‘HOME LIFE DATA’ AND CHILDREN’S PRIVACY

Report submitted together with a response to ICO Call for Evidence for Age Appropriate Design Code, 18th September

A report by Dr Veronica Barassi, Department of Media, Communications and Cultural Studies, Goldsmiths University of London/ Principal Investigator on Child | Data | Citizen Project, Funded by the British Academy http://childdatacitizen.com)

Co-signed by Gus Hosein, Executive Director, Privacy International
Supported by Jeff Chester, Executive Director, Center for Digital Democracy

INTRODUCTION

The development and domestication of AI – together with the extension of smart technologies – is rapidly transforming our homes. Powerful new applications and business models are emerging that pose a threat to the privacy of children and their families. Home automation is becoming a rapidly expanding market. A report published in January 2017 by Juniper Research – that specializes in identifying and appraising new high growth market sectors within the digital economy - estimated that smart home hardware and service, which include entertainment, automation, healthcare and connected devices is set to drive revenues from $83 billion in 2017 to $195 billion by 2021. They also estimated that the ‘big four’ (*Alphabet/Google, Amazon, Apple* and *Samsung*) companies – which at present dominate the smart home market – will further solidify their position, with Amazon securing a leading role (Juniper, 2017). Home Hubs threaten to further socialise kids to divulge their data, which is one reason design code in the home must create safeguards across all ages. By introducing the concept of **home life data** (Barassi, 2018), in this report we wish to draw attention to the fact that the data that is being collected by home hub technologies is not only personal (individual) data but it is household, family and highly-contextual data. Understanding the complexity of home life data makes us appreciate the fact that children’s data is too often intertwined with adult profiles. We believe that the ICO should include home automation and home life data in the list of areas to take into consideration when developing age appropriate code.

HOME HUBS: A COMPLEX BUSINESS MODEL

When we think about home automation, we need at first to break down the incredible variety of smart technologies that are entering our homes. The pace of technological transformation, the extensiveness and pervasiveness of the developments in the Internet of Things makes it incredibly difficult to have a comprehensive overview. Yet broadly speaking (and at the time of writing) home automation is enabled by different sets of technologies, which include:

- *artificial intelligence devices* (e.g. virtual assistants, robots that act as home assistants; artificial intelligence toys, etc.);
- *entertainment devices* (e.g. smart TVs, whole house wireless music systems; video games, etc.)
- *home appliances* (e.g. smart fridges; smart toilets; smart washing machines etc.)
- *security technologies* (e.g. smart locks; surveillance cameras; alarms, which can detect intruders and are equipped with special sensors to detect floods, fires etc.)
- *energy and utilities monitoring and measuring tools* (i.e. meters that monitor water and energy consumption, etc.)
- *lighting monitoring devices* (e.g. smart bulbs and switches that can be controlled at a distance, etc.)
- specific *solution devices* (e.g. devices that offer different specific solutions, such as support with recycling or intercom solutions, etc.).

One of the challenges that businesses are facing at the moment relates to the fact that in order to build a truly automated home all the different technologies need to communicate with one another (Zuckerberg, 2016). It is for this reason that, in the last few years, we have seen the emergence of a new *business model* developed by the so called Big Four of Home Automation (Amazon, Google, Apple and Samsung) for home automation and domestication of artificial intelligence, which we will refer here as '**home hubs'**.

The business model of home hubs is quite complex and is structured (broadly speaking and again at the time of writing) by four different dimensions:

- The first dimension is of course the *AI virtual voice assistant* (Amazon Alexa, Google Assistant, Apple Siri, and Samsung Bixby). Virtual assistants are usually operated by home speakers (however as mentioned by Prof. Leah

Lievrouw, UCLA, in a joint interview on data and privacy with Dr. Barassi in 2018, these are not only 'speakers' but also recording technologies). Virtual assistants can be integrated into a variety of home technologies (as we shall see below, and especially if companies have an open platform model). The AI assistants operate through voice recognition and are connected to specific profiles and accounts (e.g. Amazon and Google).

- The second dimension of the business model is created by the different *'services'* that users can access through the assistant. In very simplistic terms, we can understand these services as 'voice operated apps' that families can access through the interaction with their virtual assistant (e.g. Alexa Skills, Google Actions, Siri Shortcuts, Bixby Commands). These services are continuously expanding. To regain competitive advantage over its competitors, Apple, for instance, is developing Siri Shortcuts by tapping into its 2 million apps at the moment. In order to extend Alexa Skills Amazon, for instance, created the Alexa Fund, which provides up to $100 million in venture capital for companies that build Alexa Skills Kit. In the last two years Alexa Skills have increased from 5,191 in November 2016 to 30,006 in March 2018 (Kinsella, 2018).

- The Alexa fund also invests in the third dimension of the business model of home hubs: the creation of '*compatible technologies'*. All the different companies are investing in the development of their own smart technologies (e.g. Apple, Samsung) or in funding other companies that include their voice operated assistant in their own technologies (Amazon). At present we are seeing homes being built with these technologies. In 2018, for example, Amazon signed a deal with the Lennar Corporation, which is building 35,000 automated homes in Florida, which are operated by Alexa.

- The fourth dimension of the business model is defined by *mobile home apps.* These are apps that enable to control the home remotely from your phone (Alexa App; Google Home app; Apple iOS Home app; Samsung Smart Home app).

HOME LIFE DATA AND CHILDREN'S PRIVACY

When we think about home hubs and their complex business model, the question about children's data and privacy is not a simple one to tackle. There are three different problems that we encounter as we try to address this question: a) the complexity of home life data b) the newness of home data environments c) the secrecy of algorithms

The Complexity of Home Life Data. Debates about the privacy implications of AI home assistants and Internet of Things focus a lot on the the collection and use of *personal data*. Yet these debates lack a nuanced understanding of the different data flows that emerge from everyday digital practices and interactions in the home and that include the data of children. When we think about home automation therefore, we need to recognise that much of the data that is being collected by home automation technologies is not only personal (individual) data but **home life data** (Barassi, 2018) and we need to critically consider the multiple ways in which children's data traces become intertwined with adult profiles. An attention to *home life data* should include a focus to the following categories:

1) *household data* – Home hubs and smart technologies collect a wide variety of household data from shopping lists to energy consumption and gather key information on families' behaviours, choices and routines (including the ones of children).
2) *family data* – Home hubs and smart technologies' Terms and Conditions are usually focused on explaining what happens to personal (individual) data. Yet they don't refer to whether they use family data. What is becoming clear is that, to enable multi-user functions, companies are aggregating profiles (see the example of Amazon Household Profile Case Study in the Appendix). Aggregated profiles, however, constitute a risk for children's privacy. Let's imagine that you are having dinner with a friend who has a child who suffers from diabetes and you might ask Google assistant or Alexa to look for information on 'diabetes in children'. That information would be automatically stored on your profile. Let's also imagine that in the weeks to come you feel concerned about your own child getting diabetes and you start looking for information on symptoms. All these data traces would imply that you probably would be profiled as "parent" with a "diabetes interest'" (this is a guess because there is so much secrecy about the ways in which we are being profiled). If this is the case, the the question emerges naturally: if you shared your 'household' profile with your child, and you were profiled as a parent

with a diabetes interest, would your child be profiled as possibly diabetic? The problem is that we don't know the answer.

3) *biometric data* – Most Virtual Assistants and smart technologies rely on the gathering of biometric data (voice recognition or facial recognition), including the one of children. Yet privacy policies often tend to group this data under the generic umbrella term of 'biometric data' and do not differentiate the one of adults from the one of children.

4) *highly contextual data* – To function, AI technologies do not gather only personal data but contextual data. Yet the data policies of home hubs fail to discuss how companies use this data. The following examples are particularly illustrative (although a bit dated) of the ways in which developers are thinking about context:

   > "*Understanding context is important for any AI. For example, when I tell it to turn the AC up in "my office", that means something completely different from when Priscilla tells it the exact same thing. That one caused some issues! Or, for example, when you ask it to make the lights dimmer or to play a song without specifying a room, it needs to know where you are or it might end up blasting music in Max's room when we really need her to take a nap*." (Zuckerberg, building Jarvis 2016)
   >
   > "[*Hello Barbie] should always know that you have two moms and that your grandma died, so don't bring that up, and that your favorite color is blue, and that you want to be a veterinarian when you grow up,''* (Wulfeck, ToyTalk in Vhalos, 2015).

5) *messy data* – The data produced by family life is inevitably messy and full of imprecisions and overlaps. Families often do not use these technologies as they are designed to be used. This is not only because, on an average family da, technologies, profiles etc. always ooverlap, and this confuses algorithms, but also because families often input inaccurate data in their technologies, to use them tactically or because they do 'not want to share too much'. When we think about data traces and profiling than we need to ask ourselves: is this broken, inaccurate data used to profile families and children?

The Newness of Home Data Environments. Home hubs are collecting and processing different types of children data, from biometric data (voice recordings, facial recognition) to personal interests and details (entertainment data, other contextual personal data) but they are not *designed for or targeted at* children. Last year Mattel cancelled its Aristotle AI assistant for kids amidst privacy concerns. This year we are seeing a growing debate from lawmakers in the U.S. about Amazon's use of children's data in the Amazon Dot Echo for Kids. Yet we are seeing very little debate about home hubs and smart technologies that are targeted at adults but that children *encounter* (Montgomery, 2015) in everyday life, and that collect their personal data. These new data environments challenge some of the effectiveness of regulations such as COPPA or the GDPR to protect children's privacy in the automated home.

Secrecy of Data Use: Home hubs rely on a business model that is extraordinarily complex and involves an incredible plurality of companies and agents. Many internet companies accumulate vast amounts of data for unclear purposes and often share with data brokers. Without clarity we are left to presume that data from hubs will end up as part of the modern exploitative business models within the data brokering ecosystem. However, the ways in which companies gather, archive and sell home data or the ways in which they profile, sort and classify their users (including children) is still unknown because of the secrecy of algorithms (Pasquale, 2016) and lack of transparency of data brokers (FTC, 2014).

CONCLUSION:
THE IMPORTANCE OF NEW MEASURES/SOLUTIONS

It seems that companies are not recognizing the privacy implications involved in children's daily interactions with home automation technologies that are not designed for or targeted at them. Yet they make sure to include children in the advertising of their home technologies. Much of the responsibility of protecting children is in the hands of parents, who struggle to navigate Terms and Conditions even after changes such as GDPR. There is no acknowledgement so far of the complexity of home life data, and much of the privacy debates seem to be evolving around personal (individual) data. It is for this reason that we need to find new measures and solutions to safeguard children and to make sure that age appropriate design code is included within home automation technologies.

We recommend that the ICO supports further research or launches a review on the impact of home life data on children's privacy, we also recommend that the ICO includes the concept of home life data in current debates on children's data protection.

References:

Barassi, V. (2018) 'Home Life Data' chapter in the Child | Data | Citizen forthcoming book currently under review by MIT PRESS, more info at http://childdatacitizen.com

Federal Trade Commission. (2014). *Data Brokers: A Call for Transparency and Accountability*. Retrieved from https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission

Kinsella, B. (2018, March 22). Amazon Alexa Skill Count Surpasses 30,000 in the U.S. *Voicebot*. Retrieved from https://voicebot.ai/2018/03/22/amazon-alexa-skill-count-surpasses-30000-u-s/

Montgomery, K. (2015). Children's Media Culture in a Big Data World. *Journal of Children and Media*, *9*(2), 266–271. https://doi.org/10.1080/17482798.2015.1021197

O'Hara, D (2018) Conversation with Dr. Barassi by Dustin O'Hara, 3rd of May

Pasquale, F. (2016). *The Black Box Society: The Secret Algorithms That Control Money and Information* (Reprint edition). Cambridge, Massachusetts London, England: Harvard University Press.

Zuckerberg, M. (2016). Building Jarvis. Retrieved August 23, 2018, from https://www.facebook.com/notes/mark-zuckerberg/building-jarvis/10154361492931634/