Navigating the Data Avalanche: Towards Supporting Developers in Developing Privacy-Friendly Children's Apps

ANIRUDH EKAMBARANATHAN, University of Oxford, United Kingdom
JUN ZHAO, University of Oxford, United Kingdom
GEORGE CHALHOUB, University of Oxford, United Kingdom

This paper critically examines the intersection of privacy concerns in children's apps and the support required by developers to effectively address these concerns. Third-party libraries and software development kits (SDKs) are widely used in mobile app development, however, these libraries are commonly known for posing significant data privacy risks to users. Recent research has shown that app developers for children are particularly struggling with the lack of support in navigating the complex market of third-party SDKs. The support needed for developers to build privacy-friendly apps is largely understudied. Motivated by the needs of developers and an empirical analysis of 137 'expert-approved' children's apps, we designed DataAvalanche.io, a web-based tool to support app developers in navigating the privacy and legal implications associated with common third-party SDKs on the market. Through semi-structured interviews with 12 app developers for children, we demonstrate that app developers largely perceive the transparency supported by our tool positively. However, they raised several barriers, including the challenges of adopting privacy-friendly alternatives and the struggle to safeguard their own legal interests when facing the imbalance of power in the app market. We contribute to our understanding of the open challenges and barriers faced by app developers in creating privacy-friendly apps for children and provide critical future design and policy directions.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: privacy, developers, children, tracking, apps

ACM Reference Format:
Anirudh Ekambaranathan, Jun Zhao, and George Chalhoub. 2023. Navigating the Data Avalanche: Towards Supporting Developers in Developing Privacy-Friendly Children's Apps. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 2, Article 53 (June 2023), 25 pages. https://doi.org/10.1145/3596267

1 INTRODUCTION

Young children have become avid consumers of media and information through mobile devices, such as smartphones and tablets. They are spending more time on digital devices now than ever before [16, 43, 51]. According to recent studies, the vast majority of children in the United Kingdom between the ages of 3 and 17 have spent time online, primarily through mobile devices such as smartphones (72%) and tablets (69%) [17]. It has been reported that 82% of

Authors' addresses: Anirudh Ekambaranathan, University of Oxford, Oxford, United Kingdom, anirudh.ekam@cs.ox.ac.uk; Jun Zhao, University of Oxford, Oxford, United Kingdom, jun.zhao@cs.ox.ac.uk; George Chalhoub, University of Oxford, Oxford, United Kingdom, george.chalhoub@cs.ox.ac.uk.

Manuscript submitted to ACM

children between the ages of 5 and 7 spend almost 10 hours per week on the internet and a significant number of children under the age of 13 even have social media profiles [17, 51].

However, during the course of the past decade, the use of mobile apps have been under an increasing amount of scrutiny due to their large scale and systematic practice of collecting and sharing sensitive personal data [12, 14, 41, 57]. Seemingly, data tracking and loss of privacy is inevitable in the mobile ecosystem, mostly due to the presence of third-party libraries and the use of advertising SDKs [12], which are deemed an essential part of development [8, 24, 26].

As a result, privacy issues associated with large scale data tracking has become a priority for regulatory bodies responsible for protecting consumers' data and privacy. For example, the EU introduced the General Data Protection Regulation (GDPR) [67] in 2018, and in the UK the Information Commissioner's Office (ICO) introduced the Age Appropriate Design Code (AADC) [1] in 2021, forcing products and services aimed at children to prioritise data protection as an integral part of their offering. Importantly, the UK AADC takes the stance that apps aimed at children should be transparent about the privacy practices in a way that makes it accessible for children.

These recent developments have led us to examine the intersection of two related phenomena. First, prominent app marketplaces, like Google Play Store and Apple's App Store, feature special categories dedicated to children's apps. For instance, Google offers an 'expert-approved' children's app category [54], which is deemed safe and appropriate by child and media experts. However, the privacy practices of these apps and their adherence to new regulations, such as the UK AADC, remain unclear. Second, while the children's app market flourishes, developers are inadequately supported and face increasing pressure to prioritise privacy in their development practices. Recent studies reveal that developers struggle to understand the privacy practices of third-party libraries and SDKs [23, 24], translate legal requirements into technical ones [11, 65], and present legal implications understandably to children with varying digital literacy levels [68]. Our analysis primarily focuses on the UK app marketplace due to the comprehensive children's privacy framework provided by the AADC, which is not addressed by other regulations.

Being able to navigate the vast landscape of third-party libraries and SDKs has been identified as one critical challenge for developers who wish to build privacy-friendly apps for the users, or be legally compliant [24]. Third-party libraries and SDKs account for a major part of current mobile app development. These libraries are used to simplify the development process, improve security, and add additional functionality to apps [26]. However, research has shown that the use of such third-party libraries often pose substantial data tracking privacy risks to the users, particularly to children [13, 34, 49]. Third-party libraries often share various types of users' data with third-party tracker companies, to provide core app functions as well as (more often) generate personalised digital promotions for the users. Although app developers are largely aware of these practices, they still struggle to navigate the complex technical landscape of third-party libraries [23], many of which are recommended by leading app market providers or associated with deeply intertwined tracker networks [24]. Tools to aid developers are scarce and are mostly limited to generating privacy policies [50, 71], creating in-app privacy annotations [40], or focus on specific data types in apps (e.g., location) [39]. There is a lack of tools to holistically support app developers navigating the use of third-party SDKs and trackers. In light of this, in this study we aim to answer the following research questions:

RQ1 How does a tool that provides increased transparency about privacy implications of third-party libraries affect app developers' perception of their existing development practices in relation to privacy?

RQ2 What challenges and barriers do developers face in making privacy-friendly development choices and how can they be supported?

To answer these questions, we designed DataAvalanche.io, a web-based tool to provide transparency on the privacy practices of commonly used third-party libraries and SDKs. The design of DataAvalanche.io is motivated by a twofold set of considerations. First, we were guided by existing research that highlights the critical barriers faced by developers to create privacy-friendly apps, which include the ability to navigate the technical privacy implications of third-party libraries as well as the legal implications of making certain choices of these libraries. Second, we conducted an analysis of 137 'expert-approved' children's apps from the Google Play Store, to identify the technical and legal privacy implications associated with third-parties libraries used by these expert-approved apps. The empirical data resulting from these design considerations helped us inform the three key components of our tool:

(1) A **transparency** component for users to examine third-party trackers that can be associated with each SDK.
(2) Recommendations for **privacy-friendly alternatives** which developers can incorporate in their apps.
(3) A **legal** component that provides easy access to privacy policies of the SDKs and highlights any potential violations to the GDPR.

We evaluated DataAvalanche.io through semi-structured interviews with 12 children's app developers. Our study revealed that developers appreciated the transparency provided by our tool but also faced several barriers, such as challenges in adopting privacy-friendly alternatives and safeguarding their legal interests due to the power imbalance in the app market. This research contributes to our understanding of the open challenges and barriers app developers encounter when creating privacy-friendly children's apps and offers critical future design and policy directions. Specifically, this paper makes the following contributions:

- Firstly, we present the design of DataAvalanche.io, a tool that assists app developers in navigating third-party libraries. The tool's design is grounded in empirical evidence collected through studies with developers and children's apps.
- Secondly, we offer insights into app developers' perceptions of our tool prototype's effectiveness, identifying key design opportunities to support the creation of privacy-friendly apps.
- Thirdly, we conduct a rigorous privacy analysis of children-specific apps, demonstrating that popular apps promoted by app stores do not necessarily prioritise privacy.

2 BACKGROUND

2.1 Children & Privacy in the App Ecosystem

Despite the fact that children nowadays grow up in a digital environment, they do not understand digital risks as well as most adults [43], and have a poor understanding of privacy related contexts [36, 42, 70]. Privacy is a crucial element for child development and plays an important role in developing skills such as responsibility, trust, autonomy, and critical thinking [9, 53, 55, 56, 69]. The digital activities and platforms that children engage with provide opportunities for them to build these skills. As children grow older, their understanding and desire for privacy also increases [20, 36, 60]. Research has shown that children particularly struggle to fully understand *implicit/invisible data privacy risks*, e.g., how and why their data is collected by third parties [3, 25, 37], how it is processed [15], or how it can be used in the future [15, 48, 52]. While children value their privacy online [36, 62], they have grown to accept targeted advertising and analytics as part of everyday life [37], without being able to change their behaviours to affect this [37, 52].

2.2 Data Tracking

Mobile apps form a specific risk to children's privacy, because of their use of third-party libraries, which are unavoidable in today's landscape of app development [8]. Third-party libraries are widely used in the practice of app developments to simplify the development process, improve security, or add additional functionality to apps [26]. However, these libraries also often enjoy various permissions to collect sensitive data [14, 41], and have been shown to track contact information, browser history, and call logs, for purposes of targeted analytics and advertising, even if this is not their intended functionality [27]. Children's apps are not exempt from this form of data tracking. In fact apps in the 'Family' category on Google Play are shown to have the second highest number of data trackers [12]. Paid apps which do not make use of advertising do not significantly reduce the amount of data tracking and collection in their apps [10, 29, 59]. Children surrendering their data is often non-negotiable, as they are nudged into disclosing more data than is required [7, 61] or are forced to disclose their data to access a service [37, 47].

Persistent data tracking and the resulting loss of privacy can lead to concrete harms, such as the sale of sophisticated user profiles by data brokers and exposure to sexual predators [22, 48]. Other consequences include the normalization of a surveillance culture [45], identity theft, and fraud [21]. Even more significant are the risks associated with long-term life opportunities and reputation as children grow older [44]. As technology advances, the potential for misusing data accumulated throughout a person's life, from childhood to adulthood, will only increase.

2.3 Developer Incentives and Supports for Them

The state of the mobile ecosystem and the current app landscape is often attributed to app developers. While limited research has been conducted on their role in shaping it, a growing body of work is helping us understand their design motivations and challenges in creating privacy-friendly apps [6, 8, 24, 46]. It is now known that third-party libraries are widespread because they expedite and simplify development, and are more secure than proprietary software [26]. Importantly, developers often rely on targeted advertising for most of their revenue [4, 19, 38], delivered through libraries developed by advertising networks. More recent work reveals that while developers are motivated to uphold privacy protection for children, they often lack support, guidelines [23, 24], or understanding of libraries' data collection practices [8, 24]. Developers frequently use data collection and targeted analytics revenue models because they are readily available and supported by marketplaces [46]. The competitive app landscape also contributes to this, as free apps with advertising are perceived as more popular [24].

Existing tools for developers are scarce, mostly assisting in automating privacy policy creation and communicating privacy-related information to end-users [50, 71]. Studies show that existing guidelines and educational materials for privacy development practices are often high-level and impractical [2]. Developers lack tools for privacy risk analysis similar to those for security risk analysis [39]. Research on IDE plugins for clarifying privacy implications for developers exists [39, 40], but these studies do not specifically address developers' need for increased transparency and understanding of third-party SDKs' data-related behaviours.

3 DESIGNING DATAAVALANCHE.IO

To overcome some of the shortcomings of current supportive tools for developers we identified in the literature, we design DataAvalanche.io, a web-based platform that provides the much needed data privacy-related transparency for common third-party libraries and SDKs. The design of our tool is informed by a twofold set of considerations:

- Firstly, we were guided by recent research that identifies the critical barriers faced by developers to create privacy-friendly apps and navigate the technical privacy implications of third-party libraries.
- Secondly, we also conducted an in-depth analysis of 137 'expert-approved' children's apps from the Google Play Store, to identify the technical and legal privacy implications associated with third-parties libraries used by these expert-approved apps. The empirical data resulting from our analysis informed the design of our tool.

In this section we describe these considerations in detail and present the three core components of DataAvalanche.io.

3.1 Needs from Children's App Developers

To effectively address the needs of app developers, we have formulated three high level design goals for our tool (D1 - D3) drawing on our previous study [24] and literature (see section 2.3).

D1 Support developers making informed choices of third-party libraries by providing more comprehensive privacy information about SDKs and libraries, and a knowledge graph about the data tracker network.

One of the key challenges faced by developers when building privacy-friendly apps for children is navigating the privacy implications of third-party libraries and SDKs [24]. Developers may choose prominent ad vendors without fully understanding the privacy implications of these choices [5, 46, 66]. We therefore aim to provide sufficient privacy-related information about commonly used SDKs, so that it is easier for developers to access relevant privacy information (such as third-party and fourth-party trackers associated with a given SDK), and make better informed decisions.

D2 Support developers making privacy-friendly development choices by proposing privacy-friendly alternatives and configurations.

Research has shown that developers often have difficulties with configurations related to consent in children's apps [33] or identify more privacy-friendly alternative libraries or SDKs [24]. Therefore, we intend to provide concrete instructions for child-specific SDK configurations and a list of alternative SDKs with clearer documentation about their data collection behaviours, enabling developers to make more informed choices.

D3 Support developers to protect the privacy of their users by providing them with legal considerations for using third-party libraries.

Past research has also indicated that developers largely struggle with translating legal requirements into technical requirements [11, 65], which Brexit made even more complex. We aim to make privacy policies more accessible and GDPR implications more explicit. Although we don't provide comprehensive guidelines for developing children's apps, we plan to help developers visualise the countries SDKs share data with, supporting compliance with relevant laws and regulations.

3.2 Privacy Analysis of Children's Apps

In 2022, Google Play Store launched a new category called expert-approved' apps for children under 12 [54]. These apps, designed for kids and approved by teachers and children's media specialists, are deemed 'age appropriate' and 'thoughtfully designed'. We specifically chose these apps for our analysis, as Google promotes them as safe for children, while not disclosing the assessment methodology or expert selection process.

https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/overview-data-protection-and-the-eu

3.2.1 Selecting 'Expert-Approved' Apps. In March 2022, we scraped the Google Play Store children's category to create an initial dataset of 1038 expert-approved apps. The apps we scraped were directly accessible from the homepage, making it likely for users to encounter them. Subsequently, using the number of reviews as an indicator, we sorted the apps according to their popularity. We ensured that we (i) selected a diverse range of apps by selecting at least one app from each genre and (ii) avoided apps from duplicate developers and studios (as they often tend to have similar design features and one privacy policy for all their apps). We also limited ourselves to apps which were free to download (due to funding limitations) and which clearly had an age rating. We were left with 470 apps, of which we selected the top 150 most downloaded apps for manual analysis. We faced issues with 13 apps, such as non-English content, premium access, or installation problems. This resulted in a final total of 137 apps for analysis.

3.2.2 Tracker Analysis. For our analysis we used a Huawei Mediapad M5, with 4GB RAM, 32GB storage, running Android 9.0. We initialised the device using a child account, creating a female persona of 10 years old. We needed to verify this account with a second parent account using a credit card. Using a child account we were faced with several limitations. For example, we did not have access to the developer settings and needed approval from the parent account before installing certain apps. We made use of a child account to simulate what a child would experience interacting with these apps.

To understand the technical privacy behaviours of the app, we drew on existing track analysis methods. We had two main goals in our analysis:

(1) Identify the tracker network of the third-party SDKs commonly used by this group of 'expert-approved' apps for children.
(2) Understand whether there is a gap between the data tracking that is happening in apps and the latest data privacy legal requirements for technologies designed for children.

We used a third-party tool called TC Slim, which dynamically identifies third-party trackers [33]. Using the TC Slim app, we identified all associated trackers and libraries for the 137 expert-approved apps. For each third-party tracker and library, we reviewed the API documentation to determine whether they require special configurations for children's apps. In such cases, the documentation often explicitly states that apps aimed at children or part of Google's 'Designed for Families' program need to add specific lines of code, such as the Adjust library, to ensure COPPA compliance.

Besides our privacy analysis, we explored the implications of third-party tracking in apps by mapping the fourth-party tracking ecosystem. We define fourth-party entities as data-sharing partners of third-party entities. For instance, if a third-party entity like Unity shares data with its partners, these partners are considered fourth-party trackers. To achieve this, we first identified the most common trackers in our children's app dataset, which were Google and Unity. We examined their privacy policies and legal documentation available on their websites to locate data partners, understand data sharing practices, and identify the entities data is shared with. In some cases, this information was not readily available, as a list of data-sharing partners might not be present or was incompletely presented visually rather than textually. We often encountered the same partners from multiple sources, resulting in a dense network of data-sharing partners. We then manually cleaned up the data to reconcile naming differences among different companies.

Finally, we cross-referenced companies with existing public knowledge bases, such as Crunchbase and Rocket Companies, to enrich our dataset with additional metadata like company location, number of employees, target market,

https://play.google.com/store/apps/details?id=net.kollnig.missioncontrol.play
https://help.adjust.com/en/article/apps-for-children-android-sdk
https://www.crunchbase.com
https://www.rocketcompanies.com

etc. Using this method, we created a knowledge graph with over 3000 nodes, representing data-sharing relations between commonly used app SDKs and third- and fourth-party libraries.

3.3 Results of Privacy Analysis of 'Expert-Approved' Children's Apps

3.3.1 Third-Party Tracking. Our third-party trackers found that 30 out of all apps did not contain any trackers. We identified a total of 43 unique tracker libraries used in the 137 apps we analysed. Figure 1a shows the distribution of the most commonly occurring libraries which engage in tracking practices, including *advertising*, *analytics*, *fingerprinting*, and *social* functions (such as social media).

Google SDKs, such as Google Analytics, Crashlytics, and Google Ads are the most commonly occurring tracking libraries (33.3%). Surprisingly, several apps also had Facebook integrations (2.8%), while Facebook is specifically aimed at children above 13. Apart from Google SDKs, apps also made use of other attribution and analytics providers, such as AppsFlyer (3.3%), SuperAwesome (2.8%), Mixpanel (1.1%), and Amazon (8.3%).

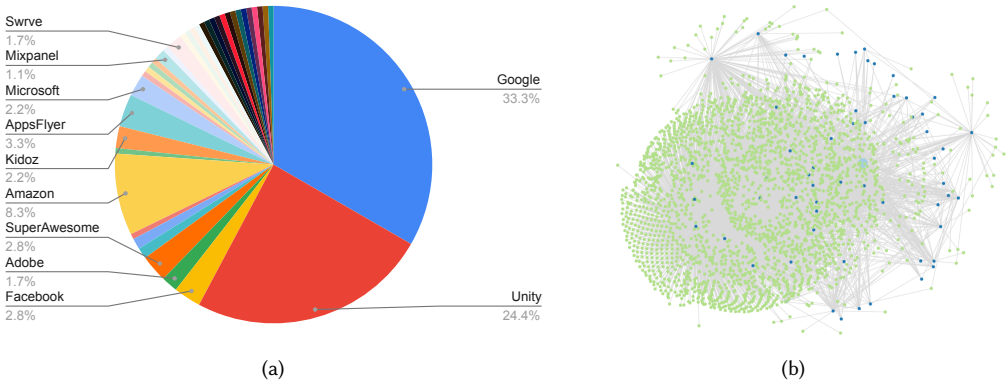

Fig. 1. (a) Distribution of most commonly occurring trackers, based on our analysis of 137 children's apps. (b) The tracker knowledge graph showing the data sharing partners or Unity. The blue nodes are the direct partners of Unity, the green nodes are their data sharing partners in turn.

3.3.2 API Configurations. Looking more specifically at the privacy policies, we found that 15 of these libraries have specific configurations for when they are to be used for children's apps. For example, Adjust and Appsflyer state in their documentation that their SDKs need to be specially configured so that the advertising ID is not transmitted. Similarly, some of the larger companies have such requirements as well, such as Google and Unity. However, we did not verify whether apps complied with this.

3.3.3 Fourth-Party Tracking. To map fourth-party trackers, we manually read through legal documents of these tracker companies. Although Google was the most prominent tracker in our analysis, with a list of 1067 sharing partners, it was impractical to manually review all these policies. Instead, we chose Unity as our root node, which had 71 data sharing partners. Using the 71 fourth-party companies from Unity as a starting point, we constructed a dense network of 2973 nodes and 14872 edges. An image of the data sharing partners of Unity alone is shown in Figure 1b. From the

https://help.adjust.com/en/article/apps-for-children-android-sdk
https://developers.google.com/admob/ios/api/reference/Classes/GADRequestConfiguration?hl=en#-tagforchilddirectedtreatment:
https://docs.unity3d.com/Packages/com.unity.ads\spacefactor\@m{}3.3/manual/MonetizationResourcesDashboardGuide.html?q=children

whole network, the companies with the most number of data sharing partners are Xandr (1942 partners), Google (1067 partners), and Centro (163 partners).

We also mapped the countries of incorporation of all companies. We found that in total data is shared with 40 unique countries. For 66% of the companies we could not determine the country information, as we could not find it in a public registry. Of the remaining companies, 17.3% are from the US, 4.3% from the EU, 0.6% from China, and 0.2% from Russia.

Our graph shows that by sharing data with a single company, such as Unity, it is theoretically possible for this data to reach thousands of companies in total. To describe this phenomenon of uncontrollable large scale data dissemination, we use the term *data avalanche*. The rest of the network can be viewed and analysed using our tool, by visiting DataAvalanche.io.

4 AN OVERVIEW OF DATAAVALANCHE.IO

Motivated by the developers' needs and the technical and legal privacy implications identified from our app analysis, we implemented the web-based tool DataAvanche.io, which consists of three key components.

4.1 D1: Empowering Developers to Make Informed Decisions about SDKs and Libraries

D1's objective is to equip developers with tools and guidance for informed privacy decisions regarding SDKs and libraries. To achieve this, We offer a **search feature** for easy access to specific SDKs/libraries and their alternatives, and provide **meta-data** (e.g., website, company, location) to understand the company's nature. Additionally, **privacy information** (e.g., privacy policy, data-sharing partners) helps developers assess the SDK's privacy practices. Additionally, it provides **privacy information**, including a link to the privacy policy and the number of data sharing partners, to help developers understand the privacy practices of the SDK. Developers can also understand data-sharing practices and their implications through a **data-sharing partners table** (see Figure 2b) and a **data-sharing partners graph**, which visualises relationships between data-sharing partners, allowing informed privacy risk decisions.

4.2 D2: Providing Developers with Tools and Guidance for Protecting User Privacy

D2 aims to provide developers with the necessary tools and guidance to protect the privacy of their users by offering features such as **child-specific configurations** and **privacy-friendly alternatives**. The child-specific configurations feature offers links to necessary configurations required by COPPA and GDPR for child users. We specifically added this because it has been shown in previous research that apps for children do not sufficiently obtain consent for data tracking, partly due to app misconfigurations [33]. Therefore, we manually examined 71 SDKs to determine special configurations for children's apps and included this information in the metadata section (see Figure 2a). As an example, the documentation for the Adjust SDK states that:

> "If your app targets users under the age of 13, and the install region is not the USA, you need to mark it as a Kids App. This prevents the SDK from reading device and advertising IDs (e.g. `gps_adid` and `android_id`). To do this, call the `setPlayStoreKidsAppEnabled` method with a `true` parameter".

For those that have special configurations, we included this information in the metadata section of our tool, along with a URL to the relevant documentation (see Figure 2a).

We also aimed to provide a list of **privacy-friendly alternatives** for popular libraries. There currently does not exist a reliable database that offers such a mapping of alternatives, and creating one is beyond the scope of this paper.

https://help.adjust.com/en/article/apps-for-children-android-sdk

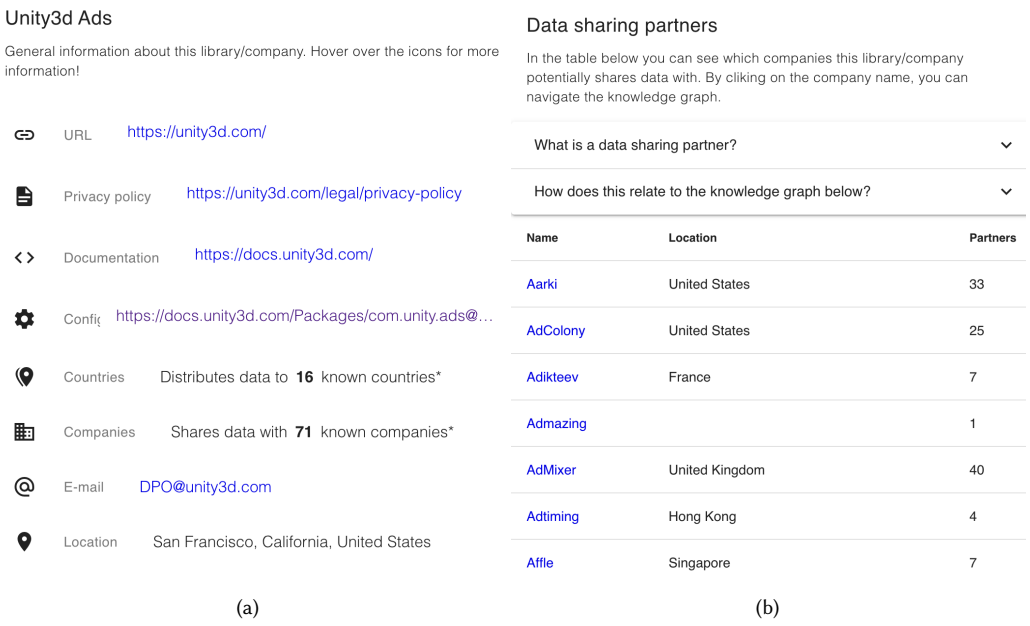

Fig. 2. (a) Privacy information provided for an example SDK *Adjust* by DataAvalanche.io. (b) Table allowing users to view the data sharing partners of the selected third- or fourth-party organisation/library.

Instead, as the goal of our research is to explore whether providing such privacy alternatives may affect app developers' development practices, we selected three libraries and manually identified alternatives, which developers can access as part of the tool.

4.3 D3: Shedding Light on the Legal Implications of Data Sharing Infrastructure

We implemented D3 through two components: **information on legal implications**, and a **visualisation of data sharing**. These components help developers understand the legal risks and requirements of using certain libraries based on data-sharing locations. The data sharing visualisation allows developers to easily see where their app's data is being shared and to make informed decisions about the privacy risks and benefits of different options (see Figure 3).

We wanted to provide transparency on the legal implications of using third-party libraries and SDKs; however, providing a comprehensive legal analysis of every SDK is beyond the scope of this study. Given our focus on the UK market, we instead focused on clarifying a specific GDPR-related privacy implication and providing easier access to privacy policies of all the SDKs wherever possible. For the former, we read through the privacy policies of the SDKS and provided clarifications of GDPR requirements for those SDKs that claim to share data outside the EU jurisdiction. This information is aimed to help developers understand that such SDKs may share data with countries where end users may not be able to enforce their privacy rights and that they should be mindful of this when making decisions about which libraries and SDKs to use.

5 EVALUATION

We evaluate our findings and tools with 12 children's app developers, serving a dual purpose:

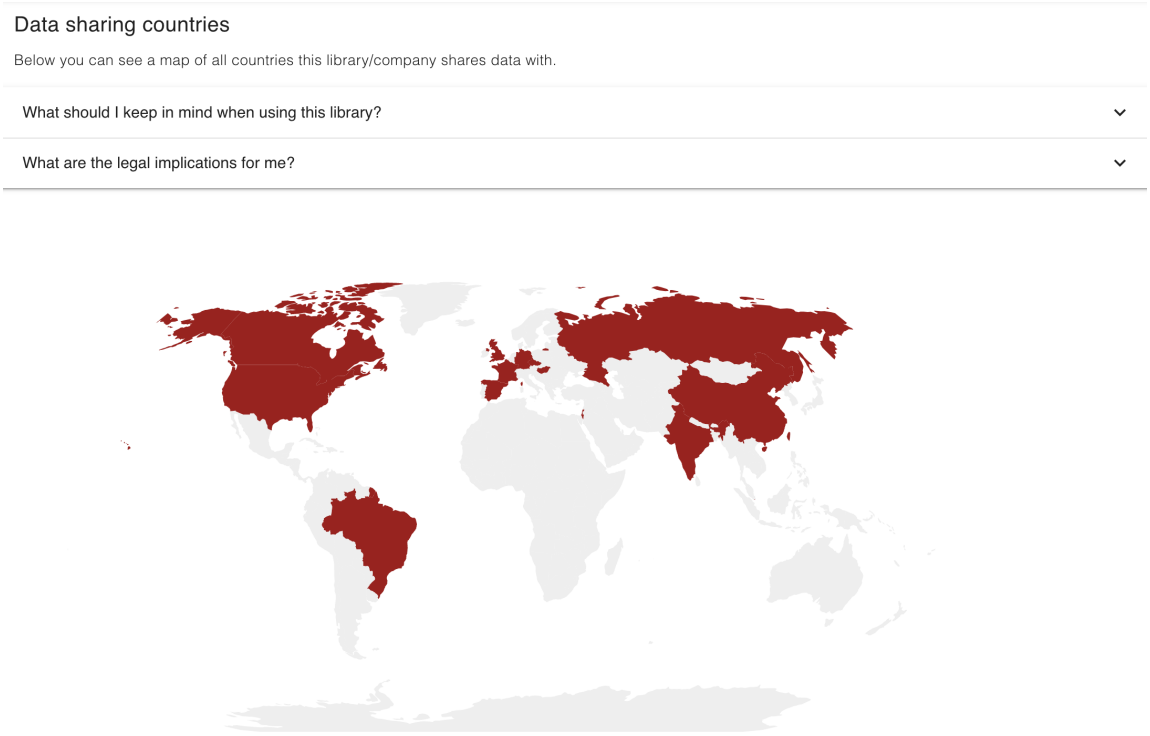

Fig. 3. In DataAvalanche.io users can view to which countries SDKs are sharing their data with. We have provided static information about what GDPR implication would be.

(1) Investigating developers' third-party library choices and privacy considerations allows us to understand factors influencing decisions. Our tool helps us understand developers' privacy-related decision-making practices.
(2) Evaluating our tool's usefulness in helping developers understand common SDKs' privacy practices enables us to assess its impact on decision-making and its role in assisting informed third-party library selection and user information. This relates to our research question about the tool's effectiveness in guiding developers regarding privacy practices of various SDKs and libraries.

To evaluate our tool, we employed semi-structured interviews and exit surveys, using the Systems Usability Scale [18], as detailed in the sections below.

5.1 Participants and Recruitment

We sought developers experienced in creating children's mobile apps. Using search engines like Google and DuckDuckGo, we employed keywords such as "Kids App Developers" and "Game Studio" to find participants. Preliminary discussions involved email exchanges or video calls, during which consent procedures and data rights were explained. We recruited 12 children's app developers who each received a £15 Amazon gift voucher. Our study protocol was approved by an Institutional Review Board.

Table 1. Table containing descriptive characteristics of the study participants. While the participants were not located in the UK, they all developed apps for the UK market.

Participant	Country	Role	Years of Exp.
P1	India	Lead Android developer	10+
P2	Brazil	Mobile game developer	5-10
P3	Armenia	App developer	2-5
P4	US	App developer	5-10
P5	Armenia	Web and Mobile developer	2-5
P6	Canada	App developer	0-2
P7	US	App developer	10+
P8	Israel	Mobile game developer	2-5
P9	Germany	App developer	5-10
P10	US	App developer	2-5
P11	Canada	App developer	5-10
P12	Bulgaria	App developer	2-5

Children's app developers were challenging to recruit, as the market for developers is relatively small compared to the market for children's apps. We stopped recruitment at 12 participants, as research suggests this number is sufficient for data saturation and stable themes to emerge [28]. Though the participants were not UK-based, they had experience developing apps for the UK and EU markets. All participants were male and had direct app development experience. Participant demographics are shown in Table 1.

5.2 Interview Procedure

We conducted 12 semi-structured interviews in the summer of 2022 using remote meeting software, such as Microsoft Teams, Google Meet, or Zoom, depending on the participants' preferences. All interviews were audio recorded with the participants' consent. We started each interview with a general introduction, during which we introduced ourselves, the research goals and objectives, and once again informed participants about the procedure. The interviews consisted of three parts.

(1) We discussed developers' app design experiences and privacy considerations, and inquired about empowering and supporting children in making privacy decisions.
(2) We shared DataAvalanche.io with participants, requesting they open the URL and share their screens. This allowed us to use the tool as a probe to uncover their thoughts on commonly used SDKs' privacy behaviours. To ensure consistent usage, we employed a cognitive walkthrough process [30, 58], commonly used for assessing system usability. During the walkthrough, participants opened DataAvalanche.io, shared screens, and briefly explored the tool while discussing initial thoughts. They chose from two pre-selected SDKs, Adjust and Unity Ads, to explore in-depth. Participants shared their thoughts and considerations as they examined metadata, privacy-friendly alternatives, legal and data sharing sections, and potential incorporation into development practices.
(3) In the third part, after these tasks were completed we asked them about the usability of the tool itself. We asked which features they liked and disliked, which features they would like to see added, and how they would see themselves using this in practice. At the end of the interview, we asked participants if they had any additional

comments or questions. After the interview, we sent the participants the exit survey. The full set of tasks of the cognitive walkthrough can be viewed in Appendix B

5.3 Surveys

For the exit survey, we made use of the Systems Usability Scale (SUS) [18] to gauge the perceived usability of the tool from the perspective of developers. The SUS is a widely recognised and standardised evaluation tool that measures the usability of a product or system. It consists of 10 statements that users are asked to rate on a 5-point Likert scale, with higher scores indicating a higher level of perceived usability.

We also asked two open-ended questions at the end of the survey to solicit additional feedback from participants. The first question asked if there were any additional features that participants would like to see in the tool, and the second question asked for any additional comments. These open-ended questions provided us with valuable insight into participants' thoughts and ideas, and helped us to identify areas for improvement in the tool.

5.4 Data Analysis

The collected audio recordings of the interviews were transcribed and anonymised. The anonymised interview transcripts and answers to the open-ended survey questions were analysed using a grounded, thematic approach [31] to develop codes and themes related to developers' *challenges, perceptions, requirements* and *values*. We iteratively coded the data, identified emerging themes, and refined and organized the themes into a cohesive pattern or structure. We examined the perspectives of different research participants, highlighted similarities and differences, and generated unanticipated insights.

The thematic coding process started by dividing the transcriptions into two equal-sized sets. The first and second authors independently analysed the first set of transcriptions to derive an initial set of codes. They then met to consolidate and reconcile codes into a common codebook. The first author then completed coding the remaining half of the transcriptions using this common codebook. Interviews lasted between 38 and 75 minutes, with the average interview being 46 minutes (SD=11).

6 EVALUATION RESULTS

Here we report the results of the evaluation of DataAvalanche.io. We first start by describing the results from our interviews. Through our analysis we uncovered how the participants perceived the privacy transparency provided by DataAvalanche.io, some open barriers, and some further requirements for the tool.

6.1 DataAvalanche.io can Increase Transparency of SDKs

Overall participants indicated that DataAvalanche.io can be positively useful for their own development. They found it particularly useful that all information is accessible through a central repository: "*It is good to have a place where you can find easily all these third-party tools and their documentation easily to access,*" (P2). They also liked that certain privacy information was presented visually, as "*we understand visuals and it helps to take it home for the developer,*" (P4). As a result, they found that this makes it easier to make decisions in choosing libraries. One developer gave an example of this: "*As you can see, there is so much difference between Amplify and Google Firebase. If we we're creating a bigger application it's better to use something like this [pointing to an SDK] so that they don't share too many data,*" (P5).

Participants found the *legal implications* provided by our tool particularly useful to address the *transparency* challenge they often face when choosing third-party libraries. Participants reflected that while they have suspcicions about what

data is being collected and for what purposes, they could not exactly put their finger on it:"*For example, when it comes to Firebase the data that we were storing, is it being used by Google or is it not? I'm not sure about it [...] I've had some concerns about the data in Firebase, but I'm not really sure,*" (P2).

They ascribe part of this problem to the complexity of privacy policies and terms and conditions, stating that they are not accessible and written in a way that the general public cannot understand it: "*Part of the problem is if you read through the disclosure that's written by lawyers, it has to be written very conservative from a legal standpoint, so it's not transparent. Nobody else but another lawyer could read it and make sense of it,*" (P7).

Interestingly, many participants also mentioned that the *legal implications* discussions provided by DataAvalanche.io would also help them amend their privacy policies or appropriately inform end users: "*freelancers would be able to share it with clients if there are concerns about data privacy. If the clients live in a country where there are strict regulations when it comes to data sharing. This could be very useful, "* (P3). Furthermore, oftentimes developers need to list all third-party services integrated into their apps in the privacy policy, alongside notable privacy practices, which DataAvalanche.io makes substantially easier to do.

6.2 DataAvalanche.io can Help Both Developers and Users

During the interviews, we asked participants about their potential use cases for DataAvalanche.io and the advantages it could offer in their development practices. While we have specific use cases in mind for the tool, we acknowledged that participants may come up with creative and more relevant use cases based on their experiences.

Firstly, participants identified several ways in which our tool could be used to benefit children and end users of apps. For example, they suggested that it could be used to generate privacy policies and present privacy information to end users in an improved way, so that they could avoid presenting users a lengthy document but a more succinct one with structured information about what data and how they are shared. They also mentioned the possibility of using it for requesting consent in apps: "*If we create an application, we can let the users know that, let's say, Firebase shares their data around the world. This way, we can [obtain] consent from the users,*" (P5).

Secondary, participants indicated that privacy is becoming increasingly important in today's industry. Many investors and startups prioritise privacy, and developers often need to demonstrate strong privacy practices to obtain funding. One of the examples they suggested was tech-for-good companies, "*developers who have the best interest of the users, [as] they would use this application,*" (P6). The increased demand for privacy also means that privacy-friendly SDKs which are now offered as alternatives will become more and more mainstream, and there is a need for a platform for them to communicate their availability: "*Privacy is a big thing now. It's becoming bigger and bigger nowadays. People are spooked about it. I think this is very important to have. It'll help you give a better description in your own privacy policies when you do write them,*" (P4).

6.3 Privacy-friendly Alternatives Are Good But Not Always Applicable

In evaluating the usefulness of our tool for developers to consider privacy-friendly alternatives, such as open-source SDKs or those known to do less data tracking, we found that participants largely welcomed the option: "*I think the privacy-friendly alternative section is probably a really key feature because it suggests to me that if I'm using a library to do notifications, I can probably find something else and preferably, ones that are open source and that I can use in my application instead of this, "* (P4)

However, this sentiment was not shared by all participants, and in most cases, privacy was not the leading factor when choosing SDKs. In fact, "*privacy concerns really just take a backseat, it is not the first thing that we think about,*" (P11). Developers formulated four reasons why they would choose a prominent SDK over a privacy-friendly alternative.

6.3.1 Financial Incentives. Firstly, participants acknowledged that apps need to earn money, and that advertising is often the most lucrative way to do this. However, they also acknowledged that some advertising networks generate more profits than others, and therefore, there may not always be a strong incentive to choose a privacy-friendly network over a more lucrative one: "*The problem was, to show an interesting ad to a user you're supposed to know them really well. Only two of the companies that could ever - I could think of right now - to know anyone in the planet more than anyone else is Facebook and Google. Others don't have that many privileges and rights, "* (P1).

6.3.2 Learning Curve and Time Investment. Secondly, participants indicated that they have already invested a significant amount of time in learning and becoming proficient with a particular SDK and its related features. Changing to a new SDK would negatively affect their productivity, workflow, and their ability to deliver the service they are offering. The challenge of learning a completely new tool, even if it is the best on the market, can be a significant obstacle: *"Whenever you use a new tool, especially analytics-wise, you have a big learning curve. You could create one of the best privacy-friendly analytics tools, but the problem is my eight years learning curve could be on Google Analytics,"* (P3)

6.3.3 Insufficient Feature Offering and Quality. Thirdly, some participants prioritised the features and quality of certain SDKs, indicating that this is one of the more important reasons for choosing an SDK, as the "*functionality of the actual third-party library is going to be the dominant factor in making a choice,*" (P7). Others were a bit more direct and stated that "*in terms of choosing a platform, I'd say the faster and the better well-known, that's the priority,*" (P2).

Additionally, participants indicated that some SDKs come built in with integration with many other tools, as is the case of Firebase, which hooks into many other Google products: "*Firebase has pretty much everything, so if you want a Google sign-up or a Facebook sign-up, or any of the third-party application sign-ups, you can do it through Google,*" (P6)

6.3.4 Education Cost. Lastly, some participants indicated that certain SDKs are more accessible due to their popularity and the availability of educational resources online. This makes them more attractive to beginner developers or developers with a limited budget: "*Why I'm using Firebase? Because it's mainly popular. Every course that I watch, such as in Udemy, everyone uses Firebase. Afterwards, every time that I need stuff like that, I use Firebase,*" (P11). We finally asked participants what would prompt them to consider alternatives. They took a practical stance and stated that ultimately, the product offering is set by customer demand. If customers demand more privacy-centric products, then that is what they will provide: "*If there is a good enough demand, developers will have no other option to offer it,*" (P1).

6.4 Challenging to Safeguard Own Legal Interests

Although participants welcomed the legal implications which our tool provided, they expressed their frustration that third-party vendors often try to have everything 'technically' disclosed, by having elaborate and complex privacy policies, and thereby have themselves legally protected. They referred to such practices as being invisibly transparent: "*companies do take information and they are pretty transparent, but I say they are invisibly transparent "* (P6).

This frustration is also reflected through participants' passive reaction to the comprehensive data tracking and sharing information we provided in the tool. While we expected our knowledge graph and data sharing maps to induce a shock factor, due to the pervasiveness and vastness of data sharing, we found that in general most participants were not very surprised. Most had a strong belief that third-party SDKs collect and share much more data than is necessary:

"*They probably know from each computer's IP what each person is browsing, and when you connect it to Google using Google Analytics. They would have access to whatever niche that your website is. They would basically gather all this data from the user using the IP, and they probably would connect it,*" (P3).

They largely accepted that data tracking is an inherent part of life nowadays, even if users take measures to protect themselves: "*people might have pushed the Do Not Track button, but at the end of the day Facebook is still going to look at other parameters for your device and associate that,*" (P11).

However, they felt limited in adjusting data sharing attributes of current SDKs, especially when third-parties are endeavoured to protect their own legal responsibility by being 'invisibly transparent': " *I would love to see some kind of filter where I could say, okay, I would love only to share within Europe to have this to get me alternatives for example,* " (P9).

6.5 Additional Requirements

Participants formulated several additional key requirements which they would like to see implemented in our tool. These include the need for navigation and filtering, to help them find the SDKs that meet their specific needs, the importance of having access to clear documentation and educational resources, the need to consider privacy implications, and the commercial considerations such as revenue and market trends. Here we summarise these requirements, while a full overview of all 17 requirements can be seen in Table 2 in Appendix C.

In terms of **navigation and filtering**, participants mentioned the importance of being able to categorise and sort SDKs based on various criteria, such as popularity, features, or compliance standards. They also highlighted the need to be able to filter SDKs by countries and the ability to mark favourites and bookmarks for future reference.

Documentation and education were also identified as key considerations. Participants emphasised the importance of having access to clear and concise documentation for SDKs, as well as educational resources such as tutorials and summaries of privacy policies. They also suggested the idea of embedding documentation directly into the tool and the possibility of community contributions to the tool.

Privacy was another important factor for participants, who highlighted the need to see more details about data sharing practices and who is benefiting from the data. They also mentioned the need to be alerted about privacy breaches and to be able to compare apps in terms of their privacy practices.

Finally, participants pointed out the **commercial considerations** they often have when selecting SDKs, such as the need to consider the revenue potential of their apps and the impact that certain SDKs might have on their market trends. They suggested the idea of being able to see trends and having the option of adding user data to a map.

6.6 Survey Results

Our survey results showed that participants generally had positive reactions to the usability of DataAvalanche.io (see Figure 4). For example, participants largely agree that *the system was easy to use, the functions were well integrated, and people would learn to use this system quickly*. However, there were some statements that received mixed responses, including the frequency of use and the ease of learning. While some participants agreed that they would use the tool frequently and that it was easy to learn, others demonstrated different opinions on *the consistency of the system* or *whether the system was cumbersome to use*. These mixed responses suggest that while the tool is generally well-designed, there may be areas for improvement to enhance the user experience further. The open-ended questions of the survey provided some answers as to what these improvements may be, but we generally found that they aligned with the additional requirements formulated in the interviews. For example, participants wanted to features to compare libraries and more intuitive visualisations of the data.

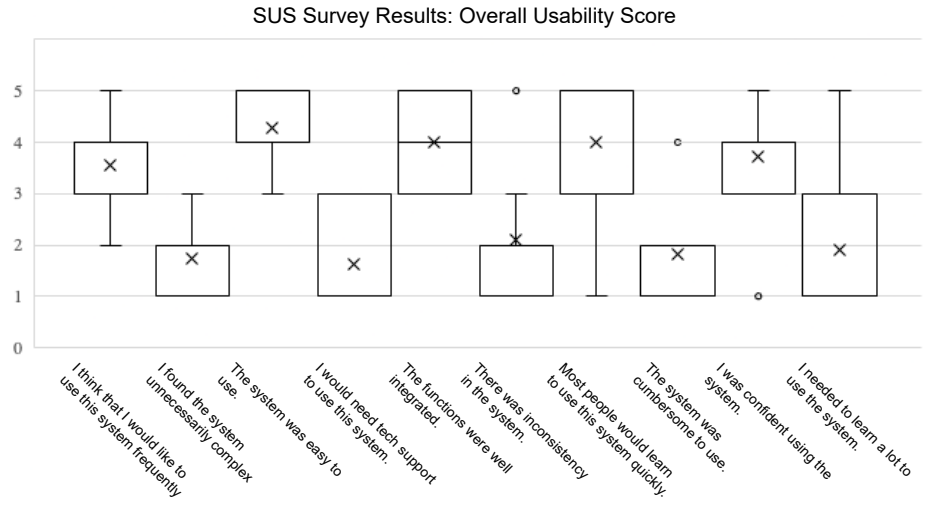

Fig. 4. Box plot of SUS survey results illustrating the usability scores for the DataAvalanche.io tool. The plot displays the distribution of scores among participants, highlighting the median, quartiles, and potential outliers.

7 DISCUSSION

In this study we critically examined the intersection of privacy concerns in children's apps, and the perspectives and support required by developers in order to effectively address this growing demand. Using developers' needs identified in previous literature and privacy information about common third-party SDKs used by 137 expert-approved children's apps, we designed and implemented a web-based tool, DataAvalanche.io, which provided transparency on the privacy practices of commonly used SDKs for children's apps for app developers to navigate and make more informed design decisions. We found that our participants perceived positively of the transparency provided by the tool, particularly the legal transparency, and they believe the tool can be useful for both development and users. However, we also identified several open barriers and challenges for such support to be practically useful for the app developers.

7.1 The Complexities of the Data Tracking Ecosystem

Our tracker analysis revealed that there is a vast network of data sharing entities outside of app stores and marketplaces that consumers and developers are not aware of, which has several consequences. Firstly, once data is transmitted to a fourth-party, it is recorded permanently and near impossible to retrieve or delete. It has been shown that even sending out a handful of GDPR Data Subject Access Requests and negotiating with companies can be an arduous task [35]. GDPR rights of consumers appear ineffective in this context, as it is a challenging task to send thousands of data subject access requests to retrieve or delete data, particularly when companies are located outside the EU jurisdiction. Secondly, there is a lack of transparency for both consumers and developers about how and where data is being shared, making it difficult for them to understand the scale and permanence of data sharing. This lack of transparency contributes to a lack of awareness of the risks associated with data sharing and can lead to the use of harmful third-party SDKs. Thirdly, despite the prevalence and ubiquity of third-party tracking in children's apps, they are still marketed as appropriate for children. For example, despite privacy concerns, Google still labels them as 'expert-approved' and promotes them as

safe and appropriate for children. However, the criteria used to approve these apps is not based on technical features or privacy considerations, but rather on content-based approval by teachers and media experts [54]. Google has made some changes in this regard, such as the introduction of the Data Safety section in apps, but it does not provide information about which parties the data is being shared with.

From a developer's perspective, we sought to raise awareness of these issues by integrating a visual depiction of the phenomenon into DataAvalanche.io. However, developers revealed that they were uncertain as to how to interpret this, and how they could apply it in practice. A potential solution to this would be to provide concrete examples of how the knowledge graph can be used in practice, such as case studies or scenarios that illustrate the advantages of using privacy-friendly SDKs and avoiding certain third-party libraries.

7.2 Challenges in Designing Privacy-Friendly Apps for Children

Through our analysis of privacy features in apps, we have gained a deeper understanding of how children need to be better supported in understanding and exercising their privacy rights. We have broadly identified three ways in which privacy is inappropriately negotiated with children.

Firstly, in many cases, privacy was treated as a legal necessity in apps, rather than as an effort to benefit children. Children were presented with or referred to long and complex privacy policies and had no other option than to give consent. This approach fails to consider the cognitive and developmental limitations of children and it is crucial that designers of children's apps consider these limitations when creating privacy-friendly apps. Secondly, we have observed cases in which privacy is entirely disregarded or not mentioned, resulting in the app potentially collecting and using personal information without the user's knowledge or consent. This can lead to a general disregard for privacy and the erosion of trust between individuals and developers. Developers must prioritise making privacy transparent by clearly providing information about data collection and usage. Lastly, privacy is often limited to parents, for example through age-assurance mechanisms being applied to privacy settings and the disclosure of privacy information. The AADC, however, advocates simplifying the complexities of privacy to a level at which children are able to comprehend and interact with it, including providing age-appropriate explanations and easy-to-use controls.

While we looked at this issue from the perspective of developers in this study, we should also consider that marketplaces should set strict requirements for user-facing privacy features in apps, as they have a role to play in guiding developers. Guidelines from marketplaces look mostly at technical elements, such as third-party tracking [24], but would benefit from informing developers about which user-facing privacy features should be included in apps and how they should be communicated to users. This can for example be done by providing wireframes and software development kits (SDKs) or creating designer guidelines. This can be similar to the approach taken by the UK ICO, which provides developers with tools such as data process mapping.

7.3 Improving Support for Developers

Previous research has shown that developers find third-party libraries and SDKs essential to their development practices [8, 26, 64]. While our tool provided privacy-friendly SDK alternatives, our evaluation revealed additional reasons why developers might not choose a privacy-friendly alternative. It is important to offer similar functionality as more well-established SDKs, along with clear and actionable guidance for developers on how to leverage the SDKs to protect user privacy. Our research also highlighted the influence of market demand in driving privacy in apps,

https://support.google.com/googleplay/answer/11416267?hl=en&co=GENIE.Platform%3DAndroid
https://ico.org.uk/for-organisations/childrens-code-design-guidance/create-data-privacy-moments-maps/

suggesting the need for a multi-stakeholder approach to improve privacy. The HCI community plays a pivotal role in engaging with these stakeholders, such as policymakers, user advocates and industry partners, to create a supportive environment that encourages developers to prioritise privacy, potentially through financial incentives, raising awareness of the importance of privacy, and developing standards and guidelines for app development.

Participants formulated additional requirements for our tools, such as being able compare and filter SDKs. We interpret these requirements as a sign of potential for our prototype to become a useful product in terms of privacy. Some requirements, such as improving the design and usability, can be easily implemented, whilst others require further research. For example, participants proposed community contributions through ratings and trust scores. Recently, privacy labels have become more popular, as evidenced by Apple's requirement for developers to disclose all personal information collected in apps, and Google's initiative allowing developers to choose to disclose which data is collected and whether it is shared with third parties. Despite the fact that these tools are designed with end users in mind, rather than developers, an incentive for developers to rate the privacy-friendliness of SDKs as well as their usefulness may encourage the adoption of more privacy-friendly alternatives. We should also bear in mind the decision making process of developers, such as the educational material, learning curve and feature offering.

Furthermore, increased transparency or privacy policies don't always lead to higher privacy awareness or behaviour change [32, 63]. Our findings show developers appreciate our tool's transparency and legal implications, but these perceptions may stem from legal compliance needs rather than privacy commitment. As an interview study, we cannot yet observe how transparency support influences development practice changes.

Finally, we would like to point out that although our tool was motivated and designed to support app developers for children, it can also be useful for the app development community in general. Our results show great potential for a tool like DataAvalance.io. However, maintaining many of the resources in the tool would require a much larger commitment. We hope our work will provide a foundation for future initiatives or community effort to further develop and sustain similar efforts.

7.4 Limitations

We would like to address several limitations of our study. Firstly, due to difficulties in recruiting, we had a limited sample size. While 12 participants generally suffice for data saturation [28], the app development process is complex, requiring different specialisations, SDKs and stakeholders including UX designers, database engineers and coders. We limited our focus to children's app developers, considering SDKs and legal implications specific for children. Possible ways to address these limitations in future studies include recruiting a more diverse sample of developers, with different levels of experience. Secondly, while developing DataAvalanche.io, several of the parts were still in the high-fidelity prototype phase. This means that the results of our study may not fully reflect the potential usability and effectiveness of a fully developed version of the tool. Future work will aim to further refine and test the tool, to better understand its impact on developers' decisions about third-party libraries and SDKs. Lastly, our sample was limited to apps from the Google Play Store and only selected apps from the 'expert-approved' category. Future research should also examine the privacy practices of apps for adults, and compare them to those of children's apps, to understand whether there are any differences in the way that data is collected and privacy rights are exercised. Additionally, future research should investigate whether there is any difference in the financial pressure placed on children between paid and free apps.

https://www.apple.com/privacy/labels/

8 CONCLUSION

In this study, we designed and evaluated DataAvalanche.io, a web-based tool that provides transparency on the privacy practices of commonly used third-party software development kits (SDKs) and libraries. We informed the design of our tool based on previous literature and our own privacy analysis of 137 'expert-approved' children's apps. We evaluated our tool through semi-structured interviews with 12 children's app developers. Our participants reacted positively to our tool but also identified important barriers to making privacy a core element of their app development. They identified additional requirements they would like to see implemented in our tool. Our findings provide important insights for designing supportive tools and policy considerations.

REFERENCES

[1] 2020. Age appropriate design: a code of practice for online services. https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf.

[2] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L Mazurek, and Christian Stransky. 2016. You get where you're looking for: The impact of information sources on code security. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 289–305.

[3] Amelia Acker and Leanne Bowler. 2018. Youth data literacy: teen perspectives on data created with social media and mobile devices. (2018).

[4] Alessandro Acquisti, Curtis Taylor, and Liad Wagman. 2016. The economics of privacy. *Journal of Economic Literature* 54, 2 (2016), 442–92.

[5] Noura Alomar and Serge Egelman. 2022. Developers Say the Darnedest Things: Privacy Compliance Processes Followed by Developers of Child-Directed Apps. *Proceedings on Privacy Enhancing Technologies* 4 (2022), 2022.

[6] Hala Assal and Sonia Chiasson. 2019. 'Think secure from the beginning': A Survey with Software Developers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 289.

[7] Jane Bailey. 2015. A perfect storm: How the online environment, social norms and law shape girls' lives. (2015).

[8] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason I Hong, and Lorrie Cranor. 2014. The privacy and security behaviors of smartphone app developers. (2014).

[9] Claire Balleys and Sami Coll. 2017. Being publicly intimate: Teenagers managing online privacy. *Media, Culture & Society* 39, 6 (2017), 885–901.

[10] Kenneth A Bamberger, Serge Egelman, Catherine Han, Amit Elazari Bar On, and Irwin Reyes. 2020. Can you pay for privacy? consumer expectations and the behavior of free and paid apps. *Berkeley Technology Law Journal* 35 (2020).

[11] Kathrin Bednar, Sarah Spiekermann, and Marc Langheinrich. 2019. Engineering Privacy by Design: Are engineers ready to live up to the challenge? *The Information Society* 35, 3 (2019), 122–142.

[12] Reuben Binns, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. Third party tracking in the mobile ecosystem. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 23–31.

[13] Reuben Binns, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2018. Measuring third-party tracker power across web and mobile. *ACM Transactions on Internet Technology (TOIT)* 18, 4 (2018), 1–22.

[14] Theodore Book, Adam Pridgen, and Dan S Wallach. 2013. Longitudinal analysis of android ad library permissions. *arXiv preprint arXiv:1303.0857* (2013).

[15] Leanne Bowler, Amelia Acker, Wei Jeng, and Yu Chi. 2017. "It lives all around us": Aspects of data literacy in teen's lives. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 27–35.

[16] Great Britain. 2018. *Children and parents: Media use and attitudes report 2018*. Ofcom.

[17] Great Britain. 2022. *Children and parents: Media use and attitudes report 2022*. Ofcom.

[18] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[19] Interactive Advertising Bureau. 2015. IAB Internet Advertising Revenue Report: 2015 Full Year Results.

[20] Stephane Chaudron, Rosanna Di Gioia, and Monica Gemo. 2018. Young children (0-8) and digital technology, a qualitative study across Europe. *JRC Science for Policy Report* (2018).

[21] Sean Coughlan. 2018. 'Sharenting'puts young at risk of online fraud. *BBC News* 21 (2018).

[22] Ratan Dey, Yuan Ding, and Keith W Ross. 2013. Profiling high-school students with facebook: how online privacy laws can actually increase minors' risk. In *Proceedings of the 2013 conference on Internet measurement conference*. 405–416.

[23] Anirudh Ekambaranathan, Jun Zhao, and Max Van Kleek. 2020. Understanding Value and Design Choices Made by Android Family App Developers. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–10.

[24] Anirudh Ekambaranathan, Jun Zhao, and Max Van Kleek. 2021. "Money makes the world go around": Identifying Barriers to Better Privacy in Children's Apps From Developers' Perspectives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[25] Lia Emanuel and Danaë Stanton Fraser. 2014. Exploring physical and digital identity with a teenage cohort. In *Proceedings of the 2014 conference on Interaction design and children*. 67–76.

[26] ENISA. 2018. Privacy and data protection in mobile applications: A study on the app development ecosystem and the technical implementation of GDPR.

[27] Michael C Grace, Wu Zhou, Xuxian Jiang, and Ahmad-Reza Sadeghi. 2012. Unsafe exposure analysis of mobile in-app advertisements. In *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 101–112.

[28] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field methods* 18, 1 (2006), 59–82.

[29] Catherine Han, Irwin Reyes, Amit Elazari Bar On, Joel Reardon, Álvaro Feal, Serge Egelman, and Narseo Vallina-Rodriguez. 2019. Do you get what you pay for? Comparing the privacy behaviors of free vs. paid apps. (2019).

[30] Bonnie E John and Hilary Packer. 1995. Learning and using the cognitive walkthrough method: a case study approach. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 429–436.

[31] Nigel King. 2004. *Essential Guide to Qualitative Methods in Organizational Research*. SAGE Publications Ltd, London. https://doi.org/10.4135/9781446280119

[32] Spyros Kokolakis. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security* 64 (2017), 122–134.

[33] Konrad Kollnig, Pierre Dewitte, Max Van Kleek, Ge Wang, Daniel Omeiza, Helena Webb, and Nigel Shadbolt. 2021. A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps. 181–196. https://www.usenix.org/conference/soups2021/presentation/kollnig

[34] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. 2021. Are iPhones Really Better for Privacy? Comparative Study of iOS and Android Apps. *arXiv preprint arXiv:2109.13722* (2021).

[35] Jacob Leon Kröger, Jens Lindemann, and Dominik Herrmann. 2020. How do app vendors respond to subject access requests?: A longitudinal privacy study on iOS and Android Apps. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3407023.3407057

[36] Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L Clegg, and Jessica Vitak. 2017. 'No Telling Passcodes Out Because They're Private': Understanding Children's Mental Models of Privacy and Security Online. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 64.

[37] Gry Hasselbalch Lapenta and Rikke Frank Jørgensen. 2015. Youth, privacy and online media: Framing the right to privacy in public policy-making. *First Monday* (2015).

[38] Ilias Leontiadis, Christos Efstratiou, Marco Picone, and Cecilia Mascolo. 2012. Don't kill my ads!: balancing privacy in an ad-supported mobile application market. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*. ACM, 2.

[39] Tianshi Li, Yuvraj Agarwal, and Jason I Hong. 2018. Coconut: An IDE plugin for developing privacy-friendly apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–35.

[40] Tianshi Li, Elijah B Neundorfer, Yuvraj Agarwal, and Jason I Hong. 2021. Honeysuckle: Annotation-guided code generation of in-app privacy notices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021).

[41] Jialiu Lin. 2013. *Understanding and capturing people's mobile app privacy preferences*. Technical Report. CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE.

[42] Sonia Livingstone, Alicia Blum-Ross, and Dongmiao Zhang. 2018. What do parents think, and do, about their children's online privacy? (2018).

[43] Sonia Livingstone, Julia Davidson, Joanne Bryce, Saqba Batool, Ciaran Haughton, and Anulekha Nandi. 2017. Children's online activities, risks and safety: a literature review by the UKCCIS evidence group. (2017).

[44] A Longfield. 2018. Who knows what about me.

[45] Deborah Lupton and Ben Williamson. 2017. The datafied child: The dataveillance of children and implications for their rights. *New Media & Society* 19, 5 (2017), 780–794. https://doi.org/10.1177/1461444816686328 arXiv:https://doi.org/10.1177/1461444816686328

[46] Abraham H Mhaidli, Yixin Zou, and Florian Schaub. 2019. " We can't live without them!" app developers' adoption of ad networks and their considerations of consumer risks. In *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*.

[47] Anca Micheti, Jacquelyn Burkell, and Valerie Steeves. 2010. Fixing broken doors: Strategies for drafting privacy policies young people can understand. *Bulletin of Science, Technology & Society* 30, 2 (2010), 130–143.

[48] Maria Murumaa-Mengel. 2015. Drawing the threat: a study on perceptions of the online pervert among Estonian high school students. *Young* 23, 1 (2015), 1–18.

[49] Finn Myrstad and Ingvar Tjøstheim. 2021. Out of Control. How consumers are exploited by the online advertising industry. (2021).

[50] Elijah Neundorfer and Alfredo J Perez. 2022. ClearCommPrivacy: communicating app privacy behavior in Android. In *Proceedings of the 2022 ACM Southeast Conference*. 248–253.

[51] Ofcom. 2018. Children and Parents: Media Use and Attitudes.

[52] Luci Pangrazio and Neil Selwyn. 2018. "It's Not Like It's Life or Death or Whatever": Young People's Understandings of Social Media Data. *Social Media+ Society* 4, 3 (2018), 2056305118787808.

[53] Jochen Peter and Patti M Valkenburg. 2011. Adolescents' online privacy: Toward a developmental perspective. In *Privacy online*. Springer, 221–234.

[54] Google Play. 2022. Expert approved apps. https://play.google.com/intl/en-GB_ALL/console/about/programs/teacherapproved.

[55] Pooja Pradeep and Sujata Sriram. 2016. The virtual world of social networking sites: Adolescent's use and experiences. *Psychology and Developing Societies* 28, 1 (2016), 139–159.

[56] Kate Raynes-Goldie and Matthew Allen. 2014. Gaming Privacy: a Canadian case study of a children's co-created privacy literacy game. *Surveillance & Society* 12, 3 (2014), 414–426.
[57] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. *Proceedings on Privacy Enhancing Technologies* 2018, 3 (2018), 63–83.
[58] John Rieman, Marita Franzke, and David Redmiles. 1995. Usability evaluation with the cognitive walkthrough. In *Conference companion on Human factors in computing systems*. 387–388.
[59] Suranga Seneviratne, Harini Kolamunna, and Aruna Seneviratne. 2015. A measurement study of tracking in paid mobile applications. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. 1–6.
[60] Wonsun Shin, Jisu Huh, and Ronald J Faber. 2012. Tweens' online privacy risks and the role of parental mediation. *Journal of Broadcasting & Electronic Media* 56, 4 (2012), 632–649.
[61] Wonsun Shin and Hyunjin Kang. 2016. Adolescents' privacy concerns and information disclosure online: The role of parents and the Internet. *Computers in Human Behavior* 54 (2016), 114–123.
[62] Cristiana S Silva, Glívia AR Barbosa, Ismael S Silva, Tatiane S Silva, Fernando Mourão, and Flávio Coutinho. 2017. Privacy for children and teenagers on social networks from a usability perspective: a case study on Facebook. In *Proceedings of the 2017 ACM on Web Science Conference*. 63–71.
[63] Daniel J Solove. 2021. The myth of the privacy paradox. *Geo. Wash. L. Rev.* 89 (2021), 1.
[64] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. 2021. Deciding on Personalized Ads: Nudging Developers About User Privacy. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. 573–596.
[65] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. 2021. Privacy Champions in Software Teams: Understanding Their Motivations, Strategies, and Challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
[66] Mohammad Tahaei and Kami Vaniea. 2021. "Developers Are Responsible": What Ad Networks Tell Developers About Privacy. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.
[67] Paul Voigt and Axel Von dem Bussche. 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing* 10, 3152676 (2017), 10–5555.
[68] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. 'Don't make assumptions about me!': Understanding Children's Perception of Datafication Online. (2022).
[69] Anna L. Wisniewski, Graeme T. Lloyd, and Graham J. Slater. 2022. Extant species fail to estimate ancestral geographical ranges at older nodes in primate phylogeny. *Proceedings of the Royal Society B: Biological Sciences* 289, 1975 (2022), 20212535. https://doi.org/10.1098/rspb.2021.2535 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2021.2535
[70] Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 2019. I make up a silly name': Understanding Children's Perception of Privacy Risks Online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 106.
[71] Sebastian Zimmeck, Rafael Goldstein, and David Baraka. 2021. PrivacyFlash Pro: Automating Privacy Policy Generation for Mobile Apps.. In *NDSS*.

A ARCHITECTURE & IMPLEMENTATION

A.1 Knowledge Graph

The core component of the tool is the knowledge graph, which contains the data sharing partners of the SDKs we analysed in the previous section. The knowledge graph is stored in the back-end and contains all the relevant information associated with the SDKs, including company and privacy information that we want to display to the user. The knowledge graph serves as the central repository of data for the tool, and enables developers to explore and understand the data sharing practices of different SDKs and libraries in the mobile app ecosystem. To our knowledge, this is the first time that a fourth-party tracker network has been examined and presented to the public. We use this knowledge graph as a fundamental part of our tool. We decided to not show the full graph, as with nearly 3000 nodes and 15,000 edges it would become confusing to view and too heavy to load in the browser. Instead, we show the data sharing partners which are one and two degrees separated from the root node. In addition to the graph, we also show the data sharing partners in a tabular manner, so that users can scroll through it textually.

A.2 Back-End Server

On the back-end, we run an NGINX reverse proxy to handle incoming requests and redirect them to the appropriate server. This is important for improving the security and performance of the tool, as NGINX can handle a high volume of requests and traffic, and also provide additional layers of protection against common attacks such as DDoS. We also use a Gunicorn WSGI web server to host the tool, which provides a standard interface for connecting web servers and applications. This allows us to easily deploy and run the tool on a variety of platforms and environments. Finally, the REST API is written in Flask, a microweb framework for Python. Flask allows us to quickly and easily build the API, and provides a number of useful features such as request routing, template rendering, and error handling.

A.3 Front-End

The front-end of the tool is responsible for rendering the user interface and interacting with the user. It communicates with the back-end through the REST API and is responsible for handling user input and rendering the visualisations and information retrieved from the back-end. The front-end of the tool is responsible for rendering the user interface and interacting with the user. It contains four main components: two visualisation modules for displaying a map of data sharing locations (*legal*) and for visualising the knowledge graph (*graphing*), an information retrieval module with a *search* function, and a component for viewing *metadata* of an SDK such as the company name, website, and size. These components work together to provide users with the necessary tools to explore and understand the data sharing relationships between different SDKs and libraries.

The front-end of the tool is implemented using Vue 3.0, a progressive JavaScript framework for building user interfaces. We chose Vue for its simplicity, ease of use, and strong developer community. Alongside Vue, we also used Vuetify, a Material Design component library, to provide a consistent and visually appealing interface for the tool. For visualisation, we made use of D3.js, a powerful JavaScript library for manipulating and displaying data in a variety of formats. D3.js allows us to create interactive and dynamic visualisations of the knowledge graph, which is crucial for helping developers explore and understand the complex data sharing relationships between different SDKs and libraries, as outlined in D3.

A.4 Architecture

Figure 5 depicts the architecture of DataAvalanche.io. DataAvalanche.io has a client-server architecture, with the front-end serving as the client and the back-end serving as the server. The front-end is implemented using Vue 3.0 and Vuetify, which provide a framework for building user interfaces and a design system, respectively. These technologies were chosen for their ease of use, performance, and support for modern web standards. The back-end consists of a WSGI web server running Flask, which is a lightweight Python web framework. The back-end is responsible for serving the REST API, which provides access to the knowledge graph stored in the server. The knowledge graph is stored in a JSON object, with each key-value pair representing an SDK and its metadata, including a list of data sharing partners. To ensure the security and reliability of the tool, an NGINX reverse proxy is used to manage the incoming traffic to the

https://d3js.org/
https://vuejs.org/
https://vuetifyjs.com/en/
https://gunicorn.org/
https://flask-rest-api.readthedocs.io/en/stable/
https://nginx.org/en/

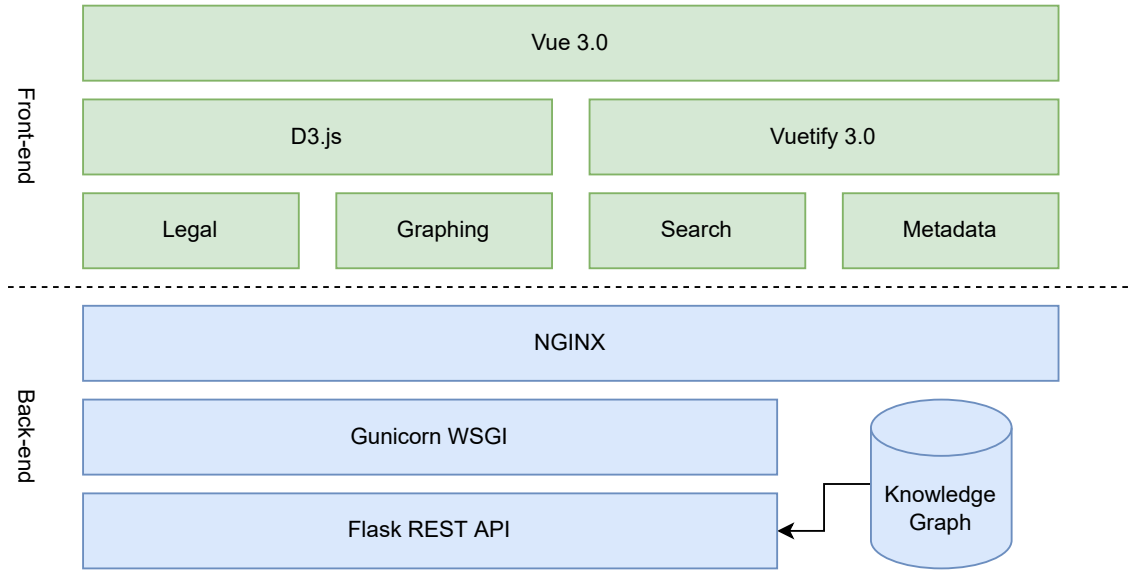

Fig. 5. System architecture of DataAvalanche.io. The front-end is implemented using `Vue 3.0` and `Vuetify`, while making use of `D3.js` for visualisation. On the back-end we run an `NGINX` reverse proxy connecting to a `WSGI` web server. The `REST API` is written in `Flask`.

back-end. This helps to protect the back-end from malicious attacks and ensures that the tool can handle high levels of traffic. We elaborate on the implementation details of this in the next section.

B COGNITIVE WALKTHROUGH OF DATAAVALANCHE.IO

For the cognitive walkthrough of our web-based tool, DataAvalanche.io, participants executed the following list of tasks:

(1) First, participants opened DataAvalanche.io and shared their screens with us.
(2) We allowed participants to glance over the tool and explore a bit on their own while voicing their initial thoughts and expressions.
(3) For the interview, we had pre-selected two SDKs, Adjust and Unity Ads, which were available as chips below the search box. We told participants to select one of the two SDKs to explore more in depth section by section.
(4) Participants next scrolled to the *metadata* section of DataAvalanche.io, where they could see important privacy and SDK information.
(5) Participants scrolled to the privacy-friendly alternatives, where we asked them about when and how they would consider switching to SDKs.
(6) Participants then scrolled to the *legal* section, where they could see a map of the countries the SDK shares data to and with. We asked participants if they were surprised by this behaviour and what their thoughts were. We also asked them if they would find this useful and how they would imagine using this in their development practices.
(7) Participants then scrolled to the data sharing partners, which showed the knowledge graph in tabular and graph format. Similarly, we asked participants what they thought of the privacy behaviours of the SDK and whether this

is what they expected. We also asked how they imagine using these features of the tool in their own development practices.

(8) Finally, users scrolled to the bottom where they had the opportunity to read FAQs.

C ADDITIONAL REQUIREMENTS

Table 2. Additional requirements formulated by our participants to improve DataAvalanche.io.

Feature	Explanation
Ads	Add information about popular ad networks and their privacy practices and help developers make informed decisions about which ad networks to use.
Categorise SDKs	Organise SDKs by their function to help developers easily find the tools they need.
Sort by popularity	Rank SDKs by popularity among developers to see the most commonly used and trusted ones.
Embed documentation	Include relevant documentation and code snippets directly in the tool to make it easier for developers to find and use the information.
Tutorials	Provide instructional materials or tutorials to help developers understand how to use the tool and make informed decisions about which SDKs to use.
Privacy breaches	Display any past privacy breaches or issues that a particular SDK has had to help developers understand the risks associated with certain tools.
App upload	Allow developers to upload their app and receive a privacy check or comparison to identify potential privacy issues and make improvements.
SDK comparisons	Compare the privacy features of SDKs to help developers make more informed decisions.
Mark favourites and bookmarks	Allow developers to mark their favourite or frequently used SDKs, or bookmark ones they want to keep track of, to quickly access the tools they need.
More data sharing details	Provide more detailed information about the types of data that each SDK shares to help developers understand the potential privacy implications.
Who is benefiting from the data	Use data or visualisations to show which companies are benefiting most from the use of certain SDKs to help developers make more informed decisions about the trade-offs they are making.
Filter SDKs by countries	Allow developers to filter SDKs by the countries that they share data with to help them select tools that align with their privacy goals and comply with regulations such as the GDPR.
Trending	Display the most popular or frequently used keywords or categories to help developers quickly find the tools they need and stay up-to-date on industry trends.
Summary privacy policy	Use natural language processing to summarise the privacy policies of SDKs to help developers more easily understand the key points and make informed decisions about which tools to use.
Community contributions	Allow developers to suggest open-source libraries or provide recommendations for privacy-friendly alternatives (e.g., GitHub links), to help developers discover new tools and contribute.
Add user data to map	Use a map to show where user demographics are, and show user distribution (e.g., percentage) per country, to help developers better understand the privacy implications of certain SDKs.
Ratings	Add ratings or a trust score to third party libraries, based on metrics such as user reviews or third party evaluations, to help developers make more informed decisions.