Before the
Federal Trade Commission
Washington, DC

In the Matter of	)	
	)	
COPPA Rule Review	)	WC Docket No. FTC-2019-0054
	)	
16 CFR part 312, Project No. P195404	)	

Comments by Angelina Wang

My name is Angelina Wang, and I am a graduate researcher at the Department of Computer Science at Princeton University. I am responding to the FTC's request for comments on its implementation of the Children's Online Privacy Protection Act ("COPPA"). Specifically, I am addressing question 24 regarding the use of audio files containing a child's voice. I will respond with the following claims:

1. The Commission should *not* amend the Rule to specifically include an exception to obtaining parental consent before collecting a child's audio file if it is a replacement for written words
2. De-identification of audio files is not effective at preventing re-identification
3. Federated learning is a better alternative for an operator to improve their products rather than de-identification

The Commission should *not* amend the Rule to specifically include an exception to obtaining parental consent before collecting a child's audio file if it is a replacement for written words

There is currently an enforcement policy statement issued by the Commission in 2017 that claims no enforcement action will be taken against an operator that does not obtain parental consent when they collect an audio file form a child, as long as it's solely as a replacement for written words. It specifies that the audio file can only be held for a brief amount of time, during which it may be used for nothing else, and then must be deleted. Although initially this sounds reasonable, I believe that this particular situation is not a valid exception to obtaining parental consent. This is because even though in theory the child's audio file will not be compromised, there is still a security risk with any such transmission of data, and the parent deserves to know if

their child's audio file is being sent to a server somewhere else off the device. The fact that the audio file is deleted after a vaguely-defined "brief" amount of time is not a good enough justification for not obtaining parental consent.

However, if an operator does not want to go through the process of obtaining parental consent for a task like speech-to-text voice recognition, I propose a safer alternative that circumvents the need to transmit the audio file in the first place, and makes obsolete the need to inform the parent of anything. There are significant advances in mobile machine learning that will allow for fully trained machine learning models to be deployed on-device rather than on the cloud. What this means is that an audio file will never even have to be transmitted into the possession of the operator in order to be translated into text; the entire process can be done securely on the device.[1] Google's new Pixel does exactly such a voice recognition process on the device, and eliminates the security risk of needing to transmit an audio file elsewhere to be translated into text.[2]

De-identification of audio files is not effective at preventing re-identification

The public request for comments specifically asks if de-identification of audio files would be effective at preventing re-identification, and the answer to this is a resounding no. There are countless examples like in the domains of mobility traces[3], credit card metadata[4], and

---

[1] Yanzhang He, et al. (2018) "Streaming End-to-end Speech Recognition For Mobile Devices." *International Conference on Acoustics, Speech and Signal Processing (CASSP) 2019: 6381-6385*
https://arxiv.org/abs/1811.06621
[2] Devin Coldewey (2019) "Google's real-time speech recognition AI can run offline on Pixel." engadget
https://www.engadget.com/2019/03/12/google-releases-offline-gboard-speech-recognizer-for-pixel/
[3] Dániel Kondor, et al. (2018) "Towards matching user mobility traces in large-scale datasets." *IEEE Transactions on Big Data 2018*
https://arxiv.org/abs/1709.05772
[4] Yves-Alexandre, et al. (2015) "Unique in the shopping mall: On the reidentifiability of credit card metadata." *Science 30 Jan 2015: Vol. 347, Issue 6221, pp.536-539*
https://science.sciencemag.org/content/347/6221/536

large sparse datasets such as movie ratings[5] where de-identification is constantly being shown to be faulty. In fact, it has even been said that de-identification in general does not work[6]. Because of the extremely high fallibility of de-identification mechanisms, and constant likelihood that it can be broken through new techniques of re-identification, I urge that this risk not be taken, especially on such sensitive data as audio files of children. There are very little benefits to be gained from de-identifying audio files, and if the purpose is to improve products, I propose a better solution in the next section that makes this need obsolete.

Federated learning as a better alternative for an operator to improve their products rather than de-identification

One of the primary reasons an operator would want to de-identify an audio file, which as discussed above is a near-impossible and bad idea, is to improve their product. I propose federated learning as a way to improve their products without ever conducting de-identification and creating a security risk. Federated learning[7] is a machine learning training technique by which a user's private data used for training never actually leaves the user's own device. In this method, a child's audio file would never need to leave the device it is on (much like the technique I discuss in the first section), but could still be used to improve the product's model. It would perform the training on device then send back the parameter updates for the model to the operator's centralized server that keeps track of the speech recognition model. By employing this

[5] Arvind Narayanan, et al. (2008) "Robust de-anonymization of large sparse datasets." *IEEE Security and Privacy (Oakland) 2008*
https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf
[6] Arvind Narayanan, et al. (2014) "No silver bullet: de-identification still doesn't work."
https://randomwalker.info/publications/no-silver-bullet-de-identification.pdf
[7] Jakub Konečny, et al. (2017) "Federated learning: strategies for improving communication efficiency."
https://arxiv.org/abs/1610.05492

technique, the need for de-identifying an audio file in order to improve a product is eliminated, along with a large security risk.

Summary

To put it all together, I propose that operators who might want to collect audio files containing a child's voice for the sole purpose as a replacement for written words should look to recent advances in on-device techniques in order to eliminate privacy and security risks. No exception to parental consent should exist for this situation, because if an operator opts not to deploy secure on-device methods for children's audio files, parents deserve to know where their child's data is going, just like in any other situation.