Pybus, Jennifer; Matheson, Katrina Nicole; Lachmansingh, Andrea

Article

Extraction-by-design: Auditing infrastructures of datafication in babytracking apps

Internet Policy Review

Provided in Cooperation with:
Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: Pybus, Jennifer; Matheson, Katrina Nicole; Lachmansingh, Andrea (2026) : Extraction-by-design: Auditing infrastructures of datafication in babytracking apps, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 15, Iss. 1, pp. 1-34,
https://doi.org/10.14763/2026.1.2087

This Version is available at:
https://hdl.handle.net/10419/339541

INTERNET POLICY REVIEW

Volume 15 **Issue** 1

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

Extraction-by-design: Auditing infrastructures of datafication in baby-tracking apps

Jennifer Pybus *York University*
Katrina Nicole Matheson *York University*
Andrea Lachmansingh *York University*

DOI: https://doi.org/10.14763/2026.1.2087

Published: 27 February 2026
Received: 7 July 2025 **Accepted:** 24 November 2025

Funding: The authors did not receive any funding for this research.
Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Citation: Pybus, J., Matheson, K.N., & Lachmansingh, A. (2026). Extraction-by-design: Auditing infrastructures of datafication in baby-tracking apps. *Internet Policy Review*, *15*(1). https://doi.org/10.14763/2026.1.2087

Keywords: Baby-tracking applications, Platform infrastructures, Datafication, Data sovereignty, Privacy

Abstract: Millions of parents across Europe use mobile baby-tracking applications. These digital aides generate intimate data sets by accessing sensitive information about (un)born babies, pregnancies, and family caregiving routines. Crucially, the tracking infrastructures embedded in these apps enable what we call extraction-by-design, whereby routine caregiving inputs are captured, standardised, and repurposed for behavioural profiling and monetisation within an opaque digital economy. We ask: (i) How do baby-tracking apps extract and circulate sensitive health data? (ii) How do tracking infrastructures enable behavioural profiling and monetisation? and (iii) What data governance interventions are needed to strengthen the current regulatory environment? We conducted a mixed-method audit of 14 of the most downloaded Android baby-tracking apps. We combine a manifest data audit (qualitative analysis of app manifests using large language models), reviews of privacy policies and data safety agreements, and walkthroughs of app interfaces to assess how data is collected, processed, and shared. All 14 apps shared data with third parties, including sensitive due dates and pregnancy loss data. Many also transmitted identifiers across borders, often without meaningful consent. We conclude by arguing for policy interventions that treat (infant) data privacy as a matter of collective public interest, including clear limitations on inference-based profiling.

Introduction

Millions of parents use mobile baby-tracking applications. Central to the appeal of this expanding market is the promise that these apps will generate insights and optimise various aspects of infant care, from sleep patterns and feeding schedules to tracking developmental milestones. Crucially, they present themselves as digital "companions" (Glow Baby, n.d.) or "reliable assistants" (Baby Care, n.d.) that draw on expert knowledge to take the "guesswork out of parenting" (Huckleberry, n.d.), thereby helping families "make data sense" (Lupton, 2020) of the countless moments that comprise a child's life. Baby-trackers are therefore marketed through a normative assumption that overwhelmed parents cannot, on their own, track and interpret the growing volume of data now associated with healthy development, thereby normalising the notion that parenthood depends on the datafication of infant care (Barassi, 2017). This framing aligns with the "logic of deficiency" identified by Neff and Nafus (2016) in research on self-tracking technologies, wherein parenting in a datafied world comes to require continuous record keeping, and even predictive artificial intelligence (AI) analytics. The *What to Expect* baby-tracking application, which boasts five million downloads globally, offers this value proposition:

> When it comes to feeding your baby, you'll need to know some specifics: What time did baby's most recent feed start? How many ounces was that last bottle? How long did baby breastfeed and on which breast(s)?…what's the color and consistency of baby's poop?… the list goes on…It's a lot to remember, especially when you're running on (very) little sleep and balancing everything else that comes with new parenthood. Fortunately, a baby-tracker like …*What to Expect* app can help (Jena, 2022).

Subsequently, this "numbering of infant lives" (Wernimont, 2019) creates continuous opportunities to generate intimate datasets rooted in access to sensitive information about (un)born babies, pregnancies, and family caregiving routines. As we discuss, these data can be leveraged through a range of monetisation strategies, circulating through a largely invisible digital economy that includes not only major platforms such as Google, Meta, and Amazon, but also offshore advertisers and data brokers. These data sharing practices raise significant concerns about the limited protections citizens have under current configurations of data governance in Europe and the United Kingdom, and contribute to growing policy concerns about the lack of transparency and accountability in how user-produced health data is accessed by mobile applications and their third parties (Felsberger, 2025; McMil-

lan, 2022; Lupton, 2016; Neff & Nafus, 2016). As prominent scholarship has noted, user-generated health data (Ostherr et al., 2017), that is, non-clinical health data that users choose to share in mediated spaces such as apps, can facilitate a range of privacy harms, including scams, targeted misinformation, and various forms of discrimination that may follow parents and their children indefinitely into the future (Witzleb et al., 2020).

Such opaque and extensive data-sharing practices sit uneasily with the principle of "privacy by design," first articulated by Ann Cavoukian (2011) and integrated into Art. 25 of the General Data Protection Regulation (GDPR) through the requirement for "data protection by design and by default" (Regulation (EU) 2016/679). This principle requires privacy protections to be built into networked systems at the design stage, rather than added after the fact. Within this regulatory context, baby-tracking apps raise ongoing legal and civil concerns. Framed as democratizing tools that challenge the unequal power hierarchies that have long shaped women's health, or as empowering vehicles for better understanding (infant) bodies (Ostherr et al., 2017), these apps are also marketed and designed within extractive data economies. Prominent scholars of female-centered technologies or FemTech, have noted that the commercial value of this biometric data is now estimated to be up to fifty times greater than that of credit data (Gilman, 2021; Mehrnezhad et al., 2024). This creates strong incentives to gain access to rich and novel datasets produced through these apps. For us, this emphasizes the need for socio-technical auditing tools capable of revealing how such applications and their tracking infrastructures operate, and with whom they share this intimate data.

We therefore propose an audit of the 14 most downloaded Android baby-tracking applications across Europe and the United Kingdom to address the following questions: (i) How do baby-tracking apps create continuous opportunities to extract and circulate infant and parental data across opaque digital ecosystems? (ii) How do the infrastructures that support tracking within these apps enable behavioural profiling and monetisation? and (iii) What data-governance interventions are needed to strengthen the current regulatory environment? To explore these questions, we employ a novel mixed-method approach based on Pybus and Mir's (2025) software development kit (SDK) data audit. Our study reveals a clear gap in the protection of infant data, with every app in our sample transmitting potentially sensitive information to third-party advertisers, with some potentially sharing a mother's due date or a pregnancy loss. The inconsistencies and instances of non-compliance we observed across stated Data Safety Agreements (DSAs), privacy policies (PPs), and back-end infrastructures point to a mobile ecosystem that privileges not priva-

cy but what we call *extraction-by-design*. Accordingly, we position this work in direct conversation with both developers and policymakers, aiming to strengthen pathways for meaningful agency and protection for parents, who should be able to decide how, and with whom, both first and third parties access, share, and produce new data about their pregnancies and their children.

The rise of intimate surveillance of children in mobile applications

Since the overturning of *Roe v. Wade* in the US, feminist legal scholars have raised renewed concerns about the business model of digital devices that capitalise on what Citron (2022) has termed *intimate health data:* sensitive information about bodies, health, sex, gender, sexual orientation, close relationships, online searches, reading habits, or even private communications. Much of this work, particularly on FemTech, aligns with critiques from critical data studies, wherein a substantial body of scholarship has documented how the development of children is increasingly commodified by a range of digital devices (Barassi, 2020). For instance, Leaver (2015) characterises the monitoring, capturing, and sharing of children's data as forms of "intimate surveillance," which normalises what Mascheroni and Siibak (2021) term "datafied childhoods." Lupton's (2019) concept of "caring surveillance" further highlights the mutually constitutive relationship between apps and motherhood, and how the construct of the responsible parent now depends on technological interventions that rely on intimate data.

From this perspective, baby-tracking apps can reinforce "norms and expectations about what constitutes a 'good' mother" (Lupton, 2019, p. 2) while embedding surveillance logics into maternal responsibilities (Barassi, 2017). This normative framing aligns with a substantial body of scholarship that examines how exploitative, datafied relationships emerge through the ways parents intentionally and unintentionally generate data about their children, and how their parenting behaviours become entangled with data-driven technologies that shape children's lives from an early age (Steinberg, 2017; Livingstone & Blum-Ross, 2020). Yet, within this growing baby-tracking ecosystem, regulatory oversight remains limited (Mascheroni & Siibak, 2021) despite being widely endorsed by healthcare professionals (Thornham, 2019). This lack of oversight is not incidental; it is embedded in the routine integration of software services such as software development kits (SDKs), which developers rely on to build and monetise their apps, often with limited transparency about what data is collected and shared with these third parties.

Infrastructures of datafication in apps: how intimate data is captured and unprotected

While often described in the literature as 'trackers,' SDKs are more than tools for data extraction. They are modular packages, predominantly provided by large platforms such as Google, Meta, Amazon, and increasingly Microsoft and ByteDance (Pybus & Coté, 2024). More specifically, these operate as generative infrastructure that continuously feed data into wider formations of platform power, both as assets for value extraction (Birch and Muniesa, 2020) and as recursive tools whose expanding reach enables new forms of insight and revenue. The dual role of SDKs is central to their appeal, alongside developers' increasing reliance on them to implement core functionalities within their apps, which includes a growing suite of monetisation services like campaign analytics, cross-device tracking, digital advertising and behavioural profiling (Pybus & Coté, 2024; Flensburg & Lai, 2022; van der Vlist & Helmond, 2021). More recently, these infrastructures have become central to the provision of generative AI services, situating them as powerful enablers of extensive user surveillance, while simultaneously catalysing value production (Lomborg et al., 2024; Pybus & Coté, 2024).

Given the privileged role SDKs occupy in apps, we argue that they provide a critical site for analysing how platform power is extended and reinforced through infrastructural arrangements (Plantin et al., 2018). Drawing on conceptualisations of platform power by Nieborg et al. (2024) and van Dijck et al. (2019), we treat SDKs as an analytical entry point for examining the mechanisms that enable platforms to consolidate control. This argument builds on an extensive scholarly debate about the interrelationship between platform power and infrastructure power (see Nieborg & Poell, 2025; Lomborg et al., 2024; Cohen, 2023). Pybus et al. (2025) advance this discussion through the concept of an "infrastructure of datafication," showing how socio-technical objects such as SDKs, establish excludable (or blackboxed) pathways for the non-rivalrous data generated within the app. More plainly, these services allow multiple actors to simultaneously access and exploit the exact same data that are generated within any given mobile application. In turn, this data for service model produces new kinds of platform-specific dependencies, particularly as developers increasingly rely on these services to build and monetise their apps. These entanglements shape how personal data is mobilised and circulated to perpetuate increasing platform dominance within the app economy that entrench platform power while leaving users with limited agency or control over their data.

Challenges to protecting privacy

The asymmetries embedded within infrastructures such as SDKs raise privacy concerns, particularly within the legal "grey area" of mobile health (mHealth) apps (Schäfke-Zell, 2021). Specifically, these generate user-produced health data, which does not receive the same level of protection as medical or clinical health data (Ostherr et al., 2017). Since user-generated health data is legally equivalent to personal data in Europe, the lack of higher privacy standards for collection, access, and use, can easily lead to privacy harms. For example, in 2023, the Portuguese consumer protection association, Ius Omnibus, brought a class action lawsuit against the period-tracking app Flo, claiming "unauthorised sharing" of "highly sensitive information" (Ius Omnibus, n.d.). However, even when data sharing is authorised through consent mechanisms, social expectations surrounding app adoption can undermine ethical consent, as the pressure to download and use an app may outweigh a person's capacity to evaluate lengthy and opaque privacy policies (Bechmann, 2014). This raises questions about whether individual opt-outs or regulatory frameworks relying on informed consent can adequately protect the public (Kuntsman & Miyake, 2022).

Further complicating regulatory overreliance on informed consent, Solove (2025) argues that privacy is a context-dependent, social practice organised around shifting boundaries that govern how information is shared, inferred, and circulated. Privacy harms arise not only from what people disclose about themselves, but also from relational inferences that can act as proxies for protected attributes, emerging through the extraction and aggregation of population-level data at scale (Amoore et al., 2024). In other words, attributes such as pregnancy status or delayed infant milestones can only be inferred from aggregated, population-level user data rather than from any single person's disclosure. This dynamic reveals the limitations of privacy frameworks that place too much responsibility on individuals (Solove, 2025), while overlooking the fundamentally relational nature of data and the ways platforms repurpose, recombine, and generate information at scale (Cohen, 2019). For app users trying to negotiate this complex tracking ecosystem, it is challenging enough trying to determine how one of these apps might use their infant's data, let alone what it might mean if a third party then combines this data with other datasets for another purpose. The privacy challenges that SDKs pose are therefore foundational to the data economy and central to the privacy vulnerabilities that involve inadequately protected user-produced health data (Scatterday, 2021; Shipp & Blasco, 2020).

Auditing apps: what data are they sharing?

To audit mobile applications, one option is to focus on the back end, where it is easier to identify which companies provide services to a developer by asking which actors are present and what they are doing (Flensburg & Lai, 2022; Blanke & Pybus, 2020; Binns et al., 2018). Alternatively, researchers may concentrate on the front end, where users engage in a variety of data-sharing activities that may compromise their privacy (Malki et al., 2024; Light et al., 2018). Other scholarship attempts to connect these perspectives, most often by examining privacy permissions, mapping the number of third parties, and identifying gaps and omissions within privacy agreements (Langton & Ng, 2025; Almeida et al., 2022). Each of these approaches raise questions about whether end users are given meaningful opportunities to consent to the capture, sharing, and monetisation of their personal and/or sensitive data, yet they give relatively little attention to the Android manifest file within an app's APK as a potentially rich digital text to audit, beyond examining the privacy permissions. As we argue, app manifests can support further investigation into the kinds of data that apps access, including the discovery of app events (Pybus & Mir, 2025), how these data practices are structured, and how they are in turn represented to end users.

Building on Pybus and Mir's (2025) methodological contributions, we connect how end users are informed about an app's access to personal data with what can be observed through a back-end manifest audit (MA), attending to the mechanisms through which data, like app events, are accessed and transmitted to third-party SDK services. We have focused on app events because they are multifaceted (Bounegru et al., 2022). They can manifest as both technical objects and semantically meaningful artefacts, depending on how they are configured. Analytics companies such as Adjust define them as "anything a mobile marketer or developer determines is helpful to measure" (n.d.). While the tracking of these events is invisible to the user, the data that arise from them are foundational to behavioural profiling analytics. For the app itself, their customisation can enhance the sensitivity of the data they collect and transmit, giving rise to legal liability. This is illustrated by two notable court cases in the US: one involving the fertility-tracking app *Premom* (Sherman, 2023) and the other involving the period-tracking app *Flo* (*Frasco v. Flo Health*, 2021). In both US cases, plaintiffs claimed that the apps improperly shared customised health-related events, wherein developers assigned highly sensitive and overly descriptive app event labels to user activities such as "R_PREGNANCY_WEEK_CHOSEN" (*Frasco v. Flo Health,* 2021, p. 30) and "Ovulation/Static/Sucess" (United States of America v. Easy Healthcare, 2023, p. 8), which were then transmitted through SDKs to advertising platforms affiliated with Facebook and

Google.

As profile-building tools, app events render personal data not only 'platform-ready' (Helmond, 2015), but also asset-ready, enabling it to be cleaned, labelled, and primed for future (re)use by platforms and data brokers. To audit this data production, we deploy a static analysis, focused on an app's software architecture in order to identify third-party integrations and the services they provide (Binns et al., 2018; Flensburg & Lai, 2022; Pybus & Coté, 2022, 2024; Lomborg et al., 2024). Static analyses are particularly valuable because they make visible the infrastructures through which data move, and the potential vectors through which user data may be accessed, repurposed, or monetised. Thus, by connecting the more visible tracking behaviours at the user interface with the less visible infrastructures that support these practices, our work confronts the infrastructural reality that Cavoukian's (2011) idealized notion of privacy-by-design is not the norm in the contemporary app economy. Instead, we posit the dominance of an extraction-by-design model, wherein baby-tracking apps are structured to extract maximum value from the myriad ways parents use mobile applications. To better understand the relationship between app developers and the third-party services that enable access to highly sensitive health data in baby-tracking applications, we examine the mechanisms through which data about parents and newborns are accessed, and the extent to which users are offered meaningful opportunities to opt out or provide informed consent.

Methodology

We employed mixed methods (see Figure 1) to audit mobile tracking infrastructures across 14 baby-tracking applications. This included: (i) the app selection based on technical and commercial criteria; (ii) an assisted manifest data audit using ChatGPT4o and Claude Sonnet 4; (iii) a qualitative examination of corresponding privacy policies and data safety agreements; and (iv) the walkthrough method established by Light et al. (2018) to account for the different kinds of health data parents could input about their children in the apps' interface. We then analysed our findings to identify how third parties access personal and health data, while accounting for any discrepancies between the different methodological interventions we have deployed.

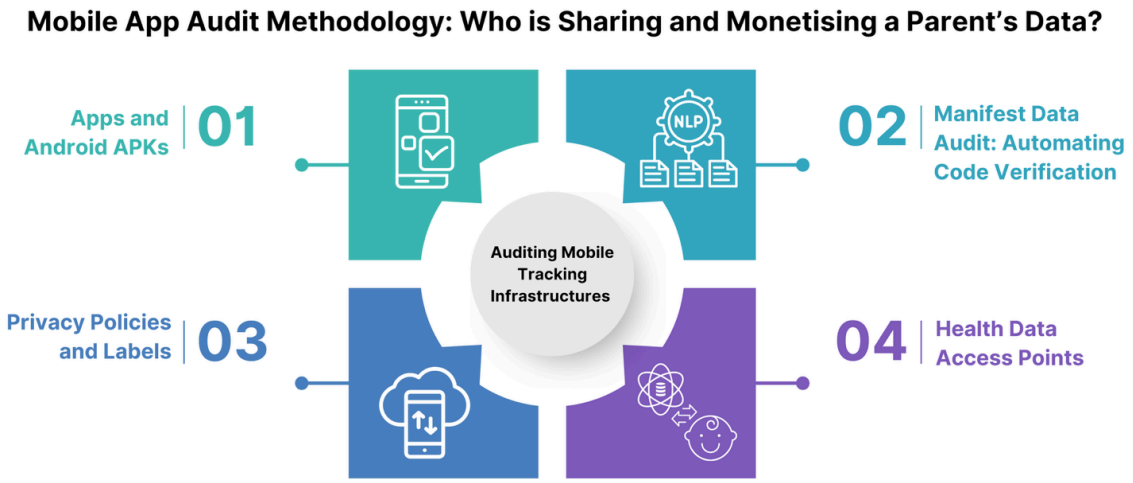

FIGURE 1: Mobile app audit methodology: who is sharing and monetising parent's data?

Baby-tracking applications and Android APKs

To select our applications, we performed a keyword search across the Google Play Stores in Europe and the UK for baby-tracking applications. We did not include iOS applications, as Android files or Android Packages (APKs) are open-source and accessible compared to Apple's closed ecosystem. We set a commercial criterion of at least one million downloads, with an exception for newer, AI-driven apps for which we used a cut-off of 10,000 downloads. This process resulted in 16 apps, but we discarded two due to inaccessible APK files. The final list is archived with the Open Science Framework (see Pybus et. al., 2026). Of our 14 apps, six advertised the use of AI to make real-time decisions, while another three offered algorithmically enhanced data visualisation tools to help parents understand and contextualise their data (See Table 1). Upon choosing our applications, we then accessed their APK files or source code from an open-source archive called APKPure (n.d.) and used an open-source decompiling tool, originally created by Google, called ClassyShark[1](Farber, 2020; Pybus, 2024). Our aim was to use this software to extract what is called an Android Manifest file for each application so we could examine which third-parties and services were being used and which privacy and advertising permissions they had enabled.

1. ClassyShark is an Android inspection tool, originally created by Google that lets you examine Android Package (APK) files. The software provides an accessible way to view the different components that make up mobile applications like classes, members, DEX files, native libraries, and most importantly for our research, manifest files (Farber, 2020). For a complete tutorial on how to decompile apps using ClassShark, see Pybus, 2024.

App	Google Play Downloads	Uses AI	AI Function	Non-AI Analysis Tools	Who Owns this App?
Baby Centre	10M+	No			Cliff Davis (US)
What to Expect	5M+	No		Milestone analysis	Cliff Davis (US)
Baby+	5M+	No			Philips Digital UK Limited (UK)
Baby Care	1M+	**Yes**	Health predictions		Wachanga (US)
Baby Daybook	1M+	No		Real time sleep predictions	Baby Daybrook, (Lithuania)
Baby Sparks	1M+	**Yes**	AI agent (Ava)		BabySparks (US)
Baby Tracker (Amila)	1M+	No			Amila (Austria)
Baby Tracker (Nighp)	1M+	No			Nighp Software LLC (US)
Baby Time	1M+	No			Simfler (South Korea)
Glow Baby	1M+	**Yes**	Milestones predictions		Glow (US)
Huckleberry	1M+	**Yes**	Sleep predictions		Huckleberry (US)
Bebememo	500K+	**Yes**	Face detection		Bebememo (US)
BabyMilestone	50K+	No		Cognitive development	Pathways (US)
Onoco	10K+	**Yes**	Sleep predictions		Onoco (UK)

FIGURE 2: Baby-tracking apps, downloads, and AI usage

The Android application manifest file is a useful document to audit because it acts as a kind of binding contract. Any personal data an application seeks to access from a user's device, or share with third parties, should be declared in this file. The manifest therefore documents the app's stated data access practices, including which third parties it integrates and what types of information may be accessed from an end user's phone (see Android Developers, n.d.). In this sense, the Android manifest resembles an airplane passenger manifest. Just as an aircraft manifest records every person on board, the Android manifest records each data access request that an app will make of a user's device. The challenge, however, is that these documents are written for machines rather than humans. We therefore engaged both ChatGPT4o (OpenAI) and Claude Sonnet 4 (Amazon) to support a qualitative reading of these files, which formed the foundation of our manifest data au-

dit.

The audit builds on the method developed by Pybus and Mir (2025) and Pybus and Coté (2024), enabling a granular investigation of the data ecosystems surrounding mobile applications. To conduct our analysis, we developed a series of prompts, drawing on best practices from the prompt engineering literature (Henrickson & Meroño-Peñuela, 2023; Marvin et al., 2024). We iteratively tested and refined each prompt until the models each produced consistent and repeatable results. We assessed accuracy by instructing the LLMs to "show with evidence," which allowed us to verify each claim directly in the original manifest file code through targeted keyword searches. Figure 2, for example, presents a screenshot from our ChatGPT4o analysis of *Baby Tracker* (NighP), where the LLM provides traceable evidence that can be used to locate the relevant entries in the manifest file, allowing the researchers to confirm that the findings were not hallucinated. In addition, we examined each company named inside the manifest to manually confirm their monetisation. In addition, the third parties were also cross checked in the APK file using ClassyShark, again to confirm any inconsistencies. As an aside, when we began this research we were not anticipating to find any health related metadata inside the manifest, however, after trialing the method, we realised that some of these events were in fact named. Thus, new prompts were developed and tested to account for these labels.

The key prompts we used to analyse the manifest files focused on the following indicators:

1. **The presence of SDKs and their associated services**, including tools for in-app profile building, market segmentation, campaign tracking, and attribution.
2. **The presence and enabled use of Advertising ID (AdID)**, a unique identifier that supports cross-app tracking and advertising-based profiling.
3. **The naming of specific intimate health activities, alarms, and intents**, such as *DueDateActivity*, *ReportLossActivity*, or *FeedingAlarm*, which correspond to functional components in the app that we examined in our walkthrough.
4. **The automatic enabling/disabling of app events** intended to collect and share behavioural and health data.
5. **Advertising permissions that enable interest-based profiling**, such as Google Topics, which are used to infer and share user interests with advertisers.

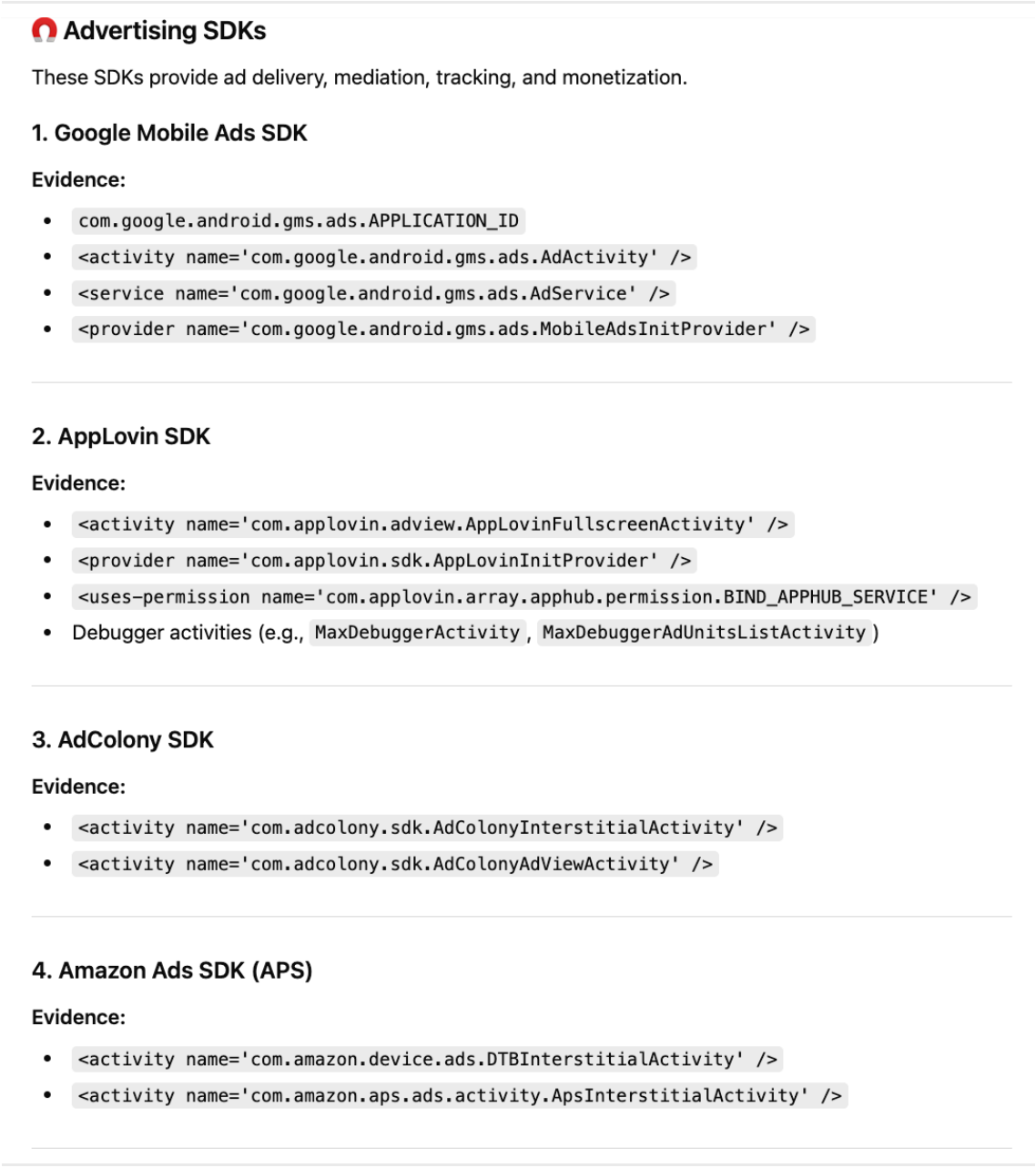

FIGURE 3: Screenshot from ChatGPT4o analysis of Baby Tracker (NighP)

To contextualise the findings of the manifest audit, we conducted a qualitative analysis of two key data governance documents that parents can access and read: (i) the *Data Safety Agreement*, a plain-language summary outlining the types of user data the app collects, shares and processes, along with the purposes for each activity; and (ii) the *Privacy Policy*, a more detailed and encompassing legal document. To search for inconsistencies, we examined how these documents represented what data was being shared about a parent's phone, including the model, operating system, and/or unique identifiers (software or hardware), or general use of the application. In addition, we sought evidence of profile building and monetisation, including the location of data processing and storage; evidence that data was

being used for in-app research and development, including to improve AI models and finally, the kinds of opt-out mechanisms provided to parents who preferred to limit data sharing. Central to this analysis was to look for inconsistencies between these two documents but more importantly to compare any discrepancies with what is represented to end users in the front-end with what we discovered in our manifest audit in the back-end.

Health data access points

To examine the various health data access points that prompt parents to disclose infant routine behaviours and pregnancy data, we employed the walkthrough method (Light et al., 2018). This qualitative framework systematically analyses an app's interface, features and user flows to uncover the sociocultural assumptions and governance logics embedded within its design. Accordingly, we downloaded and actively used each application in our sample to document the different kinds of sensitive data that parents are prompted to enter, such as what we can observe in Figure 3.

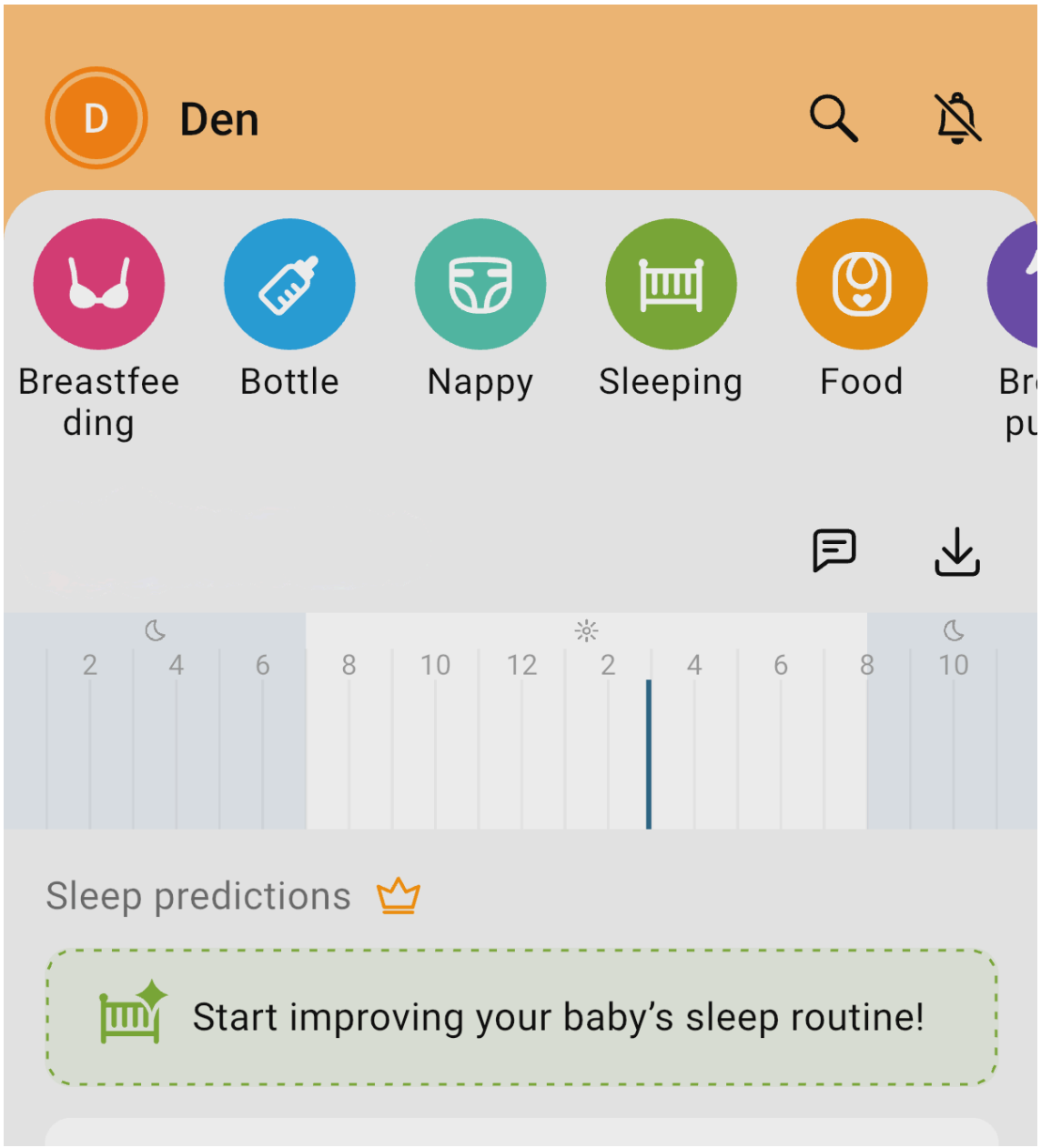

FIGURE 4: Screenshot of prompts from Baby Daybook app

Findings

Our audit of the 14 baby-tracking applications reveals extensive tracking, and with limited exception, a general failure to communicate the scope and scale of data collection to parents. Our findings also reveal a range of back-end sharing practices, highlighting that app developers can choose to limit data sharing. Moreover, this section draws attention to three key issues: first, the extensive ways in which parents share highly personal health data with a range of obfuscated advertising networks; second, how app store governance policies fall short of protecting infant data and third, the potential risks posed by app event tracking and related behav-

ioural profiling in the name of personalised advertising. Taken together, these findings point to a mobile app ecosystem in which intimate family health data is extracted by design, routinely commodified and globally circulated, often in what appears to be in contravention of GDPR regulations, with minimal transparency about the nature and extent of the data sharing.

Opt-in by default: how are parents sharing their children's data?

Our analysis of the Android manifest files demonstrates significant variation in two key areas: (i) the number of third party SDKs integrated to support monetisation strategies and (ii) the extent to which advertising identifiers (AdIDs) were enabled. These AdIDs, also known as Android Advertising IDs (AAIDs), are unique, user-resettable identifiers provided by Google Play Services. They allow advertisers to anonymously track user behaviour across apps, platforms, and devices (Google, n.d.-a), thereby supporting personalisation. When active, AdIDs make it easier for advertisers and data brokers to construct detailed, cross-referenced behavioural profiles of users by linking a parent's activity across multiple apps and devices. In doing so, they contribute to the enrichment of datasets and enable more precise targeting, both within the app and across wider digital ecosystems. While the iPhone's App Tracking Transparency (ATT) feature introduced in 2021, makes it much easier to disable AdIDs, in Android they continue to "enable a whole range of privacy harms" (Cyphers, 2022).[2]

To determine whether AdIDs were enabled, we used our LLM-assisted analysis as part of the manifest data audit, with results summarised in Figure 5. Figure 4 provides an illustrative example from Claude Sonnet 4 of this process, derived with the prompt: "Are any AdIDs enabled automatically in this app? Show with evidence from the manifest." Wachanga's *Baby Care*, for instance, sets Facebook's AdID to "true," meaning they are automatically enabled. *Glow Baby*, by contrast, sets it to "false," offering parents greater agency and control by requiring an explicit prompt during the app's set-up process.

2. While ATT has reduced the use of AdIDs in iPhone, Apple still allows the collection of over 20 other kinds of identifiers to make up for the loss of their own identifier for advertisers (IDFA) (Naprys, 2025).

Wachanga Baby Care App - Facebook SDK Advertising ID Collection

```xml
<meta-data
    name='com.facebook.sdk.AdvertiserIDCollectionEnabled'
    value='true'>
</meta-data>
```

Facebook Advertising ID Collection Setting in Glow Baby App

```xml
<meta-data name='com.facebook.sdk.AdvertiserIDCollectionEnabled' value='false'></meta-
```

FIGURE 5: Different choices app developers make, either setting AdIDs to true (on) or false (off) as seen in Claude Sonnet 4

The greater the number of embedded SDKs and enabled AdIDs, the more opportunities there are for third parties to access the data that parents produce within these apps. Figure 5 illustrates the stark differences in the decisions developers have made across our sample. At one end of the spectrum, *Baby Tracker* (NighP) uses ten SDKs, all enabled to collect data from the moment the app is launched, whereas *Baby Milestone* uses just one SDK from Google's Firebase. How these identifiers are configured offers parents with no straightforward way to opt-out of tracking, unless they already know how to adjust their phone's privacy settings through the operating system. But this expectation assumes a high level of digital literacy, placing an unreasonable burden on family members to anticipate and know how to mitigate these risks.[3] Indeed, only four apps in our sample, *Baby+*, *Baby Daybook*, *Glow Baby,* and *Baby Tracker (Amila)* did not configure their AdIDs automatically, making it easier for parents to disable their tracking services within the app. Conversely, this means that the other ten apps, representing 71% of our sample, had AdIDs automatically enabled, ready to share data with a range of third parties.

3. For an easy guide to turn off AdIDs in Android see Wallen (2024).

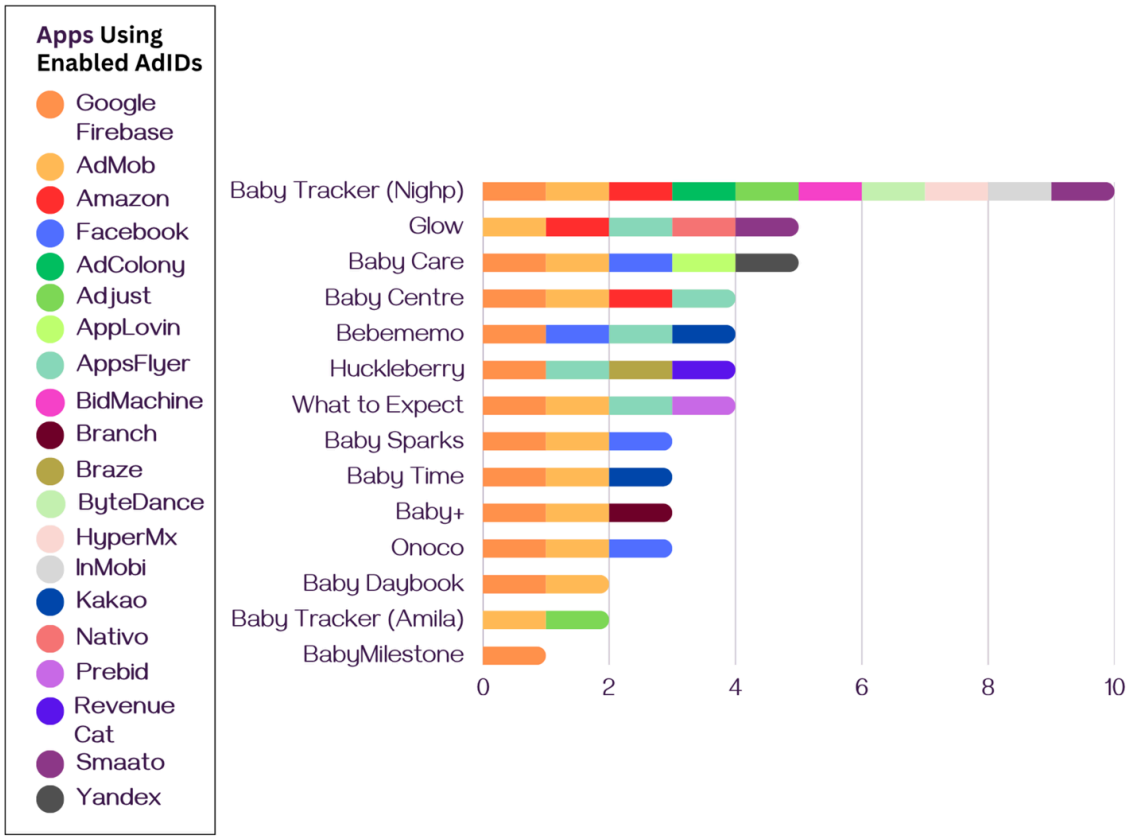

FIGURE 6: Visualisation of the automatically *enabled* advertising identifiers being used by baby-tracking apps

Next, we compared the findings from the manifest audit with each app's Data Safety Agreement, also known as the app's privacy labels, and its privacy policy, both of which are intended to support informed consent prior to downloading an app. When these documents were cross-referenced with the apps' manifest files, however, we identified notable differences between what developers disclosed between these two legal documents, versus what is technically observable in the manifest. For example, six apps listed in Table 2 claim in their privacy labels that they do not share personal data, while their privacy policies disclose data sharing with advertisers. Similarly, *Baby Time* and *Onoco* make no mention of third-party data sharing in either their DSAs or PPs, yet the manifest audit indicates active data sharing with platforms including Google and Facebook (see Figure 5). These inconsistencies are particularly significant given that the GDPR defines online identifiers, which include not only AdIDs, but a range of hardware and software identifiers, as personal data under Article 4(1), meaning that apps that fail to declare such practices may be in violation of this legislation (Regulation (EU) 2016/679). We also observed the widespread use of ambiguous language in privacy policies, such as claims that data "may be collected" or "may be shared." Such wording

makes it difficult to determine what is being collected or with whom it is shared, a practice that arguably conflicts with GDPR requirements under Article 12 for clear and transparent communication (Regulation (EU) 2016/679).

	Data Safety Agreement														PP	MA
Apps	App Activity and logs	Device Data and Phone IDs	Personal Information	Health and Fitness	Photos and Media	Location	Financial Info	App Activity and logs	Device Data and IDs	Personal Information	Health and Fitness	Photos and Media	Location	Financial Info	Personal Information	Personal Information
Baby Centre	x	x	x	x	x	x	-	x	x	x	x	x	x	-	x	x
Baby Care	x	x	x	x	x	x	-	x	x	x	x	x	x	-	x	x
What to Expect	x	x	x	x	x	-	-	x	x	x	x	x	-	-	x	x
Baby Tracker (NighP)	x	x	x	-	x	-	-	x	x	-	-	-	x	-	x	x
Baby Tracker (Amila)	x	x	-	-	-	-	-	x	x	-	-	-	x	-	x	x
Huckleberry	x	x	x	-	x	-	-	-	x	x	-	-	-	-	x	x
Baby Time	x	-	x	-	x	-	-	x	-	-	-	-	x	-	-	x
Baby Daybook	x	x	x	-	x	-	x	-	x	-	-	-	-	-	-	x
Bebememo	x	x	x	-	x	-	-	-	x	-	-	-	-	-	-	x
Glow	x	x	x	x	x	x	-	-	-	-	-	-	-	-	x	x
Baby+	x	x	x	-	-	-	-	-	-	-	-	-	-	-	x	x
Baby Sparks	x	x	x	-	-	-	-	-	-	-	-	-	-	-	x	x
Onoco	-	-	x	x	x	-	-	-	-	-	-	-	-	-	-	x
BabyMilestone	-	-	x	-	-	-	-	-	-	-	-	-	-	-	x	x
	Data Collected							Data Shared								

FIGURE 7: Data sharing disclosures across data safety agreements, privacy policies (PP), and manifest audits

Which platform policies can protect infant data?

Although Google Play requires apps to disclose relevant data practices in posted Data Safety labels, including data collection that occurs through third party SDKs (Google, n.d.-e), their requirements are partially dependent on the app's target audience. Apps categorised as "Parenting" typically include pregnancy, infant care and monitoring, and child care tools (Google, n.d.-b). Some of these apps target parents, such as the baby-tracking apps that we audited, however others target

children, such as Diveo Media's *Little Stories* or Disney's *Baby TV*. For the latter, developers must comply with the Google Play Families Policy and "disclose the collection of any personal and sensitive information from children" (Google, n.d.-d). The aim is to protect children from advertisers and other forms of profiling by safeguarding their data accessed by mobile applications. Subsequently, the infant data that parents share through baby-tracking applications falls into a policy grey area in which no additional protections are triggered, precisely because parents, rather than infants, are treated as the primary users of these apps. From a privacy protection perspective, a simple question follows: why would data belonging to infants only be protected if an actual infant is the recognised user? More to the point, given that baby-trackers rely on infant data to generate value, the Google Play Families Policy should arguably include them, especially if healthcare professionals are recommending these apps to parents (Thornham, 2019).

With this in mind, we assessed our sample against the Google Play Families Policy to determine whether developers had implemented more privacy-preserving safeguards for the data they were collecting, such as due dates, names, and baby habits. Apps that follow this policy should not be using or sharing certain device identifiers, which include: (i) *Android Advertising Identifiers* (AAID), or AdIDs, which enable behavioural profiling and advertising; (ii) *International Mobile Equipment Identity* (IMEI) numbers, hardware identifiers unique to a device that can be used to infer access to a phone number; and (iii) *Media Access Control* (MAC) addresses, unique software identifiers that can also be used to identify a specific device (Google, n.d.-c).

As shown in Table 3, none of the fourteen applications in our sample have taken the Google Play Families Policy into account. First, each app has automatically enabled AdIDs. Second, none of the privacy policies provide a complete list of the third parties accessing these identifiers. Third, when we examined other identifiers that apps are prohibited from accessing, specifically IMEIs and MAC addresses, we found that six were collecting at least one of these highly sensitive data points. Notably, *Onoco* and *Baby Care*, two apps that use AI to predict optimal sleep schedules and or interpret babies' health and development data, were collecting both. Moreover, as Table 2 shows, *Onoco* neither declared that it was collecting nor sharing any of these identifiers (see column 2). These potential violations, or at the very least a refusal to more meaningfully protect infant data, highlight the limitations of safeguarding such data only when children are treated as the primary users of an app, rather than protecting it more broadly. This analysis also underlines that profiles are being built about children from the moment of conception,

with limited policy interventions to prevent or limit these invasive data-gathering practices.

Apps	SDKs using AdIDs	IMEI	MAC ID
Baby Tracker (NighP)	10		x
Glow	5		
Baby Centre	4	x	
Baby Care	4	x	x
What to Expect	4	x	
Bebememo	4		
Huckleberry	3		
Baby Time	3		
Baby+	3	x	
Baby Sparks	3		
Onoco	3	x	x
Baby Tracker (Amila)	2		
Baby Daybook	2		
BabyMilestone	1		

FIGURE 8: How baby-tracking apps compare with Google Play Families Policy

Advertising infrastructure and the circulation of health data

While the manifest audit can reveal how apps access and share AdIDs, how health-related information is sold and resold within the broader app economy remains far less clear. In specific cases, such as *BabyCentre* and *What to Expect,* the sharing and selling of user (and baby) health data is explicitly disclosed. Both apps are owned by Everyday Health Group, a division of Ziff Davis and include a dedicated section outlining how "consumer health data" are shared with advertisers and other third-party affiliates. Both apps state that they "share [health] information with other Ziff Davis Companies and [their] partners for the purposes of targeted advertising" (*What to Expect, 2025; BabyCentre*, n.d.), openly acknowledging the sale of health data to advertisers:

> "We may also sell or transfer your [health] information to certain third parties, such as advertisers who will use this data for marketing purposes" (*What to Expect, 2025; BabyCentre*, n.d.).

Notably, these privacy policies are among the few we examined that provide a clear description of the sensitive nature of the data being monetised, including a user's due date:

> "We may share…your name, *due date*, postal address, email, other identifiers, or inferred data based on your interaction…One of the key ways in which this can occur is via *app events*, which are also automatically enabled" (*What to Expect*, 2025; *BabyCentre*, n.d.; emphasis added).

The disclosure of enabled app events communicates the presence of powerful infrastructural mechanisms that facilitate granular tracking of user interactions across the app's interface. While many SDKs come preloaded with a suite of standard app events, Firebase, for instance, includes over 500 (Firebase, n.d.), developers can also customise their own, based on the app's unique features (Pybus & Mir, 2025). To consider what data could potentially be shared with the apps in our sample, we used the walkthrough method (Light et al., 2018) to code possible health-related activities that could be transformed into app events (see Figure 6). While not all these data points will be shared and sold, the mere collection of this kind of user activity signals a troubling normalisation of the datafication of vast amounts of intimate behavioural data. Indeed, Table 2 shows that at least six apps are sharing app events while almost all of the apps are collecting them. Notably, three of the six applications that incorporate AI – *Baby Care*, *Huckleberry,* and *Glow Baby* – are leveraging this data for their own model development, with few mechanisms for opting out. This raises important questions about consent, especially when data is used to train and optimise proprietary AI systems.

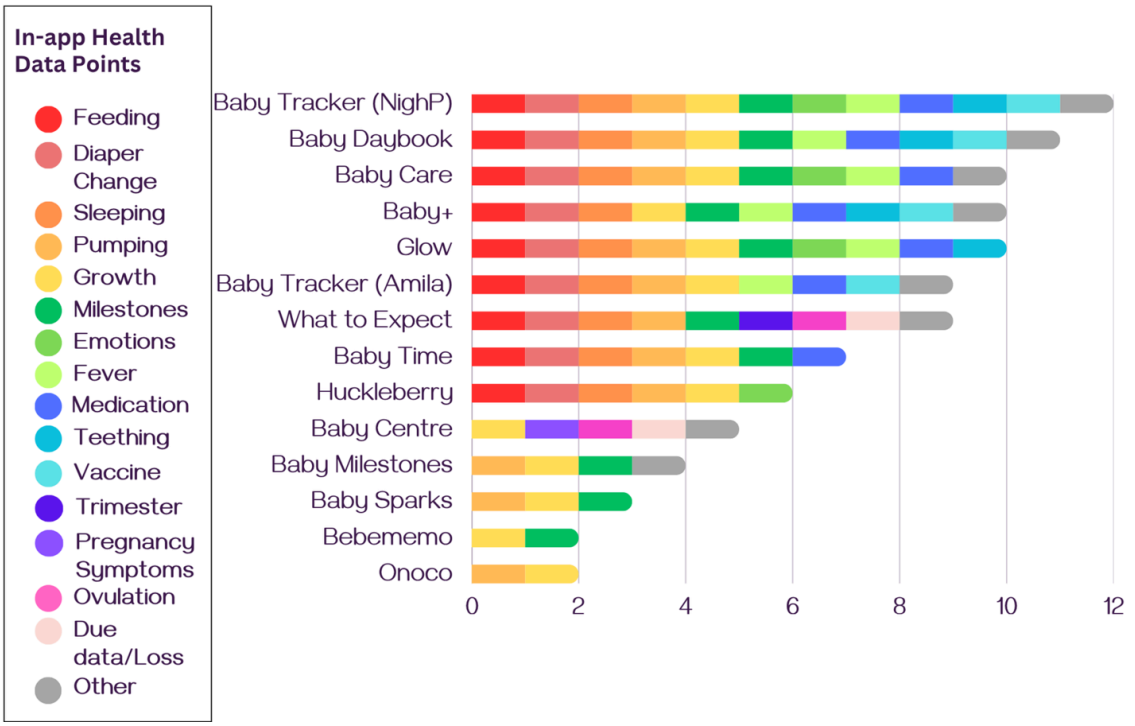

FIGURE 9: Visualisation of health-related data points discovered via the walkthrough method

To assess which apps were actively leveraging health data, we developed LLM prompts to audit the customised metadata labels that developers used to reference the screens a user views, the notifications they may receive, and or the actions that trigger different components within the app. In Android manifest files, developers declare core components such as activities, which can be used to name screens a user sees in an app; intent filters, which specify the actions these components can respond to; in addition to alarm-related declarations, including alarm-labelled actions that can enable scheduled reminders (Android Developers, n.d.). Most notably, the sharing of these labels, even anonymously through app events, can enable potentially consequential forms of behavioural tracking, as demonstrated in the cases against *Flo* and *Premom*. Several of these labels used by apps in our sample raise concerns. For example, *Glow Baby's* manifest includes an activity named "FeedingLogActivity;" *Baby Tracker* (NighP) includes customised alarms such as "AlarmReceiverPumped" and "AlarmReceiverFormula;" and the *What to Expect* app includes intents labelled as "miscarriage," "stillbirth," and "child loss." As shown in Table 4, the labelling of these app events appears to be an active choice made by developers, with only half of our sample using these highly specific health-related labels.

We argue that these naming practices warrant further research and question why these intimate labels are being used, and how parents can be assured that they are

not being shared with third parties. This is especially relevant in a historical moment where women in the US are being arrested following pregnancy loss, including miscarriages (Equal Justice Initiative, 2025), and where investigations in Poland have reportedly been initiated in cases involving miscarriage or suspected self-managed abortion (Human Rights Watch, 2023). These labels also potentially capture parents' passive engagement with apps, whereby actions such as opening a screen or clicking a notification become mechanisms for generating anonymised inferences that support exceptionally rich and sensitive behavioural profiling with user-generated health data.

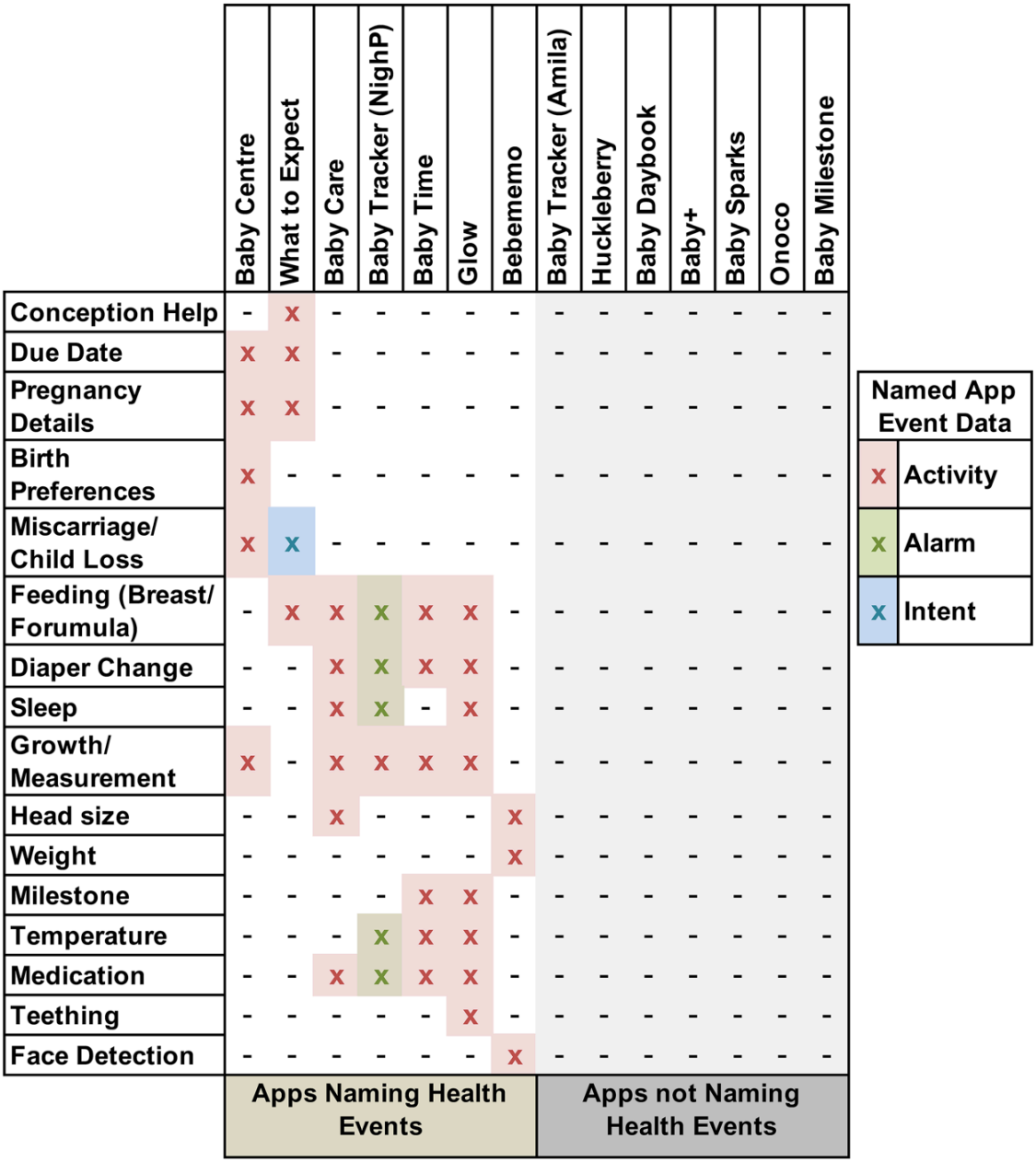

FIGURE 10: Apps with and without named health events

Opaque data flows: compromising data sovereignty

Cross border, intra corporate data flows seamlessly enabled by SDKs also raise broader concerns about data sovereignty, or the right of states and unions to determine how data generated within territorial boundaries is collected and governed, including when, where, and how it can be processed, used, sold, and stored (Polatin-Reuben & Wright, 2014). The privacy policy analysis revealed that six applications were transmitting data directly to servers based in the United States, even though the apps were downloaded via app stores in the EU and the UK. This practice could be justified through either or both Articles 45 and 46 of the GDPR (Regulation (EU) 2016/679), which permit international data transfers to private entities outside the EU when adherence to the Standard Contractual Clauses are assured (Heck & Jennings, 2024). However, we question whether it should ever be acceptable for intimate health data to be exported from the EU, especially for commercial purposes, especially in absence of routine audits and enforcement mechanisms off-shore. In terms of privacy policy analysis, five apps (BabyCentre, What to Expect, Glow Baby, Baby Milestone, and Baby Sparks) disclosed directly processing their data in the United States, however all had evidence in their respective manifests of third party SDKs from major US based platforms, including Google, Facebook, and Amazon. Additionally, some contained SDKs from Chinese (ByteDance), Indian (InMobi), Korean (Kakao), and Russian (Yandex) companies. In most cases the presence of SDKs from outside North America were either not disclosed or only partially disclosed in their privacy documents. To be clear, the national origin or domicile of third parties receiving sensitive data is not, in itself, a mandatory requirement under the GDPR. However, greater analysis, debate, and oversight are warranted, particularly as hostile political and social climates toward vulnerable groups intensify in the United States and elsewhere. In sum, we observed multiple instances, wherein European users' sensitive health data was being exploited by transnational advertising infrastructures, in nine instances without full disclosure.

Discussion

Our mixed-method app audit reveals how baby-tracking applications are designed to maximise data extraction in a range of ways. Parents who use these applications are sharing their infant's data with one or more advertising and analytics networks, often without meaningful consent or adequate protection. Of particular concern is the collection of health-related behavioural data through named app events, which can relate to habits around feeding, medication, vaccines, or miscarriage, just to name a few. The capture of this data is often obscured through vague

privacy labels, namely "app activity" (see Table 2) or disclosures buried deep within privacy policies. For instance, *BabyCentre* and *What to Expect* disclose third-party sharing of sensitive data, but do not describe that such data may include, for example, how frequently a mother views pages on postpartum depression, or the stage of pregnancy at which she accessed content related to fetal loss.

Similarly, in the case of *Baby Care*, owned by Wachanga, the app not only shares sensitive health data (Table 2) but also transmits named health event data (Table 4), shares AdIDs with five SDKs including the Russian platform Yandex (Figure 3), shares user location, and collects both software and hardware identifiers (Table 3). Like the Ziff Davis-owned apps, *Baby Care* also provides relatively clear privacy labels disclosing that health data is shared with third parties, although it says little about how this data is monetised. For us, this raises an important question: should apps like these be considered "compliant" simply because data-sharing practices align across privacy labels, privacy policies, and the app manifest, even when the information communicated remains opaque and the data sharing is extensive? For eight other apps (57% of our sample), we observed clear discrepancies in which access to personal data was not accurately represented in privacy labels, privacy policies, or both, compared with the evidence found in our manifest audit. This finding underscores the need for more stringent oversight in app stores, echoing conclusions from other studies of apps and privacy policies (Okoyomon et al., 2019; Kollnig, 2021; Malki et al., 2024). Yet regulation cannot rely on disclosure alone. Even if parents were given full transparency about how an app and its third-party SDKs collect device and app-usage data, would this be enough to make an informed decision about an app's benefits, harms, and risks? Policymakers should move beyond disclosure and take a more active role in protecting this sensitive data by pre-empting the many ways it can be exploited, regardless of how comprehensive privacy policies may appear.

Here we return to legal scholars such as Solove (2025) and Cohen (2019), who contend that extractive infrastructures such as SDKs produce collective privacy harms that exceed what individuals can meaningfully consent to or control. The use of machine learning to derive insights and inferences about end users has become commonplace in the app economy. This is facilitated in large part by SDK services such as Google's Firebase Analytics, which is present in over 73% of all Android applications (Statista, n.d.). Because the purpose of analytics services is to help developers generate profiles of users, platforms that operate these SDKs can continuously train on collective behavioural data they process across apps that have a service such as this one installed. From this perspective, privacy harms do

not emerge solely from what any one individual does or does not disclose, but from the concentration of relationally derived inferences generated across populations, communities, and vulnerable groups. In turn, the inferences used to profile a single person are derived from patterns learned from the data of millions of people (Coté & Aires, 2025). This relational dynamic reveals the limits of privacy frameworks that treat data protection as an individual responsibility rather than a collective condition, and that assume disclosure alone can meaningfully mitigate harm (Solove, 2025; Solow-Niederman, 2022; Cohen, 2019). Without understanding the downstream effects of sensitive infant and maternal health data sharing on vulnerable groups, how can families assess the potential harms that may befall them, their children, or others, because of their data sharing?

Moving forward, we propose a policy response that more adequately addresses the extraction of health data, grounded in our findings. Specifically, our audit demonstrates that infant and maternal health data is routinely captured through behavioural event data, persistent identifiers such as AdIDs, and third-party SDK sharing. Together, these create the social conditions for *inference-based profiling,* exemplified by the viral case of the American teenager whose pregnancy was revealed to her parents by targeted marketing from a store based on her recent purchases (Sag, 2025). Inference-based profiling is difficult for a user to evaluate for risks and harms because inferences are not easily understandable from the user's point-of-view. Future-oriented, collective privacy harms are also extremely difficult to anticipate or understand, especially where it pertains to contemplated discrimination against others. More to the point, inferences make it difficult for individuals to understand how decisions are made about them, especially when these are determined with proxy data that stand in for sensitive traits (Wachter & Mittlestadt, 2018).

The privacy threat of inference-based profiling necessitates stronger restrictions on the infrastructures that enable extraction, including limits on third-party advertising SDKs and the sharing of persistent identifiers in infant health apps, in addition to more robust and proactive enforcement of platform data governance rules. Unfortunately, changes to the GDPR in late 2025 have only eased privacy restrictions on data use to free up more European user data for AI training (Carpenter-Zehe, 2025). In terms of the privacy protection of families, Europe's regulatory environment remains ambiguous. Protection standards of the GDPR are eroding and cross-border data flows remain difficult to trace, and increasingly so, due to consolidating platform power. At the same time, voluntary privacy standards (for example, the Google Play Families Policy) fail to protect the sensitive health data of

infants from aggressive exploitation, because the expected app users are adults. As a result, infant and maternal health data are produced and circulated within a digital political economy that privileges extraction, rather than privacy-by-design.

Against this backdrop, alternative infrastructures are emerging, including the Data Governance Act's (Regulation (EU) 2022/868) model of data altruism. What would it mean to build infant health tracking infrastructures that prioritise parental needs and collective protection over commercial data extraction? At the very least, we hope this audit demonstrates the different approaches that app developers take in prioritising extraction vs. privacy-by-design, and can spotlight more privacy-friendly apps, such as *Baby Daybook, Baby+ and Baby Tracker (Amilia)* that limit their entanglement with infrastructures of datafication, without processing their data abroad.

Limitations

We acknowledge the limitations of our study. First, our audit focused on 14 Android applications from the EU and UK Google Play Store, yet privacy risks are also present in the iOS ecosystem and warrant examination. We also raise questions about the role of Apple's App Tracking Transparency and whether it contributes to meaningful differences between the two ecosystems. Second, our methodological approach relied on a static analysis; to more fully understand how app event data is transferred to external SDKs, future research will need to incorporate network and data analysis to build on our findings. Finally, we recognise the limitations of using large language models, namely ChatGPT4o and Claude Sonnet 4, to qualitatively audit complex manifest files that are ultimately written for operating systems rather than humans. While these generative AI tools significantly improved the interpretability of the files, the potential for hallucinations, bias and output variability remains (Pybus & Mir, 2025; Stahl & Eke, 2024). The environmental impact of LLMs is also not insignificant (Ren & Wierman, 2024), prompting further discussion on how app infrastructures might be audited in ways that are both accessible and sustainable to support critical engagement with these socio-technical ecosystems.

Conclusion

To conclude, our findings reveal that the most popular and advanced baby-monitoring applications available in the European Union and the United Kingdom are built on harmful infrastructures of datafication that are globally distributed, largely opaque, and deeply embedded in everyday care routines. These infrastructures

make it almost impossible for parents not to share intimate details about themselves and their children with third parties, including health-related information that in formal medical contexts would be subject to strict privacy protections. The audit revealed inadequate privacy disclosures and consent-gathering mechanisms, failures to adhere to platform privacy policies, third-party data sharing enabled from the moment the app is launched, and app event tracking systems that invisibly capture behavioural data in ways that can be used in consequential or exploitative ways. Taken together, these practices reflect a political economy rooted in extraction-by-design that facilitates global data sharing. In our sample, this included four apps sharing data with advertising networks in China and India (*Baby Tracker NighP*), Russia (*Baby Care*), and South Korea (*Babymemo* and *Baby Time*), and six apps (*BabyCentre, What to Expect, Baby Care, Baby Milestone, Baby Sparks,* and *Glow Baby*) exporting European and UK user data to US servers for processing. The audit therefore underscores the challenges of current data protections across the mobile app ecosystem, especially for self-tracking apps that are processing health data. Parents deserve stronger privacy regulation that recognises data-gathering and exploitation as a matter of public interest, rather than personal responsibility. As baby-tracking apps increasingly turn to generative AI to bolster their marketability, the stakes for safeguarding sensitive data have never been higher.

References

Adjust. (n.d.). *What are app events?* Adjust. Retrieved 11 January 2026, from https://www.adjust.com/glossary/events/

Almeida, T., Shipp, L., Mehrnezhad, M., & Toreini, E. (2022). Bodies like yours: Enquiring data privacy in femtech. *Adjunct Proceedings of the 2022 Nordic Human-Computer Interaction Conference*, 1–5. https://doi.org/10.1145/3547522.3547674

Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. (2024). A world model: On the political logics of generative AI. *Political Geography*, *113*, 103134. https://doi.org/10.1016/j.polgeo.2024.103134

Android Developers. (n.d.). *Android mobile app developer tools*. Android Developers. Retrieved 11 January 2026, from https://developer.android.com/

APKPure. (n.d.). *APKPure: Download APK on Android with free APK downloader*. APKPure. Retrieved 11 January 2026, from https://apkpure.com/

Baby Care. (n.d.). *Breastfeeding Tracker*. Google Play. Retrieved 12 January 2026, from https://play.google.com/store/apps/details?id=com.wachanga.babycare&hl=en

BabyCenter. (n.d.). *BabyCenter*. BabyCenter. Retrieved 27 June 2025, from https://www.babycenter.com

Barassi, V. (2017). BabyVeillance? Expecting parents,online surveillance and the cultural specificity

of pregnancy apps. *Social Media + Society*, *3*(2). https://doi.org/10.1177/2056305117707188

Barassi, V. (2020). *Child data citizen: How tech companies are profiling us from before birth*. The MIT Press. https://doi.org/10.7551/mitpress/12415.001.0001

Bechmann, A. (2014). Non-informed consent cultures: Privacy policies and app contracts on Facebook. *Journal of Media Business Studies*, *11*(1), 21–38. https://doi.org/10.1080/16522354.2014.11073574

Binns, R., Zhao, J., Kleek, M. V., & Shadbolt, N. (2018). Measuring third-party tracker power across web and mobile. *ACM Transactions on Internet Technology*, *18*(4), 1–22. https://doi.org/10.1145/3176246

Birch, K., & Muniesa, F. (Eds). (2020). *Assetization: Turning things into assets in technoscientific capitalism*. The MIT Press. https://doi.org/10.7551/mitpress/12075.001.0001

Blanke, T., & Pybus, J. (2020). The material conditions of platforms: Monopolization through decentralization. *Social Media + Society*, *6*(4). https://doi.org/10.1177/2056305120971632

Bounegru, L., Devries, M., & Weltevrede, E. (2022). The research persona method: Figuring and reconfiguring personalised information flows. In C. Lury, W. Viney, & S. Wark (Eds), *Figure* (pp. 77–104). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-2476-7_5

Carpenter-Zehe, O. (2025). *The Digital Omnibus has arrived—And here's what it really changes*. EUobserver. https://euobserver.com/digital/areee4d315

Cavoukian, A. (2011). *Privacy by design: The 7 foundational principles*. Slate. https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf

Citron, D. K. (n.d.). *The end of Roe means we need a new civil right to privacy*. Slate. Retrieved 10 January 2026, from https://slate.com/technology/2022/06/end-roe-civil-right-intimate-privacy-data.html

Cohen, J. E. (2019). Turning privacy inside out. *Theoretical Inquiries in Law*, *20*(1), 1–31. https://doi.org/10.1515/til-2019-0002

Cohen, J. E. (2023). Infrastructuring the digital public sphere. *Yale Journal of Law and Technology*, *25*(Special Issue), 1–40.

Coté, M., & Aires, S. (2025). *Futurity as infrastructure: A techno-philosophical interpretation of the AI lifecycle* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2508.15680

Cyphers, B. (2022). How to disable Ad ID tracking on iOS and Android and why you should do it now. *Electronic Frontier Foundation*. https://www.eff.org/deeplinks/2022/05/how-disable-ad-id-tracking-ios-and-android-and-why-you-should-do-it-now

Equal Justice Initiative. (2025). *Georgia woman arrested after miscarriage amid growing criminalization of pregnancy*. Equal Justice Initiative. https://eji.org/news/georgia-woman-arrested-after-miscarriage-amid-growing-criminalization-of-pregnancy/

European Parliament and Council. (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*. (OJ L 119, 4.5.2016, pp. 1–88). https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament and Council. (2022). *Regulation (EU) 2022/868 of the European Parliament and*

of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). (OJ L 152, 3.6.2022, pp. 1–44). https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng

Farber, B. (2020). *Classyshark [Java]*. https://github.com/google/android-classyshark

Felsberger, S. (with Apollo-University Of Cambridge Repository & University Of Cambridge). (2025). *The high stakes of tracking menstruation*. Minderoo Centre for Technology and Democracy. https://doi.org/10.17863/CAM.118325

Firebase. (n.d.). *Log events*. Google Analytics. Retrieved 8 January 2026, from https://firebase.google.com/docs/analytics/events?platform=ios

Flensburg, S., & Lai, S. S. (2022). Datafied mobile markets: Measuring control over apps, data accesses, and third-party services. *Mobile Media & Communication*, *10*(1), 136–155. https://doi.org/10.1177/20501579211039066

Frasco v. Flo Health, Inc. et Al., No. 3:21-cv-00757 (N.D. Cal. 23 February 2021). https://www.classaction.org/media/frasco-et-al-v-flo-health-inc-et-al.pdf

Gilman, M. E. (2021). Periods for profit and the rise of menstrual surveillance. *Columbia Journal of Gender and Law*, *41*(1), 100–113. https://doi.org/10.52214/cjgl.v41i1.8824

Glow Baby. (n.d.). *Baby tracker app*. Google Play. Retrieved 12 January 2026, from https://play.google.com/store/apps/details?id=com.glow.android.baby&hl=en_GB

Google. (n.d.-a). *Ad identifiers*. Google Play Console Help. Retrieved 30 June 2025, from https://support.google.com/googleplay/android-developer/answer/6048248?hl=en-GB

Google. (n.d.-b). *Choose a category and tags for your app or game*. Google Play Console Help. Retrieved 10 January 2026, from https://support.google.com/googleplay/android-developer/answer/9859673?hl=en#zippy=%2Capps

Google. (n.d.-c). *Data practices in Families apps*. Google Play Console Help. Retrieved 11 January 2026, from https://support.google.com/googleplay/android-developer/answer/11043825?hl=en

Google. (n.d.-d). *Google Play Families Policy*. Google Play Console Help. Retrieved 30 June 2025, from https://support.google.com/googleplay/android-developer/answer/9893335?hl=en-GB

Google. (n.d.-e). *Provide information for Google Play's Data safety section*. Google Play Console Help. Retrieved 30 June 2025, from https://support.google.com/googleplay/android-developer/answer/10787469?sjid=2921289184869994997-EU#

Heck, Z., & Jennings, J. (2024). *Another update already? New EU standard contractual clauses on the horizon to further safeguard cross border data transfers*. Taft Privacy & Data Security Insights. https://www.privacyanddatasecurityinsight.com/2024/09/another-update-already-new-eu-standard-contractual-clauses-on-the-horizon-to-further-safeguard-cross-border-data-transfers/

Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media + Society*, *1*(2). https://doi.org/10.1177/2056305115603080

Henrickson, L., & Meroño-Peñuela, A. (2025). Prompting meaning: A hermeneutic approach to optimising prompt engineering with ChatGPT. *AI & SOCIETY*, *40*(2), 903–918. https://doi.org/10.1007/s00146-023-01752-8

Huckleberry. (n.d.). *Expert sleep help for all families*. Huckleberry Care. Retrieved 11 January 2026,

from https://huckleberrycare.com/

Human Rights Watch. (2023). *Human rights crisis: Abortion in the United States after Dobbs*. Human Rights Watch. https://www.hrw.org/news/2023/04/18/human-rights-crisis-abortion-united-states-after-dobbs

Ius Omnibus v Flo Health, Inc. Retrieved 10 January 2026, from https://iusomnibus.eu/ius-omnibus-v-flo-health-inc/

Jena, H. (2022). *You're going to need to track your baby's feedings, diaper changes, sleep and more. Here's how*. What to Expect. https://www.whattoexpect.com/first-year/baby-feeding/what-to-expect-baby-feeding-tracker

Kollnig, K. (2021). *Tracking in apps' privacy policies* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2111.07860

Kuntsman, A., & Miyake, E. (2022). *Paradoxes of digital disengagement: In search of the opt-out button*. University of Westminster Press. https://doi.org/10.16997/book61

Langton, K., & Ng, R. (2025). "Tracking the Trackers" of children's first personal data in mobile applications: Using static analysis and privacy policy evaluation to explore the data-sharing capabilities and practices of baby apps. *Proceedings of the 37th Australian Conference on Human-Computer Interaction*, 845–859. https://doi.org/10.1145/3764687.3769936

Leaver, T. (2015). Born digital? Presence, privacy, and intimate surveillance. In J. Hartley & W. Qu (Eds), *Re-orientation: Translingual transcultural transmedia. Studies in narrative, language, identity, and knowledge* (pp. 149–160). Fudan University Press.

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Livingstone, S., & Blum-Ross, A. (2020). *Parenting for a digital future: How hopes and fears about technology shape children's lives* (1st edn). Oxford University Press. https://doi.org/10.1093/oso/9780190874698.001.0001

Lomborg, S., Sick, K., Flensburg, S., & Sophus Lai, S. (2024). Monitoring infrastructural power: Methodological challenges in studying mobile infrastructures for datafication. *Internet Policy Review*, *13*(2). https://doi.org/10.14763/2024.2.1763

Lupton, D. (2016). The diverse domains of quantified selves: Self-tracking modes and dataveillance. *Economy and Society*, *45*(1), 101–122. https://doi.org/10.1080/03085147.2016.1143726

Lupton, D. (2019). 'It's made me a lot more aware': A new materialist analysis of health self-tracking. *Media International Australia*, *171*(1), 66–79. https://doi.org/10.1177/1329878X19844042

Lupton, D. (2020). Caring dataveillance: Women's use of apps to monitor pregnancy and children. In *The Routledge companion to digital media and children*. Routledge.

Malki, L. M., Kaleva, I., Patel, D., Warner, M., & Abu-Salma, R. (2024). Exploring privacy practices of female mHealth apps in a post-Roe world. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–24. https://doi.org/10.1145/3613904.3642521

Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2024). Prompt engineering in large language models. In I. J. Jacob, S. Piramuthu, & P. Falkowski-Gilski (Eds), *Data intelligence and cognitive informatics* (pp. 387–402). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-7962-2_30

Mascheroni, G., & Siibak, A. (2021). *Datafied childhoods*. MIT Press. https://www.peterlang.com/document/1140627

McMillan, C. (2022). Monitoring female fertility through 'femtech': The need for a whole-system approach to regulation. *Medical Law Review*, *30*(3), 410–433. https://doi.org/10.1093/medlaw/fwac006

Mehrnezhad, M., Van Der Merwe, T., & Catt, M. (2024). Mind the femtech gap: Regulation failings and exploitative systems. *Frontiers in the Internet of Things*, *3*, 1296599. https://doi.org/10.3389/friot.2024.1296599

Naprys, E. (2025). *When you ask an app "not to track" it tracks you anyway, and the data is sold*. Cybernews. https://cybernews.com/privacy/aps-track-despite-asking-not-to-track/

Neff, G., & Nafus, D. (2016). *Self-tracking*. MIT Press. https://mitpress.mit.edu/9780262529129/self-tracking/

Nieborg, D. B., & Poell, T. (2025). Analyzing institutional platform power: Evolving relations of dependence in the mobile digital advertising ecosystem. *New Media & Society*, *27*(4), 1909–1927. https://doi.org/10.1177/14614448251314405

Nieborg, D., Poell, T., Caplan, R., & Van Dijck, J. (2024). Introduction to the special issue on Locating and theorising platform power. *Internet Policy Review*, *13*(2). https://doi.org/10.14763/2024.2.1781

Okoyomon, E., Samarin, N., Wijesekera, P., Elazari Bar On, A., Vallina-Rodriguez, N., Reyes, I., Feal, Á., & Egelman, S. (2019). On the ridiculousness of notice and consent: Contradictions in app privacy policies. *Workshop on Technology and Consumer Protection (ConPro 2019), in Conjunction with the 39th IEEE Symposium on Security and Privacy*. https://dspace.networks.imdea.org/handle/20.500.12761/690

Ostherr, K., Borodina, S., Bracken, R. C., Lotterman, C., Storer, E., & Williams, B. (2017). Trust and privacy in the context of user-generated health data. *Big Data & Society*, *4*(1), 2053951717704673. https://doi.org/10.1177/2053951717704673

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. https://doi.org/10.1177/1461444816661553

Polatin-Reuben, D., & Wright, J. (2014). *An internet with BRICS characteristics: Data sovereignty and the Balkanisation of the internet*. https://www.researchgate.net/publication/309585614_An_Internet_with_BRICS_Characteristics_Data_Sovereignty_and_the_Balkanisation_of_the_Internet

Pybus, J. (2024). *SDK Data Audit* [Data set]. Open Science Framework. https://osf.io/w96tq

Pybus, J., & Coté, M. (2022). Did you give permission? Datafication in the mobile ecosystem. *Information, Communication & Society*, *25*(11), 1650–1668. https://doi.org/10.1080/1369118X.2021.1877771

Pybus, J., & Coté, M. (2024). Super SDKs: Tracking personal data and platform monopolies in the mobile. *Big Data & Society*, *11*(1). https://doi.org/10.1177/20539517241231270

Pybus, J., Lachmansingh, A., & Matheson, K. (2026). *Baby-tracking app APK files* [Data set]. Open Science Framework. https://osf.io/c7n6t

Pybus, J., Lomborg, S., Gandini, A., & Lai, S. S. (2025). Empirical approaches to infrastructures for datafication: Introduction to the special issue. *New Media & Society*, *27*(4), 1851–1867. https://doi.or

g/10.1177/14614448251314396

Pybus, J., & Mir, M. (2025). Tracking menopause: An SDK Data Audit for intimate infrastructures of datafication with ChatGPT4o. *New Media & Society*, *27*(4), 1888–1908. https://doi.org/10.1177/14614448251314401

Ren, S., & Wierman, A. (2024). The uneven distribution of AI's environmental impacts. *Harvard Business Review*. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

Sag, M. (2025). *A Student's Guide to the Law and Policy of AI: Interests and Concerns motivating AI Regulation*. SSRN. https://doi.org/10.2139/ssrn.5958915

Scatterday, A. (2021). This is no ovary-action: Femtech apps need stronger regulations to protect data and advance public health goals. *North Carolina Journal of Law & Technology*, *23*(3), 636–668.

Schäfke-Zell, W. (2022). Revisiting the definition of health data in the age of digitalized health care. *International Data Privacy Law*, *12*(1), 33–43. https://doi.org/10.1093/idpl/ipab025

Sherman, J. (2023). *The FTC, fertility app Premom and sharing consumer health data*. Lawfare. https://www.lawfaremedia.org/article/the-ftc-fertility-app-premom-and-sharing-consumer-health-data

Shipp, L., & Blasco, J. (2020). How private is your period?: A systematic analysis of menstrual app privacy policies. *Proceedings on Privacy Enhancing Technologies*, *2020*(4), 491–510. https://doi.org/10.2478/popets-2020-0083

Solove, D. J. (2025). Artificial intelligence and privacy. *Florida Law Review*, *77*(1), 1–73.

Solow-Niederman, A. (2022). Information privacy and the inference economy. *Northwestern University Law Review*, *117*(2), 357–424.

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, *74*, 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700

Statista. (n.d.). *Android top mobile app analytics SDKs 2025*. Statista. Retrieved 10 January 2026, from https://www.statista.com/statistics/1035612/leading-mobile-app-analytics-sdks-android/

Steinberg, S. (2017). Sharenting: Children's privacy in the age of social media. *Emory Law Journal*, *66*(4), 839–880.

Thornham, H. (2019). Algorithmic vulnerabilities and the datalogical: Early motherhood and tracking-as-care regimes. *Convergence: The International Journal of Research into New Media Technologies*, *25*(2), 171–185. https://doi.org/10.1177/1354856519835772

United States of America v. Easy Healthcare, No. 1:23-cv-3107 (United States District Court, Northern District of Illinois 2023). https://www.ftc.gov/system/files/ftc_gov/pdf/2023186easyhealthcarecomplaint.pdf

Van Der Vlist, F. N., & Helmond, A. (2021). How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society*, *8*(1). https://doi.org/10.1177/20539517211025061

Van Dijck, J., Nieborg, D., & Poell, T. (2019). Reframing platform power. *Internet Policy Review*, *8*(2). https://doi.org/10.14763/2019.2.1414

Wachter, S., & Mittelstadt, B. (2018). A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI. *Columbia Business Law Review*, *2019*(2). https://papers.ssrn.com/a

bstract=3248829

Wallen, J. (2024). *Sick of ads on Android? Change these 5 settings for more privacy—Fast*. ZDNET. https://www.zdnet.com/article/sick-of-ads-on-android-change-these-5-settings-for-more-privacy

Wernimont, J. (2019). *Numbered lives: Life and death in quantum media*. The MIT Press. https://doi.org/10.7551/mitpress/11455.001.0001

What to Expect. (2025). *What to Expect privacy policy*. What to Expect. https://www.whattoexpect.com/privacy-policy

Witzleb, N., Paterson, M., Wilson-Otto, J., Tolkin-Rosen, G., & Marks, M. (2020). *Privacy risks and harms for children and other vulnerable groups in the online environment* [Research paper]. Office of the Australian Information Commissioner. https://www.oaic.gov.au/__data/assets/pdf_file/0012/11136/Report-Privacy-risks-and-harms-for-children-and-other-vuln

Published by

Alexander von Humboldt Institute for Internet and Society

CAIS Research for the Digital Age

in cooperation with

CREATe

cnrs centre internet et societe

R&I IN3 Internet interdisciplinary Institute Universitat Oberta de Catalunya

University of Tartu Johan Skytte Institute of Political Studies