Privacy Mediation for Child Digital Literacy: Building Surveillance Awareness for Society

Antony Seabra, Claudio Cavalcante, Sergio Lifschitz

[1]Departamento de Informatica - PUC-Rio

{amedeiros, claudio.cavalcante, sergio}@inf.puc-rio.br

Abstract. *Children and adolescents increasingly interact with mobile applications that silently exchange personal information with multiple analytics, advertising, and tracking services, often within broader digital environments that may also expose them to profiling, manipulation, and other forms of online risks. Because these data flows remain largely invisible to non-specialists, young users and their caregivers are often unable to recognize how everyday app usage may expose minors to these persistent forms of digital surveillance. In this paper, we propose a Privacy Mediation Architecture that transforms mobile network observations into semantically meaningful and user-interpretable privacy explanations. Designed particularly for contexts involving minors, the architecture aims to support digital literacy by making hidden surveillance practices understandable to children, adolescents, and their caregivers. An exploratory evaluation with parent-child pairs indicates that semantically mediated explanations may improve users' understanding of otherwise opaque mobile data collection behaviors, reinforcing transparency and explanation as mechanisms for privacy awareness and informed decision-making.*

1. Introduction

The widespread adoption of mobile applications has led to an unprecedented expansion in the collection, processing, and monetization of personal data. Contemporary digital ecosystems are increasingly characterized by pervasive surveillance infrastructures in which user behavior is continuously captured, analyzed, and transformed into informational assets for commercial and strategic purposes [Zuboff 2019]. In mobile environments, this dynamic is particularly pronounced, as applications frequently establish background connections with analytics providers, advertising networks, recommendation infrastructures, and data brokers [Seabra et al. 2025]. Although these interactions are often encrypted and invisible to ordinary users, they remain central to how mobile ecosystems operate. This asymmetry between organizational visibility and user understanding becomes especially critical in the context of minors. Children and adolescents are among the most intensive users of smartphones and digital platforms, yet they typically lack the conceptual and cognitive resources required to understand how invisible technical interactions relate to profiling, behavioral monitoring, and broader surveillance practices. Existing privacy protection approaches in this context have largely emphasized restrictive mechanisms, such as parental control applications or content filtering, which may reduce exposure to certain risks but often fail to foster awareness, understanding, or gradual autonomy.

At the same time, research in network measurement and privacy auditing has developed sophisticated techniques for observing and analyzing mobile traffic, enabling the

identification of tracking behaviors, third-party communication, and hidden data flows [Seabra et al. 2024]. However, these approaches are primarily designed for experts and rarely translate into user-facing mechanisms capable of supporting comprehension and reflective decision-making. Conversely, research in privacy literacy and human-computer interaction emphasizes explanation, transparency, and collaborative learning, but typically lacks direct integration with real-time technical evidence. This disconnect reveals an important gap: existing systems can either detect surveillance or educate users about privacy, but rarely combine both capabilities in a unified socio-technical artifact. We argue that effective privacy protection in surveillance-driven mobile ecosystems requires a shift from passive control toward active mediation, where privacy is understood not only as restriction, but also as a process of knowledge construction and digital literacy.

To address this challenge, we propose a *Privacy Mediation Architecture* that transforms low-level mobile network observations into semantically meaningful and user-interpretable privacy explanations. The proposed approach integrates on-device traffic observation, metadata-based inference, semantic attribution, and explanatory mediation to bridge the gap between opaque technical communication patterns and human understanding. Designed particularly for contexts involving minors and their caregivers, the architecture emphasizes literacy-oriented privacy awareness rather than silent enforcement. The main contributions of this work are fourfold. First, we introduce a deployable architecture for privacy mediation operating directly at the mobile network layer without requiring invasive device modifications. Second, we define a semantic transformation process that maps technical traffic observations into higher-level concepts such as organizations, inferred data practices, and collection purposes. Third, we propose a literacy-oriented mediation model that adapts privacy explanations to different audiences, including adolescents and caregivers. Finally, we present an exploratory human-centered evaluation assessing whether semantically mediated explanations improve the interpretability of hidden surveillance practices in mobile ecosystems.

2. Background

Contemporary digital infrastructures are often characterized by pervasive data collection practices. According to [Zuboff 2019], modern digital platforms operate within an economic logic in which user behavior is continuously captured and transformed into data assets. This model extends beyond direct user interactions, encompassing behavioral patterns, preferences, and contextual signals. A key characteristic of such systems is the asymmetry between data collection and user awareness. While organizations can collect and process large volumes of behavioral data, individuals typically lack visibility into these processes. This asymmetry has been identified as a central challenge for privacy and user autonomy in digital environments [Nissenbaum 2010]. In the context of minors, this challenge is further intensified. Children and adolescents are among the most intensive users of smartphones and mobile applications, relying on handheld devices as their primary gateway to communication, entertainment, learning, and social interaction. Yet, despite this pervasive exposure, minors rarely possess the cognitive and conceptual frameworks required to understand how these invisible data flows are collected, shared, and operationalized, making them particularly vulnerable to opaque digital surveillance practices.

Mobile applications operate as distributed systems that continuously interact with

remote infrastructures, including first-party servers, analytics providers, advertising networks, and content delivery services, often without explicit user awareness, forming complex ecosystems of data exchange [Ren et al. 2016]. Although the widespread adoption of Transport Layer Security (TLS) protocol has significantly reduced visibility into the content of network communications, encryption does not eliminate observability altogether. Metadata such as destination domains, IP addresses, connection frequency, packet sizes, timing patterns, and information exposed through mechanisms such as Server Name Indication (SNI) remain accessible and can provide meaningful insights into application behavior [Coull and Dyer 2014]. As a result, modern network measurement and privacy analysis have progressively shifted from payload inspection toward metadata-based inference, leveraging observable side-channel signals to characterize otherwise opaque communication practices. This methodological transition has become a foundational principle in contemporary mobile traffic analysis [Aceto et al. 2019].

Traditional approaches to privacy protection have focused on restricting or blocking data flows. However, recent perspectives emphasize the importance of transparency and user understanding. Privacy mediation refers to the process of transforming technical data collection mechanisms into representations that can be interpreted and acted upon by users. This perspective is closely related to the concept of digital literacy, which involves the ability to understand and critically evaluate digital systems and their implications [Eshet-Alkalai 2004]. In the context of privacy, digital literacy includes recognizing data collection practices, understanding their purposes, and assessing their consequences. For minors, mediation plays a crucial role in supporting gradual learning and autonomy. Rather than relying solely on external control mechanisms, mediation enables the development of awareness and informed decision-making, aligning privacy protection with educational objectives By grounding privacy in processes of interpretation and learning, this perspective establishes the conceptual foundation for systems that aim to make data collection visible, understandable, and actionable.

3. Related Work

Recent research has increasingly focused on auditing privacy practices in services used by children and adolescents. [Figueira et al. 2024] propose *DiffAudit*, a methodology that compares network data flows across different user groups and consent conditions. Their large-scale analysis reveals problematic practices such as data collection before consent, inconsistent policy disclosures, and extensive sharing with third-party trackers [Figueira et al. 2024]. Similarly, [Alomar et al. 2025] analyze thousands of child-directed Android applications and observe improvements in compliance over time, including increased adoption of secure communication. Nevertheless, their study shows that third-party software remains a persistent source of privacy violations and inaccurate disclosures.

In parallel, research in human-computer interaction emphasizes that children's privacy cannot be addressed solely through restrictive controls. [Alghythee et al. 2024] argue for the importance of family-centered privacy discussions and highlight the need for tools that support collaborative understanding rather than unilateral control. A further study [Liu et al. 2024] explores family-based privacy education and show that informal conversations play a central role in how children learn about privacy. Their findings indicate that families often lack a structured understanding of privacy concepts, suggesting

the need for interfaces that communicate not only actions, but also underlying principles.

In relation to mobile traffic interception and privacy auditing, prior work has demonstrated that mobile operating systems provide a feasible technical basis for on-device traffic observation. [Wu et al. 2017] proposed *MopEye*, an Android application built on top of the `VpnService` API that passively captures traffic from all applications and forwards it through socket calls, enabling per-app measurements without requiring root privileges. This work shows that local VPN-based interception is a deployable and practical design pattern for real-world smartphones. At the same time, VPN-based mediation introduces significant trust and transparency concerns. [Ikram et al. 2016] conducted a large-scale analysis of Android applications using VPN permissions and identified multiple privacy and security issues, including insecure tunneling protocols, DNS leaks, TLS interception, and traffic manipulation for tracking or advertising purposes. These findings highlight that traffic mediation is not purely infrastructural, but directly impacts user privacy and trust.

4. Methodology

This research follows a Design Science Research paradigm, in which the central contribution is the design, construction, and evaluation of a socio-technical artifact for privacy mediation in mobile environments used by minors. The proposed methodology defines a progressive transformation process through which low-level network observations are converted into semantically meaningful and user-understandable representations. As illustrated in Figure 1, the approach is organized into five main steps: mobile network signal capture, evidence formation, analytical inference, semantic mediation, and user awareness and reflection.

In Step 1, the methodology begins with the observation of outbound network communications generated by applications running on a mobile device through a locally deployed proxy that mediates traffic at the device level. This infrastructure enables controlled visibility into otherwise opaque application behavior and, when technically feasible and ethically justified, may support packet inspection. However, the proposed approach primarily relies on observable metadata rather than payload analysis in order to mitigate limitations imposed by encryption and minimize additional privacy risks. The collected signals include destination domains, connection frequency, temporal communication patterns, and associations between applications and external services. This step establishes the empirical foundation of the mediation process by making hidden interactions between mobile applications and external infrastructures observable.

In Step 2, the observed signals are semantically enriched through a knowledge attribution process that transforms raw technical traces into privacy-relevant evidence. Rather than treating network events as isolated records, the methodology interprets recurring communication behaviors in light of known tracking infrastructures and inferred communication patterns. To support this process, the system relies on a distributed knowledge ecosystem combining established public repositories, such as tracker intelligence databases, with adaptive local learning mechanisms. Known domains can be directly associated with previously identified analytics, advertising, or telemetry services, while previously unseen communication endpoints may be analyzed through behavioral classification models based on metadata patterns and destination characteristics. This allows

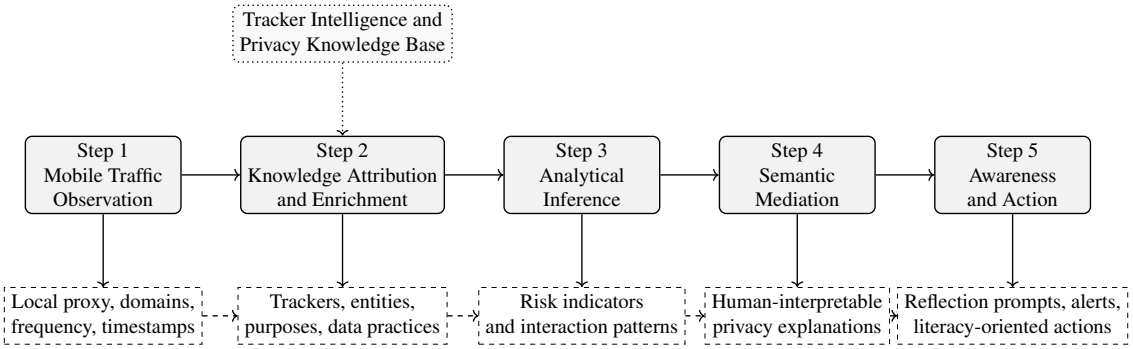

Figure 1. Overview of the proposed privacy mediation methodology.

the knowledge base to evolve incrementally as surveillance infrastructures become more dynamic and obfuscated.

In Step 3, analytical inference is applied to identify communication patterns and estimate potential privacy relevance. The methodology considers indicators such as persistence, frequency, destination category, inferred data sensitivity, and recurrence across applications. Rather than aiming at definitive legal classification, this stage prioritizes signals that warrant explanation by distinguishing ordinary operational communication from potentially intrusive surveillance-related behaviors. A heuristic scoring model may be used to aggregate these indicators into comparative privacy relevance assessments.

In Step 4, the prioritized technical evidence is translated into semantically meaningful explanations through a privacy mediation layer oriented toward digital literacy. Instead of exposing users to technical artifacts such as obscure domains or protocol metadata, the system presents human-interpretable narratives that explain who is likely receiving the data, what type of information may be involved, and for which likely purposes such exchanges occur. By converting technical observations into contextualized explanatory representations, this final step transforms passive traffic auditing into an educational socio-technical intervention aimed at strengthening privacy awareness among minors and their caregivers.

In Step 5, the semantically enriched information is presented to users through literacy-oriented explanations, alerts, and narratives. The goal is not merely to expose technical data, but to support awareness, reflection, and informed action. In the specific context of minors, this step is especially important because children and adolescents may lack the conceptual background required to interpret technical privacy signals. Therefore, the methodology emphasizes cognitive mediation: technical observations are translated into accessible explanations that help users understand the relationship between application use, data collection, third-party services, and potential privacy consequences.

Ethical considerations are embedded throughout the methodology. The approach prioritizes data minimization by avoiding payload inspection and reducing the retention of sensitive information. Whenever possible, processing is performed locally on the device, limiting additional exposure to external systems. In deployments involving minors, the methodology requires explicit parental or guardian consent and emphasizes transparency regarding both the monitored network interactions and the operation of the mediation artifact itself. These principles ensure that the proposed system does not reproduce the same

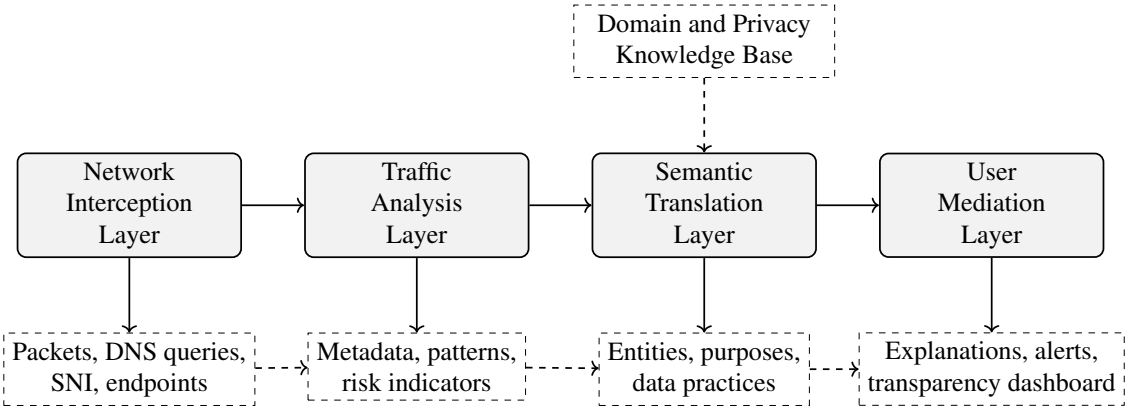

Figure 2. Privacy Mediation Architecture: layered transformation from network signals to user-understandable privacy explanations.

opacity it seeks to make visible, but instead functions as a privacy-preserving mediation layer between mobile data collection infrastructures and user understanding.

5. Architecture and Implementation

This section presents the Privacy Mediation Architecture, a layered system designed to transform raw mobile network traffic into semantically meaningful privacy representations. While the methodology defines the conceptual process through which network observations become user awareness, the architecture specifies the computational components that operationalize this process. As shown in Figure 2, the system is organized as a horizontal pipeline composed of four main layers: Network Interception, Traffic Analysis, Semantic Translation, and User Mediation. Each layer receives the output of the previous one, refines its meaning, and produces a more interpretable representation of mobile data flows.

The Network Interception Layer is responsible for observing outbound communication generated by applications running on the mobile device. This layer relies on a local VPN mechanism based on a virtual network interface, allowing traffic to be captured without requiring root access or modifications to the operating system. Its role is not to inspect sensitive payloads, but to collect observable communication signals such as DNS queries, destination endpoints, Server Name Indication values, connection timestamps, and application-level traces when available. In this way, the architecture preserves deployability in real-world mobile environments while minimizing invasive inspection.

The Traffic Analysis Layer processes the captured signals and extracts metadata that can remain visible even when communications are encrypted. This layer identifies destination domains, IP addresses, temporal patterns, connection frequency, recurrence, and service categories. These features are then used to infer behavioral patterns associated with applications and external services. For instance, repeated connections to advertising or analytics domains may indicate persistent data collection behavior. The output of this layer is therefore not merely a list of network events, but a structured set of metadata, behavioral patterns, and privacy risk indicators.

The Semantic Translation Layer constitutes the central interpretive component of the architecture. It maps technical identifiers, such as domains and endpoints, to real-

world entities, organizational ecosystems, and inferred purposes. For example, a connection to `graph.facebook.com` can be associated with the Meta ecosystem and interpreted according to known advertising, analytics, or platform integration practices. This layer is supported by a Domain and Privacy Knowledge Base, which maintains mappings between domains, organizations, service categories, data practices, and privacy-relevant purposes. Through this process, low-level network observations are semantically grounded into concepts that are meaningful to users.

Beyond entity mapping, the Semantic Translation Layer classifies interactions according to their inferred intent, distinguishing among functional communication, analytics, advertising, tracking, content delivery, authentication, and potentially harmful interactions. This classification enables the system to move from technical observability to conceptual interpretation. As a result, the architecture can explain not only that an application contacted a given domain, but also who may be involved, what type of service the domain provides, and why the interaction may be relevant from a privacy perspective.

The User Mediation Layer presents the semantically enriched information through interfaces designed to support awareness, reflection, and informed decision-making. Instead of operating only as a blocking mechanism, the architecture emphasizes explanation and transparency. The system may present alerts, summaries, timelines, application profiles, or transparency dashboards that describe which organizations receive data, what purposes are inferred, and how frequently such interactions occur. This design is particularly important in contexts involving minors, where the objective is not only to restrict behavior, but also to support the development of privacy literacy.

Overall, the architecture implements a progressive transformation from network-level signals to user-understandable privacy explanations. The Network Interception Layer observes communication, the Traffic Analysis Layer extracts patterns and risk indicators, the Semantic Translation Layer grounds technical evidence into real-world meanings, and the User Mediation Layer translates these meanings into accessible explanations. This layered organization ensures modularity, extensibility, and separation of concerns, allowing each component to evolve independently while preserving the central goal of the system: transforming opaque mobile data collection practices into interpretable and actionable privacy knowledge.

6. Evaluation

The evaluation focuses on exploring how semantically mediated explanations derived from mobile network observations are perceived by minors and their caregivers in terms of interpretability, usefulness, and privacy awareness. Rather than evaluating the system as a pure traffic auditing tool, the study examines its effectiveness as a socio-technical mediation artifact capable of translating low-level technical evidence into meaningful privacy narratives.

To this end, we conducted an exploratory user study involving ten parent-child pairs (20 participants in total), where adolescents and their caregivers were separately exposed to semantically mediated explanations generated from real mobile traffic observations captured through the proposed proxy-based infrastructure. The evaluation deliberately focused on interpretability and perceived educational value, as these dimensions are directly aligned with the digital literacy goals of the architecture.

Table 1. Examples of semantic mediation generated from captured mobile traffic evidence.

Captured Evidence	Adolescent-Oriented Explanation	Parent-Oriented Explanation
`analytics.tiktok.com`	This app may be tracking what you watch and how long you stay to learn about your interests.	The application appears to transmit behavioral usage telemetry to TikTok analytics services, potentially enabling profiling of interaction habits and engagement patterns.
`googleads.g.doubleclick.net`	This app may be contacting an advertising company to decide which ads to show you.	The application communicates with Google's advertising infrastructure, suggesting ad targeting or cross-service behavioral advertising activities.
`connect.facebook.net`	This app may be sharing information with Facebook services, even if you are not actively using Facebook.	The observed traffic indicates integration with Meta tracking infrastructure, potentially supporting cross-platform audience measurement or advertising attribution.
`api-recommend.bigdata.samsung.com`	This service may be learning what you like to suggest products or content.	The communication suggests recommendation processing involving preference inference and personalization mechanisms.

Table 1 presents representative examples of semantic mediation generated from observed network evidence. For adolescents, explanations were intentionally simplified and framed in terms of familiar digital experiences, such as advertisements, recommendations, or app behavior tracking. For example, a connection to *googleads.g.doubleclick.net* was translated as an explanation that the application may be contacting an advertising company to determine which ads should be shown. Similarly, traffic directed to *analytics.tiktok.com* was explained as the application learning what the user watches and how long they remain engaged. In contrast, explanations directed to parents adopted a more explicit privacy vocabulary, including references to behavioral profiling, audience measurement, third-party sharing, and personalization mechanisms.

Participants were asked to evaluate the generated explanations according to five criteria using a five-point Likert scale: clarity, understanding of privacy implications, perceived usefulness, actionability, and trust in the explanation. Given the exploratory nature of this study and the limited sample size, results are presented descriptively as average participant perceptions rather than inferential statistical comparisons. Figure 3 summarizes the average evaluations provided by adolescents and parents. Overall, the results suggest moderate perceptions regarding the usefulness and interpretability of the semantically mediated explanations, with consistently higher evaluations from parents compared to adolescents.

Parents reported more favorable perceptions across all dimensions, particularly regarding clarity, understanding of privacy implications, and perceived usefulness, suggesting that the explanations were generally understandable and relevant from an adult perspective. In contrast, adolescent evaluations were more moderate, especially in trust, actionability, and deeper understanding of privacy implications. These differences suggest that while simplified privacy narratives may help younger users recognize certain aspects of otherwise opaque technical behaviors, additional educational scaffolding, con-

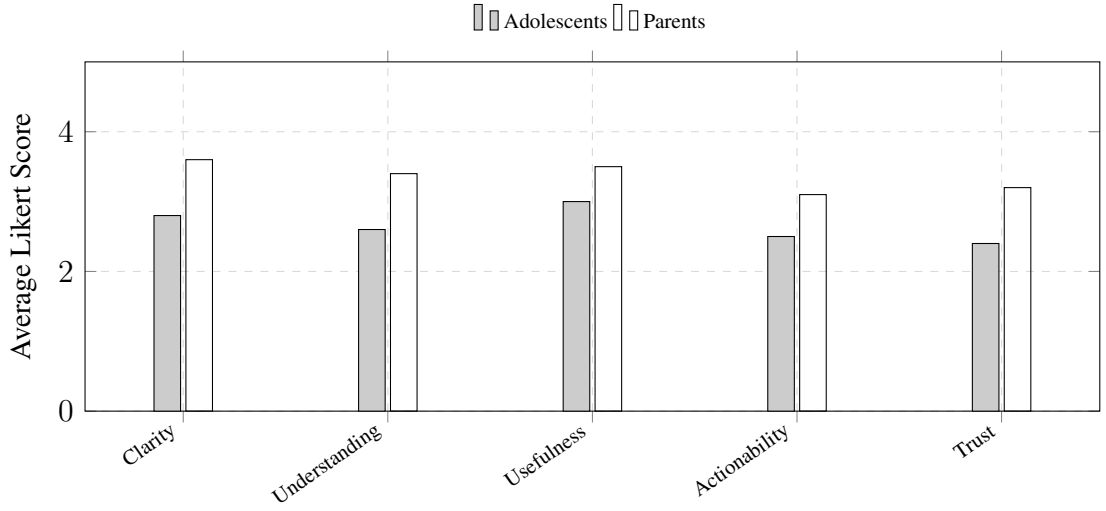

Figure 3. Average participant evaluation of semantically mediated privacy explanations in the exploratory user study (Likert 1–5).

textual examples, or more interactive mediation strategies may be necessary to strengthen comprehension, trust, and practical engagement among minors.

Although the findings do not establish causal effectiveness or generalizable behavioral impact, they provide preliminary evidence that semantically mediated privacy explanations may contribute to privacy awareness and support digital literacy discussions between minors and caregivers.

7. Conclusions and Future Work

This paper introduced a Privacy Mediation Architecture that transforms low-level mobile network observations into semantically meaningful and user-interpretable representations. By bridging technical traffic analysis with privacy-oriented semantic explanation, the proposed approach shifts privacy protection from purely restrictive mechanisms toward active user awareness and interpretive empowerment. Rather than exposing raw technical artifacts, the system translates hidden communication patterns into understandable narratives about data recipients, inferred purposes, and surveillance-related practices, particularly in contexts involving minors. The exploratory evaluation suggests that semantically mediated explanations can support users' understanding of otherwise opaque mobile data collection behaviors.

As future work, we intend to evolve the proposed architecture into an interactive educational environment centered on gamified digital literacy for children and adolescents. Instead of acting solely as an explanatory mediation tool, the system could actively engage young users through privacy-awareness missions, progressive learning challenges, and reward mechanisms that encourage understanding of surveillance practices and privacy-conscious decision-making. Parents or caregivers could be informed of participants' progress, enabling collaborative family engagement in digital literacy. More broadly, longitudinal studies should investigate whether sustained interaction with such privacy mediation mechanisms produces lasting improvements in surveillance awareness, critical digital behavior, and informed privacy choices.

References

Aceto, G., Botta, A., de Donato, W., and Pescapè, A. (2019). Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges. *IEEE Transactions on Network and Service Management*, 16(2):445–458.

Alghythee, K. K. A., Hrncic, A., Singh, K., Kunisetty, S., Yao, Y., and Soni, N. (2024). Towards understanding family privacy and security literacy conversations at home: Design implications for privacy literacy interfaces. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM.

Alomar, N. N., Reardon, J., Girish, A., Vallina-Rodriguez, N., and Egelman, S. (2025). The effect of platform policies on app privacy compliance: A study of child-directed apps. *Proceedings on Privacy Enhancing Technologies*, 2025(3):170–191.

Coull, S. E. and Dyer, K. P. (2014). Traffic analysis of encrypted messaging services: Apple imessage and beyond. In *Proceedings of the ACM SIGCOMM Conference on Internet Measurement Conference (IMC)*, pages 435–448. ACM.

Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13(1):93–106.

Figueira, O., Trimananda, R., Markopoulou, A., and Jordan, S. (2024). Diffaudit: Auditing privacy practices of online services for children and adolescents. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24)*, pages 488–504. ACM.

Ikram, M., Vallina-Rodriguez, N., Seneviratne, S., Käafar, M. A., and Paxson, V. (2016). An analysis of the privacy and security risks of android VPN permission-enabled apps. In *Proceedings of the 2016 Internet Measurement Conference*, pages 349–364. ACM.

Liu, L., Gao, L., Soni, N., and Yao, Y. (2024). Exploring design opportunities for family-based privacy education in informal learning spaces. *Proceedings on Privacy Enhancing Technologies*, 2024(3):127–143.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Ren, J., Lindorfer, M., Dubois, D. J., Rao, A., and Choffnes, D. (2016). Recon: Revealing and controlling pii leaks in mobile network traffic. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 361–374. ACM.

Seabra, A., Afonso Glatzl Junior, L., and Lifschitz, S. (2024). Surveillance capitalism revealed: Tracing the hidden world of web data collection. *arXiv e-prints*, pages arXiv–2412.

Seabra, A., Cavalcante, C., and Lifschitz, S. (2025). From surveillance to manipulation: How data collection enables targeted disinformation. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 302–313. SBC.

Wu, D., Chang, R. K. C., Li, W., Cheng, E. K. T., and Gao, D. (2017). MopEye: Opportunistic monitoring of per-app mobile network performance. In *2017 USENIX Annual Technical Conference (USENIX ATC 17)*, pages 445–457. USENIX Association.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. PublicAffairs.