What They Say and What They Do: The Need to Unpack the Datafication Practices in Children's Recommendations

GE WANG, Department of Computer Science. University of Oxford, UK
JUN ZHAO, Department of Computer Science. University of Oxford, UK

In this position paper, we argue for the need to investigate and unpack how social media platforms conduct datafication practices, especially for children. As a starting point, we reviewed and analysed on what statements were made in their data policies regarding their collection and use of data. We outlined an agenda to support research addressing if and how their statements align with things done in practice, and future research addressing the phenomena.

Additional Key Words and Phrases: datafication, data policy, child development

ACM Reference Format:
Ge Wang and Jun Zhao. 2022. What They Say and What They Do: The Need to Unpack the Datafication Practices in Children's Recommendations. In *6th International and Interdisciplinary Perspectives on Children & Recommender and Information Retrieval Systems (KidRec) Information Retrieval Systems for Children in the COVID-19 Era; co-located with ACM IDC, June 27, 2022, Braga, Portugal*. ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

Today, children are spending an unprecedented amount of time online. Ever since the COVID-19 era, children are spending more time online than ever before. A recent report by Ofcom in 2020 (Life in Lockdown report) shows that in the UK, 96% of children aged 5-15 are online, and more than half of ten-year-olds have their own smartphones or tablets [8]. This rapid adoption and increasing reliance on the online world has raised corresponding concerns about the long-term effects of *datafication*, in which children's actions are pervasively recorded, tracked, aggregated, analysed, and exploited by online services in multiple ways that include behavioural engineering, and monetisation [6, 14]. Such datafication is practically impossible to avoid, or undo through deletion [3]. Children's life are now being routinely quantified, measured and used to profile and predict practices that could in return, have short and long-term implications for them [2, 4, 14]. Such activities take place invisibly behind the scenes of apps and services, and are less well understood or discussed as risks than other kinds of more easily characterised harms, such as the collection or disclosure of particular kinds of sensitive data.

While there has almost been a mutual understanding on online platforms, especially online social media platforms are collecting and processing users' data, which includes cross-platform data sharing, profiling and the application of such profiling results on their recommendations. Little is known about the exact ways in which such datafication practices were conducted, and how children's data have been used to make decisions on their recommendation results. On the one hand side, online service platforms provide users with extremely long and complicated privacy policies and data polices, on the other hand, it is hard to tell or assess on what was stated in these policies, partly due to them being vague and hard to actually test on some of the statements they made.

Manuscript submitted to ACM

In this paper, we argue that the research community needs to consider and address the possibility of unpacking the black box of how social media platforms (and online service platforms in general) process children's data. Especially, from the perspective that whether the online service platforms are actually in compliance with what they stated in their data polices. We support our position with a detailed review of the data policies of three major social media platforms that children interact with the most, namely Instagram (Meta), YouTube (Google), and TikTok (ByteDance). We outline several potential directions for further research through proposed research questions. We hope to invoke the discussion around unpacking the datafication practices around children's recommendations, and how such practices could have long-term impact on children as they grow up.

2 DATAFICATION PRACTICES IN CHILDREN'S RECOMMENDATION

Datafication refers to the process that children's actions are pervasively recorded, tracked, aggregated, analysed, and exploited by online services in multiple ways that include behavioural engineering, and monetisation [5, 6]. Livingstone et al. [3] categorised children's online data into three main categories: *data given*, which is the data contributed by individuals during their participation online; *data traces*, which is the data left by participation online and captured via data-tracking technologies such as cookies, web beacons or device/browser fingerprinting, location data and other metadata; and *inferred data*, the data derived from analysing data given and data traces, often by algorithms, possibly combined with other data sources. At the core of the datafication practices is service providers ability to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements [1]. Such practices have been widely found in the online services children interact with on a daily basis. Google and Yahoo were found to make inference on its users based on data such as demographic data, data on interests and attitudes [10]. Facebook was found to use users' browsing history for constructing interest feeds [11]. Similarly, Research on Instagram showed that the platform have tried to nudge young teenagers towards contents that were of idealised body images [9], and studies with YouTube users found that users have been nudged to maximise their screen time based on carefully designed/selected recommended content [9]

3 DATA POLICY ANALYSIS

In this section, we present our review on the data policies of some major social media platforms that are consumed the most by children, namely Instagram (Meta), YouTube (Google) and TikTok (ByteDance). Specifically, we analysed *what they say* - how the *datafication* practices around recommendations for children were articulated in these documents, which provide the foundation for comparing these statements with our next-step investigation on*what they do*, how the datafication practices around recommendations were conducted in practice.

In terms of access to service, only Instagram requires users to sign up for an account before they can use the service. Both YouTube and TikTok do not require that, and any users could access their service as "visitors" without signing in. As a result, previous studies have found that even when both YouTube and TikTok has their "child-specifc" version (e.g. YouTube Kids), most children were still using the main platform. Thus in our review of data policies, we did not include the data policies specifically on YouTube Kids and TikTok for younger users.

3.1 "What Information We Collect"

All three platforms acknowledged users the collection of their information, although in different level of details. The types of information they collected mainly come from three categories: *device information, information provided by users* and *information from partners and other sources.*

Starting with *device information*. Apart from the standard device information such as device attributes, device operations, device signals etc, Instagram explicitly mentioned that they will combine the device information collected *across different devices a user uses*. Apart from that, all three platforms mentioned the use of Cookies alongside with the third-party tracking technologies used. To be more specific, the Cookies were mainly used on four purposes, including strictly necessary, functionality, Analytics and performance (tracking technologies to analyse how you use the Website, including which pages you view most often, how you interact with the content, measure any errors that occur and test different design ideas), which is directly related to the fourth category of Cookies - Advertising and marketing (tracking technologies help us, and third parties in some cases, to promote our services on the other platforms and websites and measure the effectiveness of our campaigns). As we were reviewing the data/cookies policies, we noticed the strong connection between all three platforms, TikTok for example, mentioned in their cookies policies, that they are using Google Analytic to connect Google searches with visits to TikTok. Also to connect activity on a Website with the Google Ads network for the remarketing of the products and services, and to generate demographic and interest reports; Similarly, TikTok were also using the Facebook Pixel, which track ad conversions from Facebook ads so as to optimise their ads, build targeted audiences for future ads, and remarket to people who have already taken some kind of action on TikTok.

Regarding *information provided by users*, this refers to the users' activities on these platforms. For YouTube (Google), this refers to terms searched for, videos watched, views and interactions with content and ads, voice and audio information, purchase activity, people with whom you communicate or share content, activity on third-party sites and apps that use our services, activity on third-party sites and apps that use our services, and chrome browsing history. On top of that, both Instagram and TikTok mentioned the data collection of "content you create or publish through the Platform" (which includes direct messages and contacts) as well as associated metadata. TikTok in particular, also collects content characteristics and features - they detect and collect characteristics and features about the video and audio recordings that are part of the User Content, for example, by identifying objects and scenery; the existence or location within an image of a face or other body parts; and the text of words spoken in the User Content.

Finally, with respect to *information from partners and other sources*. All three platforms mentioned the use of Advertising, Measurement and Data Partners. TikTok mentioned that Advertisers and measurement and data partners would share information with platforms such as mobile identifiers for advertising, hashed email addresses, and event information about the actions user've taken on a website or app. Some of our advertisers and other partners enable us to collect similar information directly from their website or app by integrating platform's API. Similarly, Instagram would collect information from their partners on users' activities off their products - including information about user's device, websites users visit, purchases users make, the ads users see and how they use their services - "For example, a game developer could use our API to tell us what games you play, or a business could tell us about a purchase you made in its shop. We also receive information about your online and offline actions and purchases from third-party data providers who have the rights to provide us with your information.".

3.2 "How We Use Your Information"

The data collection, especially those from partners and third-party websites lead to the natural question of - why platforms need these data and how they use it. Apart from the good old statement of "we collect your information to provide better services for you" (which you can pretty much expect to see as the first thing every platforms put on their data policy these days). We managed to find some interesting hints based on what the three platforms stated.

Firstly, YouTube (Google) mentioned the use of data on personalised advertisements. Although on their data policy, they stated that “We don’t share information that personally identifies you with advertisers, such as your name or email, unless you ask us to”. That does not stop them from personally identifies people as belonging to a certain group, based on their gender, age, economic status, educational status, homeowner or not and etc. According to Google, while they do not share these user profiles with their party companies, they did not mention anything regarding the share of profiles within their Google partners. And they allow third party companies to come to them with a certain group of customers they want to target at, and Google would help these companies identify these groups. Similarly, Instagram (Meta) also mentioned this process of putting tags on people. For example, an Instagram users could be tagged as a 30 years old, female, living in Hammersmith London, interested in cycling, films, cooking, and classified as iPhone user, car shopper, gamer. Such a user would be pushed to advertisers looking for desired audience who is between 18-35 years old, female, within 20 miles of my store, interested in cycling, and is mobile user. Apart from the categorising and tagging, TikTok also mentioned this concept of *inferred information*, carefully implying that users’ attributes (such as age-range and gender) and interests could be inferred based on the information they have about the users. Adding that they would only use such inference to keep their platform safe, and where permitted, to serve users personalised ads based on these inferred interests.

3.3 General Data Policies V.S. Kids Data Policies

As we mentioned before, all platforms have their “kids” (under 13s) version of data polices, which collects much less data and did much less profiling on collected data. For example, YouTube Kids (which as a kid-version for the YouTube platform) stated that they would only use unique identifies to provide contextual advertising, including ad frequency capping. The app does not allow interest-based advertising or remarketing. Similarly, TikTok stated that under 13s would have limited access their service, such as they can’t exchange messages with other users, and other users can’t view their profiles, and they would only serve contextual advertising for them. Meanwhile, Instagram completely bans the use of users under 13s.

The reasons why these kid versions of the apps were set up date back to regulations and legislative frameworks on children’s online safety. A recent analysis of AI frameworks and design codes for children identified ten common principles that were considered as the most important when providing online services for children [13]. Some of the most relevant principles here include the fairness and non-discrimination of offered services (recommendations), transparency of how services (recommendations) are being offered, non exploitation and manipulation, and to ensure the inclusion of and for all children. Comparing these principles with the above data polices, there’s a significant misalignment between what the platforms were doing and the regulations made for children. While online social media platforms could argue that their service were not intend for children by putting out the so-called kids data polices, there has been clear evidence that children, especially those under 13s are still on these platforms, and in fact, have quite heavy usage on such platforms [12]. A recent report on 2,002 US children showed that 45% of kids under 13 were already on Facebook, and 40% already used Instagram [12]. YouTube in particular, although has a ‘YouTube Kids’ version that was claimed to be for under 13s, the most recent Ofcom report still showed that YouTube remained to be the most popular video-sharing-platform among 8-12 year-olds, and more than 85% of preschoolers were found to most commonly used YouTube to watch content [7].

4 PROPOSED RESEARCH AGENDA

Our analysis showed that there is a strong need in terms of unpacking the black box of social media platforms, in terms of how they conduct datafication practices on children's data - whether such practices align with their own statements, and what will be the implications of such datafication. We therefore propose the research agenda as follows:

Inferred Age. We would like to verify social media platforms statements on "we do not know children are also using our platform". To be more specific, we would want to simulate what's it like to be a child on these three major social media platforms, thus to find indication on could social media platforms infer the personal information about a child, including his or her gender, and other inferred interests, and how would that experience be like for a young child. More importantly, would social media platforms be able to infer about a child's age after all: (1). Are social media platforms capable of inferring about a user's age? (2). What data and metrics can help us to identify this? (3). If so, what information and how social media platforms use it to make such inference?

Datafication Experience. While almost all children are using some kind of online services these days, in particular in the COVID-19 era, how online social media platforms make their recommendations is still unexplored in the sense that there's no clear model on how they make the recommendations (what data is used, how data is aggregated and analysed), and what are the implications on this processing of data: (1) How data was aggregated and analysed, can we identify some kind of profile being made for each user? (2) How can we distinguish recommendations made based on pure browsing history, and those based on profiles containing personal information (e.g. gender, economic status), or does such distinction even exist? (3) How are these profiles applied for each user, and what are the potential concerns of that?

Mitigation. To ensure better social media platform experience for children, especially in terms of their content and advertisement recommendations, we believe we need to mitigate some negative impacts that could be brought by existing datafication practices. Some potential questions for reducing such concerns include: (1) What are the negative impact brought by the datafication practices on social media platforms? (2) What design goals can help mitigate such concerns from recommendations? (3) What adjustments can be made to the current social media ecosystem (data infrastructure, algorithm, interface) to mitigate such concerns? (4) What kind of datafication experience do children want to have on the social media platforms?

REFERENCES

[1] 2020. What is automated individual decision-making and profiling? https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr

[2] Nick Couldry and Ulises A Mejias. 2019. *The costs of connection.* Stanford University Press.

[3] Sonia Livingstone, Mariya Stoilova, and Rishita Nandagiri. 2019. Children's data and privacy online. *Technology* 58, 2 (2019), 157–65.

[4] 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.

[5] Giovanna Mascheroni. 2020. Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology* 68, 6 (2020), 798–813.

[6] Ulises A Mejias and Nick Couldry. 2019. Datafication. *Internet Policy Review* 8, 4 (2019).

[7] Ofcom. 2021. Children and parents: media use and attitudes report 2020/21. https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes

[8] Ofcom.org. 2020. Ofcom Children's Media Lives: Life in Lockdown. (2020).

[9] Joint Committee on the Draft Online Safety Bill. 2021. Draft Online Safety Bill. https://committees.parliament.uk/publications/8206/documents/84092/default/

[10] Ashwini Rao, Florian Schaub, and Norman Sadeh. 2015. What do they know about me? Contents and concerns of online behavioral profiles. *arXiv preprint arXiv:1506.01675* (2015).

[11] Bernhard Rieder. 2017. Scrutinizing an algorithmic technique: The Bayes classifier as interested reading of reality. *Information, Communication & Society* 20, 1 (2017), 100–117.

[12] THORN. 2021. Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking. https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf

[13] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. Informing Age-Appropriate AI: Examining Principles and Practices of AI for Children. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–29. https://doi.org/10.1145/3491102.3502057

[14] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.