

VOICE PRINTS AND CHILDREN’S RIGHTS

Dr. Veronica Barassi, Goldsmiths University of London, Child | Data | Citizen Project (Funded by the British Academy)

Dr. Patricia Scanlon, CEO, SOAPBOX Labs www.SoapBoxLabs.com

INTRODUCTION

With the rapid growth of voice assistants, and the underlying speech-recognition technologies at home, in schools, and in different contexts across the world, children’s speech data is collected, stored and processed in ways that were not possible before. Voice biometric information (voice-prints) can be extracted from this speech data. In the Case for the UNCRC General Comment on Children’s Rights and Digital Media - written for the Children’s Commissioner for England by Sonia Livingstone, Gerison Lasndown and Amanda Third - the authors mention the fact that technologies “able to monitor conversation” can be understood as a potential challenge for children in digital environments (2017:4). With this response to the call for submissions on the *General Comment on Children’s Rights in Relation to the Digital Environment* our aim is to direct the attention of the Committee towards the fact that - in the current digital environment - speech recognition technologies are raising significant threats to children’s rights.

This submission has been written as a joint collaboration between Dr Patricia Scanlon, CEO of SoapBox Labs, a ‘privacy by design’ business that specialises in children’s speech recognition and Dr Veronica Barassi, principal Investigator on the Child | Data | Citizen project at Goldsmiths University of London. In this response we argue that speech recognition technologies hold much promise and opportunities for children: they can advance literacy and widen participation and access to information. They can also foster creative interaction, playfulness and exploration for children.

Yet in the current digital environment these technologies pose a threat to children’s rights for two main reasons:

- 1) There is a lack of transparency and shared standards for the ways in which the speech data (from which biometric data can be extracted) of children is used and processed by companies. Without transparency and shared standards, companies can use this speech data from children to potentially create unique ID profiles; extract emotion and sentiment data and catalogue other sensitive information from transcriptions of their speech - that can follow them across a lifetime. This raises critical questions about children’s right to privacy (UNCRC 16), self-expression (UNCRC Article 13), freedom of thought (UNCRC, Article 14) and non-discrimination (UNCRC Article 2).
- 2) The speech recognition technologies that are used by children in today’s digital environment are often NOT designed for them and thus can expose them to adult-centred and possibly unsafe digital experiences that go against their best interests (UNCRC Article 3) and optimum development (UNCRC Article 6).

Dr Patricia Scanlon and Dr Veronica Barassi have decided to join forces in their response to the General Comment because they would like the Committee to pay attention to the opportunities of speech recognition for children, as well as to the challenges that are currently posed by a digital environment that does not place adequate attention to the risks and privacy implications of this technology. With this response, our aim is to take a first, yet important step towards the development of a worldwide call for action involving businesses, researchers, law-makers, and non-for-profit organizations to pay greater attention to how speech recognition can impact on children’s rights.

Note on the Authors:

Dr Barassi is a media anthropologist working in the Department of Media, Communications and Cultural Studies at Goldsmiths, University of London. She is the principal investigator of the [Child | Data | Citizen](#) project, which is funded by the British Academy and relies on an advisory board of international experts. The aim of the project is to provide a rich, qualitative analysis of children's data flows in family life. Dr Barassi worked with families in London and Los Angeles, with children between 0 and 13 years of age, whose personal information online is ruled by the Children's Online Privacy Protection Act (1998). The research relied on a multi-method approach, which combined 50 semi-structured in-depth interviews, two year of auto-ethnographic participant observation, 9 months of digital ethnography of the social media of 8 families, and qualitative platform analysis (4 social media platforms, 10 early infancy apps, and 4 AI devices and Home automation hubs). Dr Barassi has published extensively in the field, she is currently writing a book for MIT Press titled *Child | Data | Citizen*, and is the author of the '[Home Life Data and Children's Privacy](#)' Report, which is now used for evidence by the Information Commissioner Office UK

Dr. Patricia Scanlon is the CEO and Founder of SoapBox Labs and has spent more than 20 years in the speech technology industry. Since 2013, SoapBox Labs has developed the world's most accurate speech recognition technology for young children aged between 3 and 12 years old. Unlike the speech recognition that supports mainstream voice assistants like Alexa and Siri, SoapBox's state of the art technology works accurately with younger voices and was specifically designed to support voice interface experiences including literacy, English language learning, games, and accessibility needs both at home and in the classroom. Respecting children's data privacy is part of SoapBox Labs DNA. Their technology uses a "privacy by design" approach and the children's voice **data they gather is never shared, reconstituted or sold**. SoapBox Labs takes pride in being transparent about when and how they use the children's voice data they collect and it is only used in one instance, to improve the quality of the speech recognition technology they offer.

SPEECH RECOGNITION OPPORTUNITIES AND PROMISES FOR CHILDREN

Speech recognition technologies can lead to many important opportunities for children in the digital environment not only in terms of learning and access to information and technology, but also in terms of self-expression and participation. With the rapid rise in use of voice operated AIs (such as virtual assistants) and the inclusion of different IoT in everyday life, these opportunities have expanded. Today through interaction with speech-recognition technologies, children have unprecedented opportunities for striving in the digital environment. We believe that it is crucial that the Committee is aware of these opportunities and the 'importance of voice' in digital environments. Voice is important for the following reasons:

Speech Recognition, Self-Expression and Emotion: Voice-Operated technologies, and especially voice-operated AIs are offering something extraordinarily new in terms of human and computer interaction. Something that differs from other types of technological engagement. According to Zuckerberg (2016) for instance the fact that we can talk to these technologies implies that we are seeking more *emotional depth*. In his post on 'Building Jarvis' (2016) he notes:

"... once you can speak to a system, you attribute more emotional depth to it than a computer you might interact with using text or a graphic interface. One interesting observation is that ever since I built voice into Jarvis, I've also wanted to build in more humor. Part of this is that now it can interact with Max and I want those interactions to be entertaining for her, but part of it is that it now feels like it's present with us. I've taught it fun little games like Priscilla or I can ask it who we should tickle, and it will randomly tell our family to all go tickle one of us, Max or Beast."

In emphasising the emotional connectedness that people feel with voice-operated machines, Zuckerberg was definitely touching on a very current research question. A survey carried out

by Google/Peerless Insights on 1,642 users of voice-activated speakers, has shown not only that people are engaging with these technologies as if they were human by saying “please”, “thank you” and even “sorry”, to their virtual assistants, but also that 41% of people said that voice-activated virtual assistants feel like talking to a friend or another person (Kleinberg, 2018).

There are clear critical questions that arise when we think about children’s interaction with voice-operated AIs, and what it means to grow up in a society where children talk to technologies as if they were quasi-humans; where they command them and attribute intelligence and emotion to them (Druga et al, 2017). Much research is needed in this regard. Nevertheless we cannot overlook the opportunities intrinsic to these technologies, the potential for deep engagement, creative interaction, playfulness and exploration for children.

Speech Recognition, Learning and Participation: Speech recognition technologies can also lead to important opportunities for literacy, learning and widening participation. These opportunities have long been noted. In 1999, Raskind and Higgins (1999) conducted a study of children aged 9 to 18 with learning disabilities and concluded that speech recognition compensated for poor writing skills and also assisted students in reading and spelling. In 2006, Adams (2006) explored the benefits of speech recognition technologies for literacy in both children and adults, and in (2011) she expanded on her findings and discussed the importance of bringing speech recognition to the classroom to foster reading fluency and engagement.

When we think about the opportunities for literacy, learning and widening participation in relation to speech recognition technologies, we need to be critically aware of issues such as digital divide, and we cannot embrace techno-solutionist understandings that see these technologies as taking the place of broad publicly funded interventions in education and learning. Nevertheless we need to appreciate the power of these technologies to foster children’s education and learning and to improve access and inclusivity for children with disabilities or less-privileged children.

VOICE PRINTS AND CHILDREN’S RIGHTS

Whilst it is essential that we appreciate the promises and opportunities for children who use speech recognition technologies, it is equally important that we recognise the risks and threats of the current digital environment. Through our collaboration we have identified the following issues:

Children’s Data Privacy, Voice Prints and Unique ID Profiles

One of the greatest threats to privacy posed by speech recognition is represented by the fact that biometric data can be extracted from voice data. People believe voice, like breath, is ephemeral - that once spoken, words disappear forever. But in the modern age of voice technology this is no longer the case. A voice print, like a fingerprint, is a uniquely identifiable piece of data, and once that data is linked to an individual it can be used to infer many things about them, from their health to their socio economic and educational background. Today, voice data is recognised as having significant inherent value but this value can be leveraged for unintended and unethical purposes by companies whose ad-driven business models put individual privacy rights at risk. In the last few years, campaign groups and lawmakers have paid particular attention to how voice-operated virtual assistants or AI toys could impact on children’s privacy. In 2017 [Mattel cancelled its Aristotle AI](#) assistant for kids amidst privacy concerns. In 2018, we witnessed a [growing debate from lawmakers](#) in the U.S. about Amazon’s use of children’s data in the Amazon Dot Echo for Kids. In 2019, the Campaign for a Commercially Free Childhood has launched a new investigation into Dot Echo for Kids that has shown that Amazon retains children’s data also after parents tried to delete it. When we think about the privacy implications of speech recognition we need to realise that building speech recognition systems has always been dependent on large data sets for accuracy: the more data companies gather and process, the more accurate and useful their data systems become.

Furthermore we need to be aware of the fact that - as Dr Barassi (2018) argued in her report on Home Life Data and Children’s Privacy submitted to the Call for Evidence for Age Appropriate Design Code of the Information Commissioner Office in the UK - voice-operated AIs gather large quantities

of *highly contextual* data of children. 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." (Zuckenberg, 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).

With the inclusion of voice-operated AIs in different contexts of children's lives the result is that companies are gathering and processing not only large amounts of children's data but also large amounts of highly contextual data that can then be mapped to unique ID profiles, which can potentially follow children across a lifetime. At present, there is a lack of transparency in the industry for the ways in which the voice prints of children are collected and processed, or whether these voice prints are actually used to create unique ID Profiles.

We believe that unique ID profiles created on the basis of the data traces collected through speech recognition can have a profound impact on children's rights. One of the main problems of these forms of Unique ID profiling is represented by the fact that data technologies often sort, profile and inform action based on group rather than individual characteristics and behavior (ethnicity, class, family etc.) (Taylor, 2017). As Barassi argued in a recent article (Barassi, 2019) children have historically been profiled on the basis of the families and the social groups they belong to. Yet today these classifications might follow them throughout their life span in ways that was not possible before. This does not only reproduce existing inequalities and stall social mobility (Eubanks, 2018) but it also impacts on their right to self-definition and moral autonomy (Niessenbaum, 2010).

Adult-Centred and Unsafe Digital Environments.

According to a Global Market Insights Report (Buthani and Bharwaj, 2018) the Smart Speaker Market size was estimated to be over 4.5 billion USD in 2017, growing at CAGR of over 20% from 2018 to 2024. One of the most problematic aspects of the growth of smart speakers is the fact that at present, as noted in the Home Life Data report by Dr Barassi (2018), 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 technologies, because they are not designed or targeted specifically at children do not have to abide to regulations such as COPPA or the GDPR to protect children's privacy in the automated home.

Speech recognition technologies - that are not designed for children - but nevertheless are technologies with which children interact on a daily basis expose them to adult-centred, and possibly unsafe, digital environments threatening their best interests and right to optimum development. As mentioned in an article appeared on TheNextWeb by Dr. Patricia Scanlon, CEO of SoapBox Labs:

"The system must respond appropriately to children. White House press secretary Sarah Huckabee Sanders recently called out Amazon after she claimed that her two-year-old child was able to use the Echo to purchase an \$80 Batman toy. It's worth noting that simply calling out the word 'batman,' even repeatedly, is not going to purchase batman for you on Amazon; the UI simply doesn't work that way. However, it does highlight an interesting issue; that home devices should not allow children to access the same environments as adults.

Alexa is not alone in her indiscriminate behavior. My eight-year-old was recently using Siri on my phone and the word 'bitch' appeared. Note that she had not actually said the word, which speaks more to its poor accuracy for kids voices. Shocked, she showed me the phone. While Siri did reply with an amusing 'there's no need for that,' Siri should have identified that this was a child speaking and responded appropriately, choosing to not print the obscenity to the screen. For voice assistants to work accurately and appropriately, they must be able to identify

children’s voices and create, implement, and adhere to proper protocols in regards to how they handle requests.” (Scanlon, 2018)

We believe that the Committee should pay attention to the risks and challenges involved in children’s engagement with voice-recognition technologies that are NOT targeted and designed for children. This would be an important step towards the development of policies and strategies that stress the importance of ‘privacy by design’.

CONCLUDING REMARKS

Speech recognition technologies hold much promise and opportunities for children, they can advance literacy, widen participation and access to information; they can also foster, creative interaction, playfulness and exploration for children. Yet in the current digital environment these technologies pose a threat to children for two main reasons. On the one hand, companies collect large amounts of highly contextual data that can be then mapped to unique ID profiles without being transparent of practices and process. On the other hand, in the current digital environment, the speech-recognition technologies used by children are not designed for them and thus possibly expose them to unsafe adult-centred digital environments. All these different, yet interconnected issues, show that currently speech recognition technologies can have a potential impact not only on children’s right to privacy (UNCRC, Article 16), freedom of expression (UNCRC Article 13) and freedom of thought (UNCRC, Article 14) but also on their right to non-discrimination (UNCRC Article 2), best interest (UNCRC Article 3) and optimum development (UNCRC Article 6). We believe that the Committee should address the following problems of the digital environment:

- Lack of transparency in the ways in which voice data is used, and processed.
- Regulations replete with un and under defined terms like “data” and “immediately” and processes that involve, for example, deleting audio but retaining transcriptions and derived information.
- The lack of clarity in the ways consumers can have data deleted permanently and the inability currently to do this through an automated, or even, online process.
- The lack of power government watchdog groups have globally to track and enforce regulations.

We believe that SoapBox Labs sets an example and an important precedent in the industry. Yet we also believe that important actions need to be taken to make sure that speech recognition does not become a threat to children’s rights:

- That tech companies replace the convoluted legalese in their privacy policies with transparent lay man language that explains exactly what happens to voice data once it is collected by each company.
- That these companies agree to regular privacy audits by an independent organisation such as the [Campaign for a Commercial Free Childhood](#)
- Oblige companies collecting children’s voice data (including through technologies that are not designed or targeted at children) to respect COPPA and GDPR regulations, and to develop protections and technologies that guarantee that children are not exposed to adult-centered, and potentially unsafe environments.
- Oblige companies collecting children’s voice data for commercial purposes such as advertising to be upfront and transparent about their objectives and make opting out (without deleting an account) an upfront, straightforward and definitive one step process.

SoapBox Labs Media Contact: Niamh@SoapBoxLabs.com

Media Enquiries can also be directed to: y.barassi@gold.ac.uk

References:

- Adams, M. (2006). The promise of automatic speech recognition for fostering literacy growth in children and adults. In M. McKenna, L. Labbo, R. Kieffer, R., & D. Reinking (Eds.), *Handbook of literacy and technology, Vol. 2* (pp. 109-128). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Adams, M. (2011). *Technology for Developing Children's Language and Literacy: Bringing Speech Recognition to the Classroom*. New York, NY: Joan Ganz Cooney Center at Sesame Workshop.
- Barassi, V. (2018). *Home Life Data and Children's Privacy* [Call for Evidence Submission Information Commissioner's Office]. London UK: Goldsmiths University of London.
- Barassi, V. (2019, March 11). Coerced Digital Participation. Retrieved May 11, 2019, from Child Data Citizen website: <http://childdatacitizen.com/coerced-digital-participation/>
- Buthani, A., & Bhardwaj. (2018). *Smart Speaker Market Statistics - Industry Trends, Share Report 2024*. Retrieved from Global Market Insights website: <https://www.gminsights.com/industry-analysis/smart-speaker-market>
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- Kleinberg, S. (2018). 5 ways voice assistance is reshaping consumer behavior. Retrieved July 30, 2018, from Think with Google website: <https://www.thinkwithgoogle.com/consumer-insights/voice-assistance-consumer-experience/>
- Livingstone, S., Lnsdown, G., & Third, A. (2017). *The Case for a UNCRC General Comment on Children's Rights and Digital Media* [Children's Commissioner].
- 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>
- Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
- Scanlon, P. (2018, March 3). Voice assistants aren't built for kids — we need to fix that. Retrieved May 11, 2019, from The Next Web website: <https://thenextweb.com/contributors/2018/03/03/voice-assistants-arent-built-kids-need-fix/>
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), 2053951717736335. <https://doi.org/10.1177/2053951717736335>