indicators.** While fingerprinting falls outside the primary scope of this work, we perform an exploratory analysis of the JavaScript resources loaded on the browser during our experiments. Following prior work findings [1, 19, 30], we search for APIs commonly associated with browser fingerprinting across most providers. We find such APIs invoked in scripts served by all providers. Although the presence of these APIs alone does not establish active fingerprinting practices, it indicates that providers possess the technical capability to potentially derive high-entropy device characteristics. This observation is consistent with recent analyses of DeepSeek, which have reported the use of fingerprinting methods for tracking and attribution purposes via Fengkong Cloud domains reported in §5 [41].

6.2.2 Mobile Implementations. Mobile apps can access and expose a broader range of identifiers and sensitive data than those typically available in browser environments, enabling stronger forms of attribution and cross-platform user recognition. Because third-party SDKs run within the host app's process and inherit its privileges, the permissions granted to the app by the user can effectively be piggybacked by third parties [24, 27, 28, 50]. Across the evaluated conversational AI services, we observe the following dissemination to third-party services:

- **Account-linked Identifiers and Email Addresses.** Four mobile clients disclose account-linked identifiers to third parties. Two transmit email addresses: Perplexity to RevenueCat, and Grok to TikTok and Twitter Analytics. Additionally, Grok transmits HEMs to both TikTok and Twitter Analytics.
- **Mobile Advertising Identifiers (MAIDs).** Three mobile clients transmit the Android Advertising ID (AAID) to third parties. We find 3 distinct third parties receiving AAIDs: Adjust (Copilot), AppsFlyer (Grok), and Singular (Perplexity). While AAIDs are designed to support advertising attribution and audience measurement, we observe cases where they are transmitted alongside other persistent identifiers. In particular, Grok shares AAIDs with AppsFlyer together with user IDs, while Perplexity shares AAIDs with Singular alongside user and installation IDs [57]. Such combinations allow ATSes to associate a resettable advertising identifier with persistent account- or installation-linked identifiers, thereby defeating the privacy properties of resettable IDs [28, 40, 53]. Even after users reset the AAID, install- and account-level identifiers persist and can be used to re-bind the newly assigned AAID on subsequent events.
- **SDK-Specific IDs.** Adjust, Auth0, Braze, Datadog, RevenueCat, Sentry, Sift Science, and Singular generate their own proprietary IDs for user recognition, session management, attribution, and analytics. In most cases, these SDK-generated IDs are transmitted independently of conversational artifacts. However, Claude forwards conversation IDs to Datadog alongside Datadog-generated and account-linked IDs, creating a direct link between conversational activity and multiple proprietary analytics IDs.
- **Other.** Claude transmits both geolocation coordinates and the Android ID (SSAID) [18] to Sift Science, while Meta AI transmits the Android ID to its own servers. While these attributes are often collected for analytics, security, or fraud-prevention purposes, they can increase the persistence and uniqueness of user identification and re-identification when combined with other IDs. We additionally observe web-style tracking IDs on the mobile side: Grok's WebView-based integrations transmit Meta's `_fbp` and TikTok's `_ttp` cookies to the same advertising and analytics endpoints observed on the web clients. However, unlike the web client, we do not observe conversation URLs or other artifacts transmitted alongside these cookies.

Table 6: Default website permalink access control across providers and account tiers.

✘ Public without Opt-out ● Public with Opt-out ◐ Owner-only ○ Unsupported
Extra controls: ★ *Whitelist* 😷 Incognito

Provider	Guest	Free	Paid	Controls
ChatGPT	◐	◐ 😷	◐ 😷	
Claude	○	◐ 😷	◐ 😷	
Grok	○	● 😷	● 😷	
Deepseek	○	◐	○	
Perplexity	✘	◐ 😷	◐ 😷	★
Gemini	◐	◐ 😷	◐ 😷	
Copilot	◐	◐ 😷	◐ 😷	
Mistral	◐	◐ 😷	◐ 😷	
Meta AI	○	◐	○	

While all tested services include the ability to share chats with everyone and afterward revoke this access, some providers expose additional access control mechanisms, such as the ability to whitelist specific users (★). We further annotate tiers with 😷 when the provider offers incognito chats (not saved in the account's history).

Together, these data dissemination practices undermine the privacy guarantees that resettable IDs are intended to provide on mobile platforms. A single Meta Pixel call carrying a conversation URL alongside a `_fbp` cookie can bind a chat to a Meta profile; hashed emails shared with TikTok or Singular can be matched against known addresses [25]; and AAIDs transmitted alongside persistent installation IDs can survive user resets. As a result, conversational activity reaches third parties already linked to long-term and cross-platform user profiles.

6.3 Accessibility of Conversation Permalinks

The dissemination of persistent conversation URLs or conversation IDs to third parties becomes particularly concerning when the referenced resources are accessible without access controls. Unlike traditional web tracking metadata, conversation permalinks can directly expose user interactions, including prompts, generated responses, interests, concerns, and other contextual information, potentially on a persistent basis, as discussed in §3.

To evaluate this risk, we systematically test permalink accessibility by opening each conversation URL in a separate private browsing session without authentication. Table 6 summarizes the access-control mechanisms observed across providers and account tiers. While some providers restrict conversations to their owners unless explicitly shared—an aspect that will be studied in §6.4—others expose them through publicly accessible URLs by default or rely on opt-out mechanisms to restrict access.

Notably, Grok takes the most permissive stance: free and premium tier conversation permalinks are accessible by default with the option to opt-out (●). Perplexity always makes guest-tier conversations public (✘). However, on April 3rd, 2026, they stopped sharing the URL with third-party trackers like Meta, possibly in response to a USA class action [33]. No significant differences in permalink accessibility were observed between web and mobile implementations across all conversational AI providers.

We do not attempt to determine the rationale behind these design decisions. However, to the extent that public accessibility is intended to simplify sharing or improve usability, our findings suggest that these choices unnecessarily increase users' exposure to privacy risks by expanding the set of entities capable of accessing conversational content.

6.4 Conversation Sharing

Conversation-sharing functionality further amplifies accessibility risks—with all services providing mechanisms to generate publicly accessible conversation URLs (see in Table 4). We observe that all conversational AI services expose shared conversations without requiring authentication when users exercise this functionality. Here, the presence of 9 third parties within conversation-sharing pages gives them full visibility into these conversations.

As part of Grok's conversation-sharing functionality, we observe TikTok collecting screenshots of conversations, while Meta and TikTok both collect the user's latest prompt. These transmissions expose verbatim portions of the conversation to third parties, as illustrated in Figure 11 in the Appendix.

6.5 Conversation Access Evidence

To investigate whether disseminated conversation resources are subsequently accessed once uploaded, we deploy canary URLs embedded in conversations (by directly sharing the URL in the prompt) and in uploaded files (by embedding the URL in uploaded .docx and PDF files). We also test a variant in which the prompt explicitly instructs the service to refrain from accessing URLs. We use default canarytokens.org [5], but also deployed it on a self-hosted canary infrastructure [58] with a generic domain name to reduce the likelihood that the URL is identified as a canary.[6]

We observe evidence of external access only for Grok, where we record 70 different canary activations over periods ranging from hours to days after their initial submission. These activations originate from 70 IP addresses spanning 48 ASes (Autonomous Systems) across 14 countries and 4 continents. While the conversations were performed in the EU, 65.7% of the connections are triggered from machines located in the USA.

In contrast to Grok's repeated post-interaction activations, access by the remaining services was generally limited to the initial interaction. For DeepSeek, Copilot, Mistral, and Claude, some canary URLs were activated once at the time of submission, with requests originating from cloud providers such as Amazon Web Services (AWS) and Google Cloud Platform (GCP). Perplexity repeatedly accessed canary URLs—even when explicitly instructed not to do so—from a range of IP addresses associated with its Perplexity-User crawler and hosted on AWS infrastructure in the USA.

However, the absence of observed accesses should not be interpreted as evidence that no access occurs, as such observations ultimately depend on the visibility provided by our instrumentation and the specific mechanisms used by downstream systems to access such resources. In fact, it is possible that other services enable server-side access to users' conversations, but we do not have visibility into this practice as discussed in §8. Yet, the repeated retrievals observed for Grok demonstrate that conversation-derived resources may remain subject to subsequent automated access long after the initial interaction.

7 Legal Context

This section discusses how the observed tracking, data-sharing, and conversation-accessibility mechanisms reported in §5 and §6 may be situated within the regulatory framework established by the ePrivacy Directive and the GDPR, which govern the use of tracking technologies and the processing of personal data in the European Union. Our objective is not to provide a definitive legal assessment or determine compliance, as such determinations require access to internal documentation, contractual arrangements, technical implementations, and organizational measures that are not publicly available. Rather, we identify legal provisions and regulatory guidance that appear relevant to the practices observed in our study and discuss potential areas of tension that may warrant further scrutiny by providers, regulators, and researchers.

7.1 ePrivacy Directive

According to Article 5(3) of the ePrivacy Directive ("**ePD**") [22], the use of cookies for analytical, advertising, or profiling purposes —a category that unequivocally includes Meta and TikTok tracking pixels or Google Analytics cookies— requires prior informed user consent. The EDPB has clarified in detail how the prohibition under Article 5(3) ePD equally extends to more modern and sophisticated technologies than cookies [21], such as (i) URL and pixel tracking, (ii) collecting unique and persistent IDs, (iii) instructing the browser (through client-side code) to send such information to a server-side API. Additionally, compliance with applicable consent requirements includes: prohibition of pre-ticked boxes, prohibition of cookie walls except for an equivalent alternative, parity between accept and reject controls, and prohibition of dark patterns leading the user to an unreflective or manipulated acceptance.

7.2 GDPR Regulation (EU) 2016/679

When AI service providers disclose chat information to third parties, both (i) the disclosure and (ii) any subsequent access by such third

[6] This decision is informed by preliminary experiments. Some LLMs identified and declined to follow URLs associated with known canary token domains. A self-hosted deployment with a generic domain name was used to avoid this.

parties constitute processing activities requiring their own specific transparent information and legal grounds, provided that they are not strictly necessary for the performance of the contract between the provider and the user, and that they exceed the user's legitimate expectations. Therefore, according to the GDPR [23], conversational AI services are legally obliged essentially:

(1) **To inform about their processing in a clear and understandable manner.** The user's legitimate expectations would hardly include the possibility that their conversations could be accessible to third parties such as Meta or Google. OpenAI, Anthropic, Perplexity and xAI use generic and ambiguous expressions to refer to the user's "user prompts" without specifically citing them when informing about third-party access: "*user content*" (OpenAI), "*conversations*" (Anthropic), "*service interaction info*" (Perplexity) and "*user info*" (xAI). Not only the EDPB, but also the CJEU in Meta Platforms Ireland (C-757/22, paragraph 54) [13] imposes literally on the data controller the obligation to inform the data subject linking for each particular processing operation (i) the breakdown of personal data processed, (ii) its purpose, and (iii) the legal basis claimed. It is significant that even when OpenAI provides this information in such a structured and linked manner, it does not reference the practices observed in this study.
(2) **To have a legal basis for processing.** The same applies to third parties accessing this data for their own purposes. This basis cannot be, as said, the "performance of the contract" between the user and the AI provider. It is highly debatable whether training the service with user interactions is "objectively necessary" for the provision of the service. We must bear in mind the constant assurances from dominant companies such as OpenAI or Anthropic regarding their policy of not training models on user interactions within the enterprise tier. What is not done with corporate users can hardly be "objectively necessary" for the same processing in the free tier.
It seems even more difficult to legitimize as "necessary for the provision of the service" the access to third parties to the title and/or URL of the chats *together with* persistent IDs and tracking pixels for purposes unrelated to the operation of the tool, using adtech-specific mechanisms. The EDPB clarified this point in its binding decision 3/2022 against Meta IE (paragraph 118) [20] by indicating that if such processing was necessary, the contract terms should provide (*mutatis mutandis*) (i) the obligation of the provider to give third parties access to information about the chats, and (ii) contractual penalties in case of non-compliance with such "obligation", elements that do not appear in the terms and conditions. This processing could only be carried out through explicit consent (as advocated by the EDPB) or legitimate interest. Even assuming the applicability of the legitimate interest, in the cases analyzed, the mandatory prior information on (i) processing and (ii) the possibility of exercising a right of opposition is conspicuously absent.
Additionally, this data processing should be covered by some exception to the prohibition of processing special data categories. It is a known and widespread practice among users (and encouraged by providers) that users make inquiries about their health, psychological state, intimate matters, etc, making the results of this study even more concerning. The longitudinal processing of these conversations over a significant period allows the inference of all kinds of special data categories even if not specifically protected (economic capacity, appetite for financial risk, vulnerability to compulsive purchasing).

Doctrine "SRB / Scania." The *"silver bullet"* used by online services in this type of scenario is typically one of the following: (a) **anonymization** or (b) **pseudonymization** under conditions that prevent the data recipient from re-identifying users without disproportionate efforts in terms of time, material resources, or personnel (SRB/Scania doctrine). The first thing to clarify is that these arguments would only serve (assuming hypothetically that they applied) to remedy the lack of a legal basis: the obligation to inform users about third-party access in the terms imposed by the CJEU in "EDPS vs SRB" (C-413/23 P, paragraph 110) [14] could have been breached.

(1) **Anonymization**. The robustness and effectiveness of any anonymization process can only be assessed on a case-by-case basis, and it would be necessary to see the actual definition of anonymization and whether what has been done is sufficient.
(2) **Pseudonymization / SRB-Scania doctrine**. It is highly unlikely that large platforms like Meta or Google could benefit from this doctrine, given their vast personal databases and their business model, which consists precisely and unequivocally of combining different data sources to better personalize targeted advertising. It should not be forgotten that the "insignificant risk of re-identification of the data subjects" must be assessed in light of the "context, purpose, and effects" CJEU Gesamtverband Scania (C-319/22, paragraph 70) [12] of the processing carried out by the party accessing the pseudonymized data.

8 Discussion

Our findings reveal that conversational AI services expose substantially richer and sensitive information than conventional web and mobile services to tracking firms. Beyond collecting persistent IDs, several providers disseminate conversation-derived artifacts that encode the content, topic, or context of user interactions to third-party tracking services. When combined with stable IDs and permissive access-control mechanisms, these disclosures create new opportunities for third parties to associate sensitive conversational data with long-term user profiles and, in some cases, directly access the referenced conversation content. Therefore, our study raises broader questions about the exposure of conversational data, the effectiveness of existing privacy controls, and the accountability mechanisms governing these disclosures.

Conversation artifact leakage to the tracking ecosystem. Our results show that conversational AI has inherited the advertising and analytics infrastructure of the web and mobile platforms, bringing conversation data into existing tracking pipelines. Across 6 of 9 web services and 3 of 8 mobile apps, third parties receive conversation artifacts like prompts, titles, and conversation permalinks (§6.1). These disclosures often occur alongside persistent IDs, including synced advertising cookies and HEMs (§6.2), making conversations linkable to long-term and cross-platform user profiles. Beyond the privacy implications, several third-party organizations, including Google and Meta, are embedded in competing AI products, raising

questions about the collection of conversational data, including model responses, by direct competitors.

Privacy controls offer limited protection. The privacy controls available to users have limited impact on the underlying data flows. Even when non-essential cookies are rejected, 80.8% of third-party trackers remain active. Subscription tier is similarly ineffective, with free and premium accounts interacting with largely identical third parties. More fundamentally, the strongest controls are also least accessible to the users most exposed to tracking. On the web, only ChatGPT's paid tier supports persistent consent preferences, yet roughly 95% of ChatGPT users remain on free tiers [51], and OpenAI has announced advertising for free and low-cost users [52]. On mobile, no equivalent to web cookie-consent mechanisms exist despite exposing similar tracking infrastructure as discussed in §5.

Scope and generalizability. While our evaluation focuses on leading consumer-facing services, the leakage vectors we observe stem from embedding standard web and mobile trackers directly into conversational interfaces. Consequently, these concerns extend beyond consumer B2C services to the broader ecosystem of LLM-powered web applications, custom customer-support chatbots, and third-party AI wrappers that rely on identical web and mobile analytics pipelines.

Exposure translates into access and liability. Beyond transmission, we also find evidence that the disclosed content is subsequently retrieved. Our canary-token experiments indicate that at least some shared conversation artifacts are subsequently accessed. We observe repeated retrievals of conversations shared through Grok from infrastructure distributed across 14 countries, with 65.7% of accesses originating in the United States despite the conversations being conducted in the EU (§6.5), adding data-sovereignty considerations to the underlying privacy concerns. However, the significance of the exposure does not depend on whether subsequent access is directly observed. Conversational metadata reaches trackers through structured URL parameters and permissive sharing defaults (§6.3), reflecting deliberate design rather than incidental byproducts of page rendering. As the CJEU established in *Fashion ID* [11] (C-40/17, paragraph 78), enabling third-party access is itself a processing decision that triggers the provider's obligations, regardless of whether the data is ultimately read (§7). The exposure, and the provider's accountability for it, therefore arise from the design itself. Tooling and methodological constraints and research challenges, discussed next, limit our full visibility into this behavior and other server-side accesses. Addressing the challenges we faced is non-trivial, and we hope this work motivates the research community to tackle them.

8.1 Limitations

Our study covers nine conversational AI services, a subset of a broader, rapidly growing ecosystem. While selective in breadth, we focus on services with large active user bases. We exclude enterprise and governmental tiers, which providers market with distinct data-handling commitments whose validation we leave to future work. Our findings coverage is constrained to the limitations of black-box analysis techniques, including susceptibility to dynamic-analysis evasion, and we were unable to extract traces for Gemini's mobile app. Furthermore, our evaluation focuses on discrete interaction sessions and does not capture cumulative privacy risks unique to long-lived conversations featuring persistent multi-turn memory. Nor does it cover unstudied scenarios such as native desktop clients, voice-first interfaces, or domain-specific vertical AI tools.

The presence of a third party does not by itself imply that data is transmitted under all conditions only for advertising or tracking purposes, as some SDKs offer support for operational functions such as user attribution or in-app payments. Establishing processing purpose requires legal and contractual analysis beyond the scope of an empirical study. Large providers can also track users through their own first-party domains and server-side tracking, consistent with disclosures in their privacy policies, and our methodology cannot distinguish such tracking from functional traffic. We therefore report the presence of well-known first-party-tracking enablers such as Google Tag Manager. Finally, we measure from a fixed set of vantage points and user profiles, and practices in other geographies, account states, or user contexts may differ. Moreover, as the ecosystem is rapidly evolving and subject to frequent updates, our findings represent a point-in-time lower bound on the extent of third-party integration and PII leakage in these services.

8.2 Mitigation

To mitigate some of these risks, both web browsers and mobile operating systems provide users with privacy and tracking controls. Privacy-enhancing browsers such as Brave block ads, analytics, and tracking resources at the network level by default [60]. However, these could be defeated by advanced server-side and first-party tracking techniques. Similarly, mobile operating systems implement permission models that control app access to sensitive resources such as location, contacts, microphones, cameras, and MAIDs. Platforms such as iOS and Android also provide mechanisms to limit ad personalization and tracking through features such as App Tracking Transparency (ATT), advertising ID reset capabilities, and privacy dashboards that expose portions of application data access behavior.

Building on these foundations, recent research explores automated permission management to give users meaningful control over agent data access [68], an approach that consumer AI services could adopt to extend existing user controls over third-party trackers. A complementary direction may enforce website-provided policies through targeted sandboxing of browser-using agents [37], complementing user-facing permission controls with system-level confinement. Providers are also better positioned to act, for instance by not exposing conversation permalinks publicly by default and by withholding automatically generated titles and permalinks from third parties. Some providers have begun to expose dedicated controls. OpenAI, for example, recently introduced marketing-cookie controls [65], although the same update also permits using cookies to promote its products on other sites. From a user perspective, available mitigations include rejecting non-essential cookies and disabling chat sharing where platforms offer such controls. Nevertheless, these measures remain partial and platform-dependent.

9 Related Work

A growing body of research focused on auditing the security of LLMs and their agentic extensions, uncovering vulnerabilities across

model training, information flows, tool integrations, and interactions with external services [8, 36, 56, 66, 69]. Our study focuses on privacy risks of generative AI in user-facing systems and follows the dissemination of conversational data itself alongside user identifiers, including prompts, conversation titles, or chat IDs. Prior work relevant to this setting falls into three areas:

Third-party LLM Apps. An emerging app ecosystem layered atop foundation models exposes external services to the model via structured tool interfaces. OpenAI's GPTs and Anthropic's MCP-based connectors are some of the few existing LLM app platforms. Prior work characterizes LLM apps as prompt-based wrappers combined with custom data-processing API integrations [31], examining their privacy posture without focusing on third-party integration [6]. Subsequent work audits OpenAI's GPT app ecosystem, analyzing the natural-language specifications of GPT Actions to assess the data collection practices they declare [67]. Our work takes a more fundamental scope, studying the native AI layer of consumer services, which sits at the core of LLM app marketplaces.

Browser Agents and Extensions. Commercial agentic systems such as ChatGPT Atlas, Perplexity Comet, and AI-powered browser extensions like Sider and Monica act as external agents that drive the browser on behalf of their users. Prior studies enumerate the privacy risks of browser agents driving web sessions [55, 60], and audit popular GenAI browser extensions to measure data collection, processing, and sharing practices, as well as user profiling along demographic and interest attributes [61]. A complementary line of work characterizes browser and behavioral fingerprints of browsing agents [63], and evaluates same-origin policy enforcement across agentic browsers [54]. In contrast, we study assistants embedded directly in first-party services rather than installed as extensions, reaching a far broader user base by default. For instance, ChatGPT exceeds 1B installs just on the Play Store (cf. Table 1), compared to 5M for Monica, the most popular extension studied in [61].

First-Party Website and Mobile Services. A central contribution of our work is systematically studying and characterizing data leakage from LLM services embedded in first-party websites and mobile applications, including conversational artifacts. Prior studies have examined third-party tracking in web-based AI services [32], primarily focusing on identifying trackers and characterizing their conventional data collection practices. Yet this work does not fully characterize the privacy risks arising from the generation, disclosure, and accessibility of conversation-derived information across both web and mobile AI services, covering their consent mechanisms and the impact of subscription tiers on tracker's behavior. By jointly analyzing both web and mobile clients, conversation-derived artifacts, identity-linkage mechanisms, consent configurations, subscription tiers, and access-control policies, our work provides the first comprehensive assessment of privacy risks in mainstream conversational AI services. To probe these risks we adopt persona-based auditing with sensitive attributes, an established paradigm to uncover deep ad-targeting behavior [39].

10 Conclusions

This paper presented the first systematic measurement of third-party tracking in consumer conversational AI, analyzing nine prominent assistants across web and mobile clients under different consent states and subscription tiers. We showed that the tracking infrastructure of the web and mobile ecosystems—tracking pixels, SDKs, persistent identifiers, and server-side forwarding—has been carried into conversational AI, exposing conversation artifacts such as prompts, generated titles, and permalinks to third parties with advertising-based business models. These disclosures frequently occur alongside persistent identifiers, including advertising IDs, hashed email addresses, and tracking cookies, enabling conversations to be linked to long-term user profiles. We further found that cookie consent and subscription tier provide limited protection, while permissive sharing defaults leave conversation permalinks publicly accessible. Using canary tokens, we confirmed that shared conversations are subsequently accessed from distributed third-party infrastructure. Together, these findings highlight tensions between current practices and obligations under the GDPR and ePrivacy Directive, exposing broader shortcomings in consent, access control, and transparency mechanisms and underscoring the need for stronger safeguards and accountability for conversational data flows.

Acknowledgments

This research was co-funded by RED2024-154240-T (EMACS) and by the European Union and the European Cybersecurity Competence Centre under grant agreement No. 101309318 (REAL-PETS). G. Oliveira, JM. Santa Olalla Gómez, and T. Jackevicius acknowledge the financial support received through the CYBERACTIONING project, co-funded by the European Union under the Digital Europe Programme (Grant Agreement No. 101123445), which supported this research carried out as part of the Master's Programme in International Cybersecurity and Cyberintelligence (MICAC). N. Vallina-Rodriguez and G. Suarez-Tangil have been appointed as 2019 Ramón y Cajal fellows (RYC2020-030316-I and RYC2020-029401-I, respectively) funded by MICIU/AEI/10.13039/501100011033 and the ESF Investing in your future. G. Suarez-Tangil and M. Sanchez's work was supported by a 2025 Leonardo Grant for Scientific Research and Cultural Creation from the BBVA Foundation (SHIFT, reference LEO25-1-19720). The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect those of any of the funding agencies. The BBVA Foundation accepts no responsibility for the opinions, statements, and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors.

Ethical Considerations

This research involves observing privacy risks in widely used conversational AI services, so we took care to act responsibly throughout the process. We did not exploit any vulnerability, access other users' conversations, or collect personal data from real users. All tests were carried out on accounts we owned only for research purposes, using a fixed prompt we authored ourselves, and traffic was captured locally through the browser's developer tools.

During our risk analysis, we determined that most of the issues reported in this study reflect intentional product-level practices rather than exploitable security vulnerabilities. Grok was the only service for which we identified a security concern that could be exploited by an external party, due to weak access controls on conversation resources. This distinction guided our disclosure strategy. Our broader goal is to inform users, support regulatory oversight, and encourage providers to address the identified privacy risks and practices. Accordingly, we followed a responsible disclosure process for the exploitable security issue, providing xAI with an opportunity to investigate and remediate it prior to publication. We also notified relevant regulatory authorities of the broader privacy findings. The timeline of our disclosures was as follows:

- **23 March 2026** — We first noticed tracker activity while analyzing the network traffic of Perplexity and Grok.
- **6 April 2026** — We expanded the testing to cover major platforms in a systematic way, across different login states, cookie consent choices, account tiers, and privacy modes.
- **13 April 2026** — We notified our findings to the relevant Data Protection Authorities (DPAs) in the European Union and UK.
- **17 April 2026** — We notified xAI via their vulnerability disclosure email address about Grok's conversation leaks due to their lack of access control mechanisms.
- **4 May 2026** — We released part of our findings public on the project website, LeakyLM.
- **27 May 2026** — The Spanish DPA, AEPD, cited our work requesting to elevate the investigations for a plenary EDPB meeting to be held on June 6 2026 [3].
- **15 Aug 2026** OpenAI updated ChatGPT's privacy policy, explicitly mentioning third-party trackers; but we cannot confirm whether this action was directly influenced by of our findings.
- **10th Sept 2026** — Grok still uses publicly accessible permalinks. No official response has been received to date since our responsible disclosure.

Open Science

To encourage reproducibility, we provide all research artifacts in https://github.com/guinucool/pbst2027.

AI Use

The authors used generative AI-based tools to revise the text, improve flow and clarity, correct typographical and grammatical errors, and create latex table structures. We have manually verified the integrity of this output. We do not use AI-based tools as part of our research methodology and data analysis pipeline, nor to generate figures. The literature review is of our own.

References

[1] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The web never forgets: Persistent tracking mechanisms in the wild. In *Conference on Computer and Communications Security (CCS)*.

[2] AdExchanger. 2026. Daily News Roundup: OpenAI, Advertising, and AI Commerce. https://www.adexchanger.com/daily-news-roundup/tuesday-03032026/ Accessed: 2026-05-31.

[3] AEPD. 2026. La Agencia promueve ante las autoridades europeas de protección de datos que se estudie si algunos sistemas de IA permiten a terceros acceder a las conversaciones. https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/la-agencia-promueve-ante-las-autoridades-europeas-proteccion-estudie-ia. Accessed: 2026-09-7.

[4] Google Developers Blog. 2025. Under the Hood: Universal Commerce Protocol (UCP). https://developers.googleblog.com/under-the-hood-universal-commerce-protocol-ucp/ Accessed: 2026-05-31.

[5] Canary 2026. Canarytokens. https://canarytokens.org/. Accessed: 2026-05-27.

[6] Juan-Carlos Carrillo, Jose Luis Martin-Navarro, Rongjun Ma, and Jose Such. 2026. Personal Data Flows and Privacy Policy Traceability in Third-party LLM Apps in the GPT Ecosystem. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[7] Certificate Transparency. 2026. crt.sh: Certificate Transparency Search. https://crt.sh/. Accessed: 2026-05-29.

[8] Jeffrey Yang Fan Chiang, Seungjae Lee, Jia-Bin Huang, Furong Huang, and Yizheng Chen. 2025. Why are web ai agents more vulnerable than standalone llms? a security analysis. *arXiv preprint arXiv:2502.20383* (2025).

[9] Cliqz GmbH and Ghostery GmbH. 2026. WhoTracks.Me. https://whotracks.me Accessed: 2026-05-24.

[10] Aldo Cortesi, Maximilian Hils, Thomas Kriechbaumer, and contributors. 2010–. mitmproxy: A free and open source interactive HTTPS proxy. https://mitmproxy.org/

[11] Court of Justice of the European Union. 2019. Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62017CJ0040 Case C-40/17, ECLI:EU:C:2019:629.

[12] Court of Justice of the European Union. 2024. Gesamtverband Autoteile-Handel e.V. v Scania CV AB. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62022CJ0319 Case C-319/22, ECLI:EU:C:2024:845.

[13] Court of Justice of the European Union. 2024. Meta Platforms Ireland Ltd v Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62022CJ0757 Case C-757/22, ECLI:EU:C:2024:598.

[14] Court of Justice of the European Union. 2025. European Data Protection Supervisor (EDPS) v Single Resolution Board (SRB). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62023CJ0413 Case C-413/23 P, ECLI:EU:C:2025:645.

[15] Criteo. 2025. Agentic Commerce Is Emerging, Just Not the Way Most People Expect. https://www.criteo.com/blog/agentic-commerce-is-emerging-just-not-the-way-most-people-expect/ Accessed: 2026-05-31.

[16] Anthony Desnos, Geoffroy Gueguen, and Sebastian Bachmann. 2015. Androguard: Reverse engineering, malware and goodware analysis of android applications... and more (ninja!). https://androguard.github.io/androguard/. Accessed: 2026-05-25.

[17] Yana Dimova, Gunes Acar, Lukasz Olejnik, Wouter Joosen, and Tom Van Goethem. 2021. The cname of the game: Large-scale analysis of dns-based tracking evasion. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[18] Android Official Documentation. 2025. Settings.Secure. https://developer.android.com/reference/android/provider/Settings.Secure#ANDROID_ID Accessed: 2026-05-31.

[19] Peter Eckersley. 2010. How Unique Is Your Web Browser?. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[20] European Data Protection Board. 2022. *Binding Decision 3/2022 on the dispute submitted by the Irish Supervisory Authority on Meta Platforms Ireland Limited and its Facebook service (Art. 65 GDPR)*. Technical Report. European Data Protection Board. https://www.edpb.europa.eu/system/files/2023-01/edpb_bindingdecision_202203_ie_sa_meta_facebookservice_redacted_en.pdf Adopted pursuant to Article 65 GDPR.

[21] European Data Protection Board. 2024. *Guidelines 2/2023 on Technical Scope of Art. 5(3) of ePrivacy Directive*. Technical Report. European Data Protection Board. https://www.edpb.europa.eu/system/files/2024-10/edpb_guidelines_202302_technical_scope_art_53_eprivacydirective_v2_en_0.pdf

[22] European Parliament and Council of the European Union. 2002. Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (ePrivacy Directive). Official Journal of the European Union, L 201, 31 July 2002, pp. 37–47. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02002L0058-20091219 Article 5(3), as amended by Directive 2009/136/EC.

[23] European Parliament and Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119, pp. 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[24] Álvaro Feal, Julien Gamba, Juan Tapiador, Primal Wijesekera, Joel Reardon, Serge Egelman, and Narseo Vallina-Rodriguez. 2021. Don't accept candy from strangers: An analysis of third-party mobile sdks. *Data Protection and Privacy: Data Protection and Artificial Intelligence* 13 (2021), 1.

[25] Federal Trade Commission. 2024. No, hashing still doesn't make your data anonymous. https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/07/no-hashing-still-doesnt-make-your-data-anonymous. Accessed: 2026-05-31.

[26] Imane Fouad, Cristiana Santos, and Pierre Laperdrix. 2024. The Devil is in the Details: Detection, Measurement and Lawfulness of Server-Side Tracking on the Web. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[27] Julien Gamba, Álvaro Feal, Eduardo Blazquez, Vinuri Bandara, Abbas Razaghpanah, Juan Tapiador, and Narseo Vallina-Rodriguez. 2023. Mules and permission laundering in android: Dissecting custom permissions in the wild. *IEEE Transactions on Dependable and Secure Computing* 21, 4 (2023), 1801–1816.

[28] Aniketh Girish, Joel Reardon, Juan Tapiador, Srdjan Matic, and Narseo Vallina-Rodriguez. 2025. Your Signal, Their Data: An Empirical Privacy Analysis of Wireless-scanning SDKs in Android. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[29] Alejandro Gómez-Boix, Pierre Laperdrix, and Benoit Baudry. 2018. Hiding in the crowd: an analysis of the effectiveness of browser fingerprinting at large scale. In *Proceedings of the ACM Web Conference (WWW)*.

[30] Umar Iqbal, Steven Englehardt, and Zubair Shafiq. 2021. Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors. In *IEEE Symposium on Security and Privacy (S&P)*.

[31] Umar Iqbal, Tadayoshi Kohno, and Franziska Roesner. 2024. LLM platform security: Applying a systematic evaluation framework to OpenAI's ChatGPT plugins. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*.

[32] Muhammad Jazlan, Ethan Wang, Yash Vekaria, and Zubair Shafiq. 2026. Tracking Conversations: Measuring Content and Identity Exposure on AI Chatbots. *arXiv preprint arXiv:2604.27438* (2026).

[33] John Doe. 2026. Complaint for Damages and Demand for Jury Trial: Doe v. Perplexity AI, Inc. Complaint filed in the Superior Court of California. https://cdn.arstechnica.net/wp-content/uploads/2026/04/Doe-v-Perplexity-Complaint-3-31-26.pdf Filed March 31, 2026. Available online.

[34] Pierre Laperdrix, Gildas Avoine, Benoit Baudry, and Nick Nikiforakis. 2019. Morellian analysis for browsers: Making web authentication stronger with canvas fingerprinting. In *Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*.

[35] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczynski, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Network and Distributed System Security Symposium (NDSS)*.

[36] Shen Li, Liuyi Yao, Lan Zhang, and Yaliang Li. 2025. Safety Layers in Aligned Large Language Models: The Key to LLM Security. In *International Conference on Learning Representations (ICLR)*.

[37] Luoxi Meng, Henry Feng, Ilia Shumailov, and Earlence Fernandes. 2025. cellmate: Sandboxing browser ai agents. *arXiv preprint arXiv:2512.12594* (2025).

[38] Meta for Developers. 2025. Meta Pixel Reference. https://web.archive.org/web/20250531104925/https://developers.facebook.com/docs/meta-pixel/reference/.

[39] Maaz Bin Musa and Rishab Nithyanand. 2022. ATOM: Ad-network Tomography. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[40] Trung Tin Nguyen, Michael Backes, and Ben Stock. 2022. Freely given consent? studying consent notice of third-party tracking and its violations of gdpr in android apps. In *Conference on Computer and Communications Security (CCS)*.

[41] NowSecure. 2025. NowSecure Uncovers Multiple Security and Privacy Flaws in DeepSeek iOS Mobile App. https://www.nowsecure.com/blog/2025/02/06/nowsecure-uncovers-multiple-security-and-privacy-flaws-in-deepseek-ios-mobile-app/. Accessed: 2026-05-28.

[42] OpenAI. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt Accessed: 2026-05-31.

[43] OpenAI. 2025. How people are using ChatGPT. https://openai.com/index/how-people-are-using-chatgpt. Accessed: 2026-09-8.

[44] Amogh Pradeep, Muhammad Talha Paracha, Protick Bhowmick, Ali Davanian, Abbas Razaghpanah, Taejoong Chung, Martina Lindorfer, Narseo Vallina-Rodriguez, Dave Levin, and David Choffnes. 2022. A comparative analysis of certificate pinning in Android & iOS. In *Proceedings of the Internet Measurement Conference (IMC)*.

[45] Exodus Privacy. 2024. Homepage. https://exodus-privacy.eu.org/en/. Accessed: 2026-05-31.

[46] Ole André Vadla Ravnås and contributors. 2014. Frida: Dynamic instrumentation toolkit for developers, reverse-engineers, and security researchers. https://frida.re/

[47] Abbas Razaghpanah, Arian Akhavan Niaki, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Johanna Amann, and Phillipa Gill. 2017. Studying TLS usage in Android apps. In *Conference on emerging Networking EXperiments and Technologies (CoNEXT)*.

[48] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, Phillipa Gill, et al. 2018. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In *Network and Distributed System Security Symposium (NDSS)*.

[49] Joel Reardon, Kenneth A. Bamberger, and Serge Egelman. 2024. Anonymity, Consent, and Other Noble Lies: An Empirical Study of the Data Economy. https://ibl.law.uiowa.edu/anonymity-consent-and-other-noble-lies-empirical-study-data-economy. Accessed: 2026-05-31.

[50] Joel Reardon, Álvaro Feal, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, and Serge Egelman. 2019. 50 ways to leak your data: An exploration of apps' circumvention of the android permissions system. In *Proceedings of the USENIX Security Symposium*.

[51] Reuters. 2025. OpenAI Projected at Least 220 Million People Will Pay for ChatGPT by 2030, The Information Reports. Reuters. https://www.reuters.com/technology/openai-projected-least-220-million-people-will-pay-chatgpt-by-2030-information-2025-11-26/ Accessed: 2026-05-31.

[52] Reuters. 2026. OpenAI to introduce ads to all ChatGPT free and Go users in US. https://www.reuters.com/business/media-telecom/openai-expand-ads-chatgpt-all-free-low-cost-users-information-reports-2026-03-21/

[53] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, Serge Egelman, et al. 2018. "Won't somebody think of the children?" examining COPPA compliance at scale. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[54] Franziska Roesner and David Kohlbrenner. 2026. Agentic Browsers and the Same-Origin Policy. In *International Conference on Learning Representations (ICLR)*.

[55] Jaechul Roh, Eugene Bagdasarian, Hamed Haddadi, and Ali Shahin Shamsabadi. 2026. SPILLage: Agentic Oversharing on the Web. *arXiv preprint arXiv:2602.13516* (2026).

[56] Avishag Shapira, Parth Atulbhai Gandhi, Edan Habler, and Asaf Shabtai. 2025. Mind the web: The security of web use agents. *arXiv preprint arXiv:2506.07153* (2025).

[57] Singular. [n. d.]. Android SDK: Setting a User ID. Singular Developer Documentation. https://support.singular.net/hc/en-us/articles/35636052267803-Android-SDK-Setting-a-User-ID Accessed: 2026-05-31.

[58] Thinkst Applied Research. 2026. Dockerized Canarytokens. https://github.com/thinkst/canarytokens-docker. Accessed: 2026-05-27.

[59] uBlock Origin Team. 2026. uBlock Origin. https://ublockorigin.com Accessed: 2026-05-24.

[60] Alisha Ukani, Hamed Haddadi, Ali Shahin Shamsabadi, and Peter Snyder. 2025. Privacy Practices of Browser Agents. *arXiv preprint arXiv:2512.07725* (2025).

[61] Yash Vekaria, Aurelio Loris Canino, Jonathan Levitsky, Alex Ciechonski, Patricia Callejo, Anna Maria Mandalari, and Zubair Shafiq. 2025. Big Help or Big Brother? Auditing Tracking, Profiling, and Personalization in Generative {AI} Assistants. In *Proceedings of the USENIX Security Symposium*.

[62] Tim Vlummens, Aniketh Girish, Nipuna Weerasekara, Frederik Zuiderveen Borgesius, Gunes Acar, and Narseo Vallina-Rodriguez. 2026. Bridges to Self: Silent Web-to-App Tracking on Mobile via Localhost. In *Proceedings of the USENIX Security Symposium*.

[63] Ethan Wang, Zubair Shafiq, and Yash Vekaria. 2026. FP-Agent: Fingerprinting AI Browsing Agents. *arXiv preprint arXiv:2605.01247* (2026).

[64] Nipuna Weerasekara, José Miguel Moreno, Srdjan Matic, Joel Reardon, Juan Tapiador, Narseo Vallina-Rodríguez, et al. 2025. Tracking without borders: Studying the role of webviews in bridging mobile and web tracking. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*.

[65] Wired. 2026. OpenAI Enables Cookies by Default for Free ChatGPT Users. https://www.wired.com/story/openai-enables-cookies-by-default-for-free-chatgpt-users/ Accessed: 2026-05-31.

[66] Fangzhou Wu, Ning Zhang, Somesh Jha, Patrick McDaniel, and Chaowei Xiao. 2024. A New Era in LLM Security: Exploring Security Concerns in Real-World LLM-based Systems. arXiv:2402.18649 [cs.CR] https://arxiv.org/abs/2402.18649

[67] Yuhao Wu, Evin Jaff, Ke Yang, Ning Zhang, and Umar Iqbal. 2025. An in-depth investigation of data collection in llm app ecosystems. In *Proceedings of the Internet Measurement Conference (IMC)*.

[68] Yuhao Wu, Ke Yang, Franziska Roesner, Tadayoshi Kohno, Ning Zhang, and Umar Iqbal. 2025. Towards Automating Data Access Permissions in AI Agents. arXiv:2511.17959 [cs.CR] https://arxiv.org/abs/2511.17959

[69] Zhonghao Zhan, Huichi Zhou, Zhenhao Li, Peiyuan Jing, Krinos Li, and Hamed Haddadi. 2026. How Adversarial Environments Mislead Agentic AI? *arXiv preprint arXiv:2604.18874* (2026).

A Conversation Titles

This section provides illustrative examples of how different AI services automatically synthesize user intent to generate chat session titles. As shown in Table 7, the web implementations of ChatGPT, Claude, and Grok result in different outcomes regarding brevity, capitalization, and information density when condensing user prompts.

Table 7: Examples of automatically generated conversation titles derived from concise user prompts in the web implementations of ChatGPT, Claude and Grok.

Prompt	LLM	Generated Title
What are the symptoms of early-stage Parkinson's disease?	ChatGPT	Early-stage Parkinson's Symptoms
	Claude	Early-stage Parkinson's disease symptoms
	Grok	Early Parkinson's Disease Symptoms
My salary is $85k. How much mortgage can I afford in NYC?	ChatGPT	Mortgage Affordability in NYC
	Claude	Mortgage affordability on $85k salary in NYC
	Grok	$85k NYC Salary: $280k-$350k Mortgage

B CSP Relationships

We detail the observed CSP relationships identified across the analyzed ecosystem. Specifically, Table 8 shows the number of different conversational AI providers that integrate these trackers into their CSP policies, highlighting the most prevalent third-party advertising and analytics domains.

Table 8: Observed CSP relationships.

Organization	Domain(s)	# Providers
Google Tag Manager	googletagmanager.com	6
Google Analytics	google-analytics.com	5
Google Ads	googleadservices.com	4
Google DoubleClick	doubleclick.net	3
Meta Pixel / Conversions API	connect.facebook.net, facebook.net	4
TikTok Analytics	analytics.tiktok.com	2
Reddit Ads / Analytics	pixel-config.reddit.com, redditstatic.com	2
Microsoft Bing	bat.bing.com, bing.com	2
Intercom	intercom.io, intercomcdn.com	2

C Consent Implementations

To provide context on Privacy policies and the differences in implementation between providers, Figures 4, 5, 6 show snapshots of what cookie consent banners looked like at the time of capture.

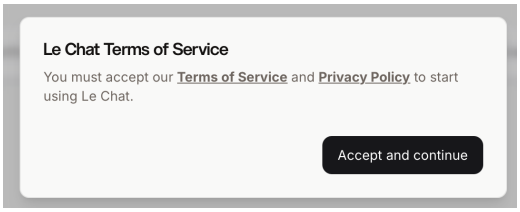


Figure 4: Mistral web consent banner.

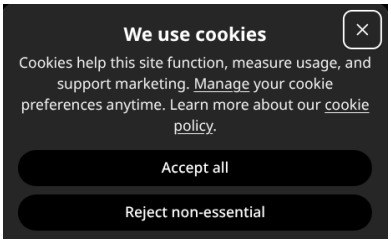


Figure 5: ChatGPT web consent banner.

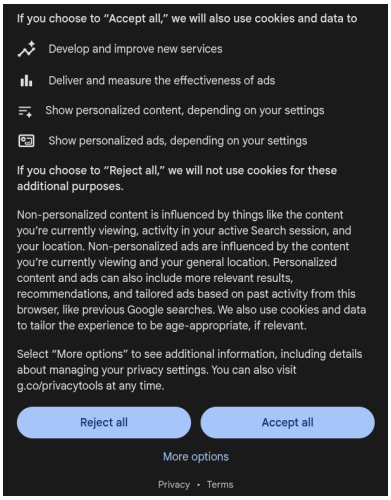


Figure 6: Gemini web consent banner.

D Sample data sent

We list representative samples of data transmitted during interactions in Figures 7, 8, 9, 10, 11. The examples illustrate the information observed in requests sent to analytics, telemetry, and advertising endpoints, including conversation identifiers, titles, prompts, URLs, user identifiers, and tracking pixels. Sensitive values have been redacted for readability and privacy.

```
TikTok Pixel --- POST analytics.tiktok.com/api/v2/pixel/
    act
"user":{"anonymous_id": "<_ttp>", ...},
"page":{
  "url": "https://grok.com/c/<convUUID>?rid=...",
  ...
},
"meta":{"title": "<convTitle>", ...}

Meta Pixel PageView --- GET www.facebook.com/tr/?...
ev           = PageView
dl           = https://grok.com/c/<convUUID>?rid=...
fbp          = <_fbp>
pmd[title]   = <convTitle>

Meta Pixel SubscribedButtonClick (share) --- GET www.
    facebook.com/tr/?...
ev               = SubscribedButtonClick
dl               = https://grok.com/c/<convUUID>?rid=...
cd[buttonText]   = Share
cd[pageFeatures] = {"title":"<convTitle>"}
fbp              = <_fbp>

sGTM relay --- POST sgtm-prod-985009374134.us-central1.
    run.app/data?event=sent_3_chat_messages
"page_location": "https://grok.com/c/<convUUID>?rid=...",
"page_referrer": "https://accounts.x.ai/",
"page_title":    "<convTitle>",
"common_cookie": {
  "_fbp": "<_fbp>",
  "_ttp": "<_ttp>"
},
"_dcid_temp": "<_dcid>"

Twitter Pixel --- GET analytics.twitter.com/1/i/adsct?...
pt                = <convTitle>
tw_document_href  = https://grok.com/c/<convUUID>?rid=...
twpid             = <_twpid>

Google Ads --- GET www.googleadservices.com/pagead/
    conversion/...
url     = https://grok.com/c/<convUUID>?rid=...
ref     = https://accounts.x.ai/
tiba    = <convTitle>
em      = tv.1~em.<userHash>
```

Figure 7: A single Grok conversation propagates to seven advertising and analytics trackers. Every recipient receives the same conversation identifier and the automatically generated title. Meta (`_fbp`), TikTok (`_ttp`), and Twitter (`_twpid`) cookies are synced across the flows—the Meta and TikTok cookies also travel server-to-server through the sGTM relay, invisible to ad blockers—and Google Ads additionally receives hashed user information (`em`).

```
GET region1.google-analytics.com/g/collect?...
cid = <cid>
dl  = https://gemini.google-b197145817.com/app/
dt  = <convTitle>
```

Figure 8: Google Analytics collects conversation Title from Gemini.

```
POST browser-intake-us5-datadoghq.com/api/v2/spans
"resource": "/organizations/{organization}/
    chat_conversations/{chat}/share"
"http.url": "https://claude.ai/api/organizations/<orgUUID
    >/
chat_conversations/<convUUID>/share",
"usr": {
"id": "<userUUID>",
"account_uuid": "<userUUID>",
"organization_id": "<orgUUID>",
"subscription_level": "free" },
"session":     { "id": "<sessionUUID>" },
"view":        { "id": "<viewUUID>" },
"user_action": { "id": "<actionUUID>" },
"device":    { "model": "AOSP on sargo", "ram_mb": 3577
    },
"os":        { "name": "Android", "version": "12" }
```

Figure 9: Datadog span emitted by the Claude *mobile* app for a conversation-share action. The Anthropic organization, conversation, and share UUIDs travel in the resource URL alongside account identifiers; further session, view, and per-action UUIDs let the recipient stitch every gesture into the same authenticated profile. The share resource is created without explicit user interaction.

```
POST browser-intake-us5-datadoghq.com/api/v2/rum?dd-api-
    key=...
{
  "type": "long_task",
  "application":{"id":"<applicationUUID>"},
  "view":   {"url":"https://claude.ai/share/{id}", ...},
  "session":{"id":"<sessionUUID>","type":"user"},
  "usr":    {"anonymous_id":"<anonUUID>",...},
  "long_task":{
    "scripts":[{
      "source_url":"https://claude.ai/share/<shareUUID>",
      "invoker":   "https://claude.ai/share/<shareUUID>"
    }]
  },
  ...
}
```

Figure 10: Datadog RUM beacon emitted by the Claude *web* client. The shared-conversation permalink is reported as the long-task script source and invoker, alongside application and session identifiers.

```
TikTok Pixel --- POST analytics.tiktok.com/api/v2/pixel/
    act
"page":{
  "url":          "https://grok.com/share/<shareUUID>?rid
      =...",
  ...
},
"open_graph":{
  "og:image":     "https://grok.com/share/<shareUUID>/
      opengraph-image/<shareUUID>",
  "og:image:alt": "<convScreenshotPreview>",
  ...
},
"meta":{
  "title":            "<convTitle>",
  "meta:description": "<UserPrompt>"
}

Meta Pixel PageView --- GET www.facebook.com/tr/?...
ev              = PageView
dl              = https://grok.com/share/<shareUUID>?rid
    =...
fbp             = <_fbp>
pmd[title]      = <convTitle>
pmd[description] = <UserPrompt>
```

Figure 11: A shared Grok conversation propagates extra information about the conversation to Meta and TikTok advertising and analytics trackers. Both recipients receives the same shared conversation identifier, the automatically generated title and the last user prompt. Meta (`_fbp`) and TikTok (`_ttp`) cookies are synced across the flows. Additionally, TikTok receives a screenshot of the most recent part of the conversation, as shown in Figure 12.

E Conversation Screenshots Leakage

We captured Grok's screenshot as transmitted to TikTok when a shared conversation is accessed, as shown in Figure 11. The screenshot captures a leaked user conversation in Grok, as it is illustrated in Figure 12.

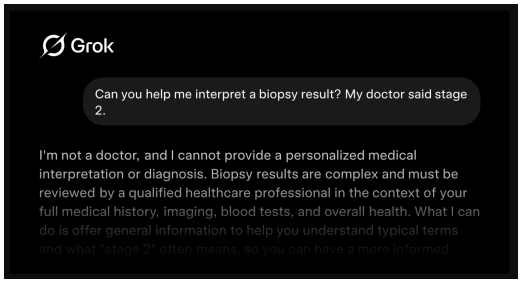


Figure 12: An example of the exact screenshot that gets sent to TikTok by Grok on shared interactions.