Thesis Project Portfolio

Designing Secure and Usable Wake Words

(Technical Report)

Regulating the Gold Rush: Using Analogies to Legislate Data Privacy in Smart Speakers

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Joshua Sahaya Arul

Spring, 2021

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Improving and Ensuring Data Privacy in Smart Speakers

One of the first things I learned as a computer scientist was how to develop apps for Amazon Alexa. Since then, smart speakers have continued to grow and are trending towards becoming a household staple. Unfortunately, the number of privacy and security concerns about smart speakers and data privacy have grown in parallel. I chose the topics for both of my projects in hopes that people, including myself, could feel comfortable using smart speakers in the future, confident that their data is protected. In our technical project, my team and I developed a method for smart speaker companies to pick wake words that are less likely to misactivate and record audio of unwitting users. In my STS research paper, I demonstrate how analogies can be a practical strategy for US legislators and citizens to regulate smart speakers effectively.

My team and I wanted to tackle the problem of smart speaker misactivations by developing alternative wake words. A wake word is the word or set of words uttered by a user to address the smart speaker (e.g. “Alexa”, “OK Google”, “Hey Siri”). A smart speaker continually listens for a user say its wake word, and once detected, it begins recording. The issue is that there are many phrases that are mistakenly detected as wake words (i.e., misactivation), such as “OK cool” for “OK Google,” thus leading to smart speakers activating and recording private conversations. Our team hypothesized that the most phonetically dissimilar words would reduce misactivations. Iterating through the English dictionary, we calculated and summed each word’s phonetic dissimilarity to snippets of a large corpus. We then selected the most dissimilar words that were still user-friendly to say and remember. We found that our process yielded wake words (e.g. “okay jaguar,” “okay liberace”) that had drastically fewer misactivations, especially compared to the popular wake words in use today. We hope that companies can use our process to allow users to select wake words that will reduce misactivations, without having to compromise convenience.

Although this feature would improve smart speaker data privacy, it may not be sufficient to increase consumers’ trust of smart speaker companies. In my STS research paper, I use Schwarz-

Plaschg's method of analogies to compare smart speakers to smartphones, a technology that received similar backlash, and the European Union's General Data Protection Regulation (GDPR), the first modern attempt to regulate data privacy. Using these analogies revealed how anonymizing data is insufficient, that court rulings can establish meaningful precedent, how fragmented government regulation can undermine itself, and the need for the US government to act upon smart speaker regulation now. Based upon these findings, I encourage US policymakers to create their own analogies example? to understand smart speakers and write effective regulation so that consumers can be confident that their data is in good hands.

One lesson I learned from both projects is that the work is almost never done. In the beginning of the year, I was worried that my research projects would not produce promising results. I was listening to a podcast by JJ Redick, an NBA veteran player I began listening to over the COVID pandemic, and he said something that resonated with me: "There was never any sense that I was done accomplishing things. You've never arrived. You're always becoming." There is rarely a flag, prize, message of congratulations, or anything that says "Hey, you've made it!" Often, we think that certain achievements are the "arriving" moment, such as making it to the NBA, graduating college, or in this case, obtaining significant results. The truth that I found during the research process was that you never arrive. As time progresses and sociotechnical systems change, there is always more to consider and more to improve upon. My STS research paper only scratches the surface of relevant analogies, and my technical project could have incorporated additional rankings based on human trials. I believe this why the "Future Work" section is a signature of most good research papers. To future engineering students, I suggest that you not worry about "arriving" at what you believe is the ideal conclusion, but instead to focus on and trust the process.