An Investigation into the Privacy Concerns of Digital Virtual Assistants

A research Paper Submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia – Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

By

Benjamin Barrett

Fall 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: date:
Benjamin Barrett

Approved: date: 11/30/20
Richard Jacques, Department of Engineering and Society

Introduction

In 2011, Apple released a new type of technology in the I-Phone: Siri (Clifford, 2017). Siri was the first of many "Digital Voice Assistants" (DVAs) that would perform tasks based on verbal or text user input. Most major technology companies followed Apple in creating voice assistants for their platforms. Amazon has Alexa, Microsoft has Cortana and Google has the Google Assistant. These voice activated assistants allow users to enter an almost science fiction like reality where the computer will do what you tell it to. In 2019, 50% of internet consumers used Digital Voice Assistants in some way. This number is up 8% from 2018 (Stoddart, 2019) and is perhaps not surprising because of these assistants featuring on phones and computers that people use every day. The next revolution after Siri, was instigated by Amazon in 2014 when it launched the Amazon Echo. This device was mostly a speaker when it launched, and its major functionality revolved around playing music based on some spoken user input (Etherington, 2014). The Echo was a revolutionary product because it took Digital Voice Assistants out of traditional computing hardware (phones and laptops) and sold it essentially standalone. This device paved the way for other stand-alone personal assistant devices and has also facilitated an explosion in smart home devices. In 2019, 24% of consumers reported owning a standalone DVA (up 9% from 2018) and 14% reported wanting to buy one in the next year (Stoddart, 2019). However, these devices, while popular, are not without their perceived issues. One issue, privacy, has plagued companies for years after their release. Standalone DVAs in particular have incredibly high stakes when it comes to privacy. These devices, if they fully become "domesticated", have the potential to eavesdrop on every conversation we have. The development of these standalone devices has thus been filled with some controversy and developers have had to react to the outcry of privacy conscious users as well as the actions of the government. This paper will attempt to review the development of these standalone DVAs through an STS lens.

The development of the Amazon Echo

When an engineer creates a product, usually the product is developed with a certain type of user in mind. In his paper *Configuring the User*, Woolgar describes how engineers create technology using their own perception of what the ideal user should be (Woolgar, 1991). So the question is: what was Amazon's "ideal user" for the Echo? The answer to this question might be brought to light when considered inside the context of Amazon at the time. The Amazon Echo followed the release of Amazon's Fire phone, which failed to penetrate into the smartphone market. It was initially just thought of as a speaker, albeit one that can play music based on the user's voice. It was soon realized that the Echo could become the hub of the so-called "smart home", controlling various Internet of Things (IoT) devices around the house like smart locks and lights. Additionally, the user could order things through Amazon with the Echo, a functionality with an obvious goal (Brustein, 2016). It is possible to infer that the customer in Amazon's mind at this time is someone that loves living inside of an interconnected network of components that make life easier. They are willing to talk to this device and say which music they enjoy, or what things they want to order from Amazon.

A brief interview by one of the employees who worked at Amazon during the development of the echo is enlightening. As the Echo was being developed, the main problems that employees were worried about were the response latency and the issue of the Echo not being able to hear if the user was playing music as well (Brustein, 2016). They were additionally working under increased pressure because of the Amazon Fire phone failure. It almost seems like

Amazon was not even considering what a typical user would think about having this device in their home. Notably absent from these discussions were questions of user privacy.

One of the potential issues that might have hindered Amazon from thinking about privacy in this way is technical. From an implementation standpoint, it is much easier to treat every Amazon Echo not as an individual decision maker, but more as a communicator. Rather than have every individual unit do its own natural language processing and then response generation, it seems much easier to have each individual connect to a central processor that can then simply relay a response. Privacy is much easier to implement if the data never leaves the home and never touches Amazon's servers, but it just wasn't practical. This model of computation may have been different than what users expected and may have been what caused some of the outraged responses in future years. In hindsight, it might have been helpful to consider the ramifications of decisions like storing user data in the cloud that would make Amazon the target of scathing articles and lawsuits.

Privacy concerns brought about by the development of DVAs

The early release of Amazon's Echo was relatively quiet, in fact, it wasn't until 2016 that the Echo had its first run in with privacy litigation. This would be through a court case that would draw national attention. Arkansas native James Bates was accused of murdering his friend one night in late 2015. The prosecution did not have quite enough evidence to convict him, and present in the home was an Amazon Echo. The Echo was believed to have been activated at some point before or after the incident. Amazon initially refused the warrant to release the information, but relented when Bates agreed to have the recordings released. The case was dismissed in 2017 and Bates was not convicted (Chavez, 2017). This first brought to the public eye the fact that Amazon was storing all of this data in the cloud and that it could be potentially accessed by strangers.

The next major story that got a lot of attention broke in 2019. This began with a Bloomberg report that Amazon employs "thousands of people around the world to help improve the Alexa digital assistant" (Day et. al, 2019). These employees will listen to an audio recording from the Alexa device and then provide some labeling data to help improve the algorithm's performance. This data also isn't anonymized. The report states that transcribers can access first names, device serial numbers, and account numbers. It was also revealed that other companies are doing similar things. Google was reported doing something similar, contracting employees to analyze recorded data that could have sensitive information within (Clauser, 2019). These reports sparked backlash against Amazon and Google. People began to realize that these companies are taking their data and essentially hoarding it. As a result, Google and Amazon needed to change their processes.

Policies relating to DVA user privacy

As the technology is still relatively new, there have not been many legislative pieces that relate specifically to DVAs. However, it is still worth discussing, as a solution to this privacy problem may well need to be enacted in the halls of government. The first piece of legislation to consider is the EU General Data Protection Regulation (GDPR). This legislation was released in 2016 and covered a wide range of requirements that companies must apply for European consumers. These include mandating explicit consent for data collected, allowing users to request their data from the company, and allowing users to delete their data if they see fit. The consequences of not complying would be steep fines (Dwoskin, 2018). This sweeping legislation was enormous in its scope. Many web-based companies had to change their privacy policies and procedures in anticipation of the law's adoption. Companies like Spotify and Apple created ways for users to download their data. Google announced a rewrite of its privacy policy, making it easier to understand, hopefully eliminating legalese. This legislation is a good start, but does not give users total control about what is happening to their data. Indeed, some companies, like Facebook, have not given users the ability to opt out of all of the data that is being collected (Dwoskin, 2018).

While the GDPR covers the EU, there is no national policy or legislation that protects user privacy. The closest thing to the GDPR in the US is the California Consumer Privacy Act (CCPA). While technically only covering California, its existence nevertheless requires internet companies to change how they approach data privacy generally. This act is extremely young, as it went into effect Jan. 1 2020. The act requires companies to tell users what data they collect and delete your data when asked (with a few exceptions). The act also allows for lawsuits in the case of data breaches (Myrow, 2019). This act forms the major policy that internet companies like Amazon and Google need to consider when doing business in the US.

Changes related to privacy concerns

The world is very different from what it looked like in 2014. The first few years after the Amazon Echo was released, not a lot changed. However, in the past 3-4 years, many companies have had to change their privacy policies and provide more public information to run their businesses and products. Many companies have had to change their policies, but Amazon is an excellent example of this. Two separate privacy policies exist on Amazon's website. The main page is the current privacy policy (Amazon, 2020), and there is also a prior version of the privacy policy, dated 2017 (Amazon, 2017). While being significantly longer, the 2020 privacy policy features a number of subsections that did not exist in 2017. Notable are "For What

Purposes Does Amazon Use Your Personal Information" and "California Consumer Privacy Act". The latter section is of course aimed at the new legislation put into place at the beginning of 2020. The former section however, reveals something much deeper. In 2017, Amazon didn't think it explicitly needed to tell consumers the purposes behind using personal information. In 2014, during the development of the Echo, Amazon didn't think twice about using personal data generated by the platform. However, by 2020, it realized that people care a lot about their privacy, especially when it comes to devices that live directly in the home.

In addition to just privacy policy changes brought about by laws, major tech companies have had to implement different privacy-focused abilities into their existing standalone DVAs. The revelations in 2019 about contractors listening in to recorded commands caused Amazon and Google to implement ways for users to delete their data, daily if needed (Clauser, 2019). Countless privacy settings have been revamped in response to the public outcry. Other recent allegations surround Amazon's Echo Dot for kids. In 2019 it was alleged that Amazon "fails to provide sufficient information about what personal data its device collects from users ages 13 and under" (Shah, 2019). The results of this complaint are not immediately evident, but it is virtually guaranteed that Amazon and similar companies will be facing complaints like these for years to come on their current course.

Conclusion

When it developed the Amazon Echo, the developers weren't thinking about how consumers would react to the product from a data privacy perspective. In 2020, nearly 6 years later, it is obvious that user privacy should have been taken more seriously. As the public became more aware of the full extent of data use circulating through these devices, this slowly became apparent. In response to this spread of concern, the GDPR and the CCPA were passed to help hold companies accountable for their data privacy practices. The many changes to privacy policies and changes to the product itself, be it privacy settings or privacy guarantees, show how users can affect fundamental change in the products they use. This saga also illustrates major takeaways that companies can use while inventing the next generation of IoT devices, devices that will become interlaced with existing standalone DVAs. The most important of these takeaways is to consider the modern internet consumer's sense of privacy when designing a device that will become a permanent fixture of their home. This by no means will make the problem go away, but will put companies better positioned to deal with privacy challenges. It also shows the importance of privacy acts like the GDPR, these acts provide users with an ability to control their data and their privacy, which is to everyone's benefit.

References

Clifford, C. (2017, July 29). Here's how Siri made it to your iPhone. Retrieved October 24, 2020, from www.cnbc.com/2017/06/29/how-siri-got-on-the-iphone.html

Stoddart & Johnson (2019, August 19). Everyone has a voice: Voice Assistants on the rise. Retrieved October 24, 2020, from www.accenture.com/gb-en/insights/software-platforms/digital-voice-assistants-rise

Woolgar, Steve. "Configuring the User: The Case of Usability Trials. In A Sociology of Monsters" Essays on Power, Technology, and Domination (1991)

Etherington, D. (2014, November 6). Amazon Echo Is A $199 Connected Speaker Packing An Always-On Siri-Style Assistant. Retrieved October 25, 2020, from https://techcrunch.com/2014/11/06/amazon-echo/

Brustein, J. (2016, April 19). The Real Story of How Amazon Built the Echo. Retrieved October 25, 2020, from https://www.bloomberg.com/features/2016-amazon-echo/

Chavez, N. (2017, December 2). Arkansas judge drops murder charge in Amazon Echo case. Retrieved October 25, 2020, from https://www.cnn.com/2017/11/30/us/amazon-echo-arkansas-murder-case-dismissed/index.html

Day et al. (2019, April 10). Amazon Workers Are Listening to What You Tell Alexa. Retrieved October 25, 2020, from https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio

Clauser, G. (2019, August 8). Amazon’s Alexa Never Stops Listening to You. Should You Worry? Retrieved October 25, 2020, from https://www.nytimes.com/wirecutter/blog/amazons-alexa-never-stops-listening-to-you/

Dwoskin, E. (2018, May 24). New privacy rules could spell the end of legalese — or create a lot more fine print. Retrieved October 25, 2020, from https://www.washingtonpost.com/news/the-switch/wp/2018/05/24/new-privacy-rules-could-spell-the-end-of-legalese-or-create-a-lot-more-fine-print/

Myrow, R. (2019, December 30). California Rings In The New Year With A New Data Privacy Law. Retrieved October 25, 2020, from https://www.npr.org/2019/12/30/791190150/california-rings-in-the-new-year-with-a-new-data-privacy-law#:~:text=Images%2FIkon%20Images-,California's%20new%20digital%20privacy%20law%20takes%20effect%20on,1.&text=1%2C%202020%2C%20all%20Californians%20will,nationally%20recognized%20data%20privacy%20expert.

Amazon. (2020, January 1). Amazon.com Privacy Notice. Retrieved October 24, 2020, from https://www.amazon.com/gp/help/customer/display.html?ie=UTF8&nodeId=468496&ref_=footer_privacy

Amazon. (2017, August 29). Changes to the privacy notice prior version. Retrieved October 24, 2020, from https://www.amazon.com/gp/help/customer/display.html?ie=UTF8&nodeId=16015091

Shah, S. (2019, May 9). Amazon's smart speaker for kids violates GDPR, cybersecurity experts say. Retrieved October 25, 2020, from

https://www.spglobal.com/marketintelligence/en/news-insights/trending/EzKaYX3R6W7KdJWQz0gP7w2