April 2022

Datafied Childhood: How Children Perceive Datafication Risks Around Them

UNIVERSITY OF OXFORD

KOALA
Project Report 5

Executive Summary

It is clear that the world of digital is here to stay. Online technologies are increasingly important for children, providing access to vital education, socialisation, participation, well-being, and entertainment resources.

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.

In the winter of 2021 we conducted one-to-one interviews with 48 children aged 7-13 in the UK on their understanding of the datafication practices online. Our research has shown that children have a good understanding of certain datafication practices, such as the existence of data collection and content recommendation. However, at the same time, we identified that children had less awareness concerning some more advanced datafication practices, such as cross-platform data sharing, and online profiling.

Key Findings

1. **Children care significantly about various aspects of datafication.** They already possess rudimentary conceptual understandings of it, and a significant willingness to take action to shape it to their desires, possibly even more than adults.

2. **Children do not always comprehend online datafication practices to a full extent**, especially in terms of who are involved in the data processing - developers? computer? algorithms? or companies?

3. **Children do not always recognise who owns their data online,** and sometimes think platforms and companies have the right to their personal data.

4. **Children do not always recognise data is being transmitted across platforms,** and have difficulty in recognising that data collected from them on one platform (e.g. YouTube) could also be transmitted to and used by other platforms (e.g.Amazon).

1 Introduction

Today, children are spending more time online than ever before. Online technologies are increasingly important for children, providing access to vital education, socialisation, participation, well-being, and entertainment resources [1]. In the UK, for instance, 96% of children aged 5-15 are online, and more than half of ten-year-olds go online using their own devices [2] Thus, the Internet has now become seen as an essential enabler for children to learn, have fun and grow, especially during a pandemic, in which much of ordinary life has shifted online [2].

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 [3]. At the core of this datafication is online services' ability to make ***data inference*** on users, that is to analyse data, supported by algorithms, with the aim to evaluate certain personal aspects relating to a natural person[1], in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements [4]. Such datafication is practically impossible to avoid, or undo through deletion [1]. 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 [5]. 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.

We thus feel that it is important to understand children's perceptions and how they interpret those online datafication practices, not only because it might identify common barriers and knowledge gaps in children, but also lead to better ways to empower and protect them online. We report our results based on one-to-one interviews with 48 children, aged 7-13, from UK schools undertaken between November and December 2021. Through situating our findings under a critical algorithmic literacy framework, our findings provided some immediate indications regarding how we could better support children in the datafied society through more transparency and autonomy-supportive designs.

2 Methodology

Given our focus on investigating children's perceptions of datafication practices online. We chose the YouTube platform to be used as an example. The reason we selected YouTube as the exemplar datafication platform is because it contains a variety of data processing practices, and is familiar by most children for it being one of the most consumed entertainment platforms among children around the world [2].

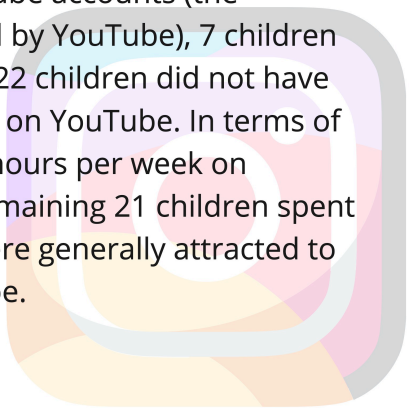
We conducted one-to-one semi-structured interviews with children to elicit their responses to a collection of tasks that attempted to recreate their everyday experiences on YouTube, followed by a collection of scenarios that reflected different types of datafication practices in relation to how YouTube could process and make use of their data.

3 Key Findings

Through careful data analysis of interview recordings, we have achieved some positive findings with respect to children's level of awareness. However, at the same time, we have also identified critical gaps in children's understanding.

Children's Overall Experience and Usage on YouTube

Most children (37/48) in our study owned their own device (phone, tablet, computer), while the rest used their parents' devices or shared with their siblings. A fair amount of children (19/48) had and would sign in to their own YouTube accounts (the minimum age to have an account is 13 years-old, as claimed by YouTube), 7 children used their parents' YouTube accounts, while the remaining 22 children did not have their own account and would not sign in to an account at all on YouTube. In terms of usage, around a third of children in our study spent over 5 hours per week on YouTube, 11 others spent 3 to 5 hours per week, and the remaining 21 children spent 1 to 3 hours per week on YouTube. Children in our study were generally attracted to game videos, animations and educational videos on YouTube.



Children do not recognise who's involved in the datafication practices

The majority of the children reported that they believed that it was not `human' that `manually' processed their data. They would describe the operators as some kind of `machine' using terms such as `a computer', `a robot', `an AI' or `010101'. On the other hand, only a minority of children (11/48) had some awareness that algorithms were used to process their data. Interestingly, when trying to explain how algorithms follow a set of rules, many children did not recognised the `data-driven' aspect of algorithm, but instead described algorithms as following very exact rules. For example, when asked about why they think algorithms could generate gender-stereotypical contents for people, many children tend to think that algorithms were set to only give certain content to certain gender. Only very few children recognised that there might be inherit bias within the data input.

Children do not recognise the data transmission across platforms

A key theme emerged was children's perception about where these data about them went: the majority of them believed that the data were collected by a certain platform and would be only used by this platform locally. Many of these children reported that they thought only their data resulting from their interactions on a particular platform would be used to make inference about them. To be more precise, almost all children only mentioned the possibility of YouTube using data generated from their interactions on YouTube (e.g. videos watched); none, meanwhile, mentioned the possibility of YouTube using other data sources, such as what they did on other websites.

Children were unsure about their data ownership

There was a strong theme of discussion around what data children thought as their `personal data'. Interestingly, while almost all children strongly believed that data such as their location information and their age and gender was their own personal data, a majority of children regarded their behavioral data (e.g. videos watched, terms searched, channels subscribed) as `not personal to me' and could be accessed and quantified by platforms. Furthermore, we found an interesting common thinking among children regarding who owns their behavioral data online. A surprising amount of children (32/48) thought that it was YouTube (or the platform) that own these data, or have the rights to these data

Recommendations for Parents & Caretakers

1. **Continue to talk to children about being careful online**, because they are indeed facing challenges in their use of digital devices every day and they really care about protecting their personal space online!

2. **Remind children that everything they do online will leave a permanent record.** Such data will be transmitted across the Internet and build up a permanent record - It can't be erased.

3. **Remind children that the things they see on the Internet could be directly targeting them.** Online profiling is a common practice nowadays, which involves targeting users based on collected data of them.

4. **Do not be afraid of talking about digital technologies with children.** They would enjoy learning with you and sharing their knowledge with you!

References

[1] Livingstone, Sonia, Mariya Stoilova, and Rishita Nandagiri. "Children's data and privacy online: growing up in a digital age: an evidence review." (2019).

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

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

[4] ICO. What is automated individual decision-making and profiling. (2020).

[5] ICO. Nick Couldry and Ulises A Mejias. The costs of connection. Stanford University Press. (2019).

Acknowledgement

We thank all the families and children who have contributed their time and knowledge to our study. Without their support, this study would not have been successful!

The report is based on a full academic publication **'Don't Make Assumptions About Me': Understanding Children's Perception of Data Inference Online**, as part of the project of KOALA (https://koala.web.ox.ac.uk/home): Kids Online Anonymity & Lifelong Autonomy, funded by EPSRC Impact Acceleration Account Award, under the grant number of EP/R511742/1.

Authors

Ge Wang (ge.wang@cs.ox.ac.uk)

Jun Zhao (jun.zhao@cs.ox.ac.uk)

Nigel Shadbolt: (nigel.shadbolt@cs.ox.ac.uk)