Design Assignment 2 -- Ethical Analysis:

Massoud, Yahya 300147944
Zhang, Lingfeng 300134245
Zhou, Yiqiang 300129168

Design Topic: Amazon Echo Smart Speaker

Design Problem: How much personal data Amazon should collect about its customers?

Design Description:

Amazon Echo, a brand of smart speakers developed by Amazon, is an voice-based smart device that connects its users to the AI personal assistant service Alexa. Echo will listen to you and respond when you say one of the wake-words, "Alexa" or “Hey Alexa”. Echo has built-in features such as: music playback, setting timers, playing podcasts and audiobooks, voice interaction, etc. Besides, third-party companies can add their customized features, which are called “Alexa Skills” to Echo using API and developers’ guides provided by Amazon. The "Alexa Skills" enable Amazon Echo to have further capabilities like "adding events to your Google Calendar and even ordering pizza[1]." Consider the fact that the Echo can be placed in any room/space in a home. Features can be added into the Echo to ensure it is recording/sharing data appropriately.

Key Value Tension: (Customer) Parents [Privacy] vs Amazon [Data Collection]

Motivation:

Although most of the users buy Amazon Echo devices to facilitate and accomplish their daily tasks, people should have enough privacy when they are at home. Especially when Echo devices are placed in certain rooms that people tend to value the privacy in those rooms most, such as bedrooms or bathrooms. When a person is having a private conversation with their partner in their bed, these conversations are supposed to be private and should not be known by any other people. And obviously, people would like to have full privacy when it comes to the bathroom as well. However, Amazon wants to collect data from its customers whenever it is possible. On one hand, Amazon Echo by its nature, has to "listen to" the people in the room all the time in order to be triggered when the "magic word" appears [2]. On the other hand, Amazon needs to collect the data in order to improve the quality of their services, the overall user experience, and to advertise more efficiently [3]. Often is the case that after the device is triggered, recorded data will be uploaded to Amazon's cloud server in order to provide useful feedback to the users. And the data will also be used to "improve your [users] experience and our [Amazon] services” [3].

When looking through value tensions, we found that Amazon Echo devices treat all rooms equally and are always "listening to" the trigger/wake words no matter where they are placed, and will upload the data after being triggered. This might become problematic, especially when Echo devices are placed in some "sensitive" areas or places in someone’s home.

Ethical Issue: Privacy

Since Echo devices are supposed to respond timely, their microphones are active all the time and their processors are analyzing what the microphones receive. That alone is a little uncomfortable when people are having private conversations. That process of "identifying the trigger word" is mostly done locally so it is less a problem to most of the people. However, Echo devices can be mistriggered and sometimes even if users intentionally trigger the devices, unwanted voices might accidentally appear. In those cases, the recorded voices will be uploaded to Amazon's cloud servers, which could lead to a bigger privacy issue.

Each and every person has to have complete privacy when it comes to their homes. Assuming no modifications to the current way of how Alexa operates, private conversations are prone to be recorded without explicit user permission [5] and stored on the cloud without being completely anonymized, as each conversation could be traced back to the persons involved in it using the device’s serial number [4].

Amazon needs more customer speech data to do further natural language processing analysis due to improving the quality and stability of Alexa. Sometimes, customers' speech include some personal information which can be used to identify a certain person, customer privacy may be violated without anonymity processing. Also, knowing that Amazon Echo could listen without being invoked and without notifying its user [5], would make the people worry more about their private conversations to be exposed to Amazon.

Design Decisions:

The following are potential design decisions that could be made and we will discuss the value outcomes of each of these decisions:

1. User tells each Echo device what kind of room it is being placed:

Users can tell Alexa what kind of room they are setting and Alexa can analyze different room contexts and handle corresponding tasks, some unrelated speech, like personal conversations should not be sent to Amazon cloud for neither processing nor storage.

2. User set different "privacy level" for each Echo device:

Users can set "privacy level" according to various situations, either manually using physical buttons or voice-based instruction. If the privacy level is high, Amazon should not store these personal information. How Amazon deals with the collected data depends on the customer's preset privacy level. For each level, there might be some features that the user will not be able to use.

3. Amazon develops technology that detach the data from each identified user:

This mechanism can be applied by differential privacy technology. Differential privacy is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals in the dataset. For example, differentially private algorithms are used by some government agencies to publish demographic information or other statistical aggregates while ensuring confidentiality of survey responses, and by companies to collect information about user behavior while controlling what is visible even to internal analysts. Roughly, an algorithm is differentially private if an observer seeing its output cannot tell if a particular individual's information was used in the computation [6].

4. Develop advanced architecture of Alexa which can process speech words on-device (edge-computing), or using third-party services:

The problem of Alexa is this smart device has to submit the customer's speech to Amazon cloud service, which will store customer's data in the cloud server, process the query, and respond back to the customer. If speech recognition and natural language processing models are lightweight, this system architecture can be applied in the Amazon Echo product directly, meaning all the preceding steps can be done on-device. This strategy can protect customer's private information completely. If possible, customers can donate their own Alexa data to Amazon for advanced improvement and research. In another way, some remote processing services like Solace Publish/Scribe service [10] can be applied in Alexa. This service can process inputs in the remote end without storing these inputs information, and the processing speed is quicker than traditional cloud service. This technology can protect customers' privacy successfully.

5. Constraining Alexa to work only under confirmed user consent:

Section 184 of the Criminal Code in the Canadian law prohibits the wilful interception of private communications but provides an exception if one of the participants consents to that interception [7]. One other important thing is to enforce Amazon to let the AI-enabled voice assistant, Alexa, to declare itself whenever it is actively listening, while eliminating false alarms during the day[8] and being activated only either by hardware button being pressed or by using the wake-words, then having Alexa re-confirm that the user really wants to request something from it. This procedure may affect the user experience and make it less smooth, but the most important concern here is the privacy of the user. These suggestions should give the users a feeling that their conversations within their private space are actually private, which is a right to every human being to have, especially at home.

6. Only store information which Alexa can respond to:

Sometimes customers' words are meaningless for Alexa and Alexa does not have a confident response for the customer, most of these "meaningless" words tend to be private information. These information which ALexa cannot respond to should be removed from the Amazon cloud and these "meaningless" information sometimes is helpless for the improvement of Alexa. This design decision can reduce the possibility of storing customer's privacy in Amazon cloud.

References:

[1] CRAIG LLOYD. (2018, February 7.) [The Best Third-Party Alexa Skills on the Amazon Echo] (https://www.howtogeek.com/256707/the-best-third-party-alexa-skills-on-the-amazon-echo/)

[2] Amazon.com. (2019) [Alexa and Alexa Device FAQs] (https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230)

[3] Ilker Koksal. (2018, December 11.) [How Alexa Is Changing The Future Of Advertising] (https://www.forbes.com/sites/ilkerkoksal/2018/12/11/how-alexa-is-changing-the-future-of-advertising/#fa004311d4dc)

[4] Valinsky, Jordan. (2019, April 11th). Amazon reportedly employs thousands of people to listen to your Alexa conversations. Retrieved from https://www.cnn.com/

[5] McCue, TJ. (2019, April 19th). Alexa is listening all the time: Here’s how to stop it. Retrieved from https://www.forbes.com/

[6] Dwork C., McSherry F., Nissim K., Smith A. (2006) Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi S., Rabin T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg

[7] FindLaw. Can I record private conversations? Retrieved from https://criminal.findlaw.ca/

[8] Heater, Brian. (2017, May 15th). Amazon is helping to remove false wake words from Alexa’s vocabulary. Retrieved from https://techcrunch.com/

[9] Su, Jeb. (2019, May 16th). Why Amazon Alexa is always listening to your conversations: Analysis. Retrieved from https://www.forbes.com/

[10] Solace Publish/Scribe. https://solace.com/samples/solace-samples-jms/publish-subscribe/