Amazon Alexa and Its Challenges to Reach More Households

Pankaj Pathak, Rati Shukla, Himani Jain, Vikash Yadav, Parashu Ram Pal, and Rishabh

Abstract Voice assistants became an important feature in the current smart device world. Taking instructions from human and providing services to them has increased the data traffic in the IOT systems due to which telecom players are getting benefitted. However, these devices are connected to the network all the time and without proper security measures will cause security breaches in the voice assistant systems like Alexa, Siri, Google Now, Cortana, etc. Alexa Amazon's smart speakers enabled with AI rapidly is adopted by households and inculcated in various daily life activities. Though the technology is innovative but with its dark side and its challenges of security vulnerability. The case study highlights major cybersecurity attacks on the voice assistants and the challenges which Alexa and other voice assistant devices are facing for their market expansion. This study enables to critique the security vulnerability in Amazon Alexa AI voice assistant and other such products. It helps to have profound understanding of key market and industry-based challenges in the current markets.

P. Pathak
Symbiosis Institute of Digital and Telecom Management Symbiosis International (Deemed University), Pune, India
e-mail: pankajpathak@sidtm.edu.in

R. Shukla
MNNIT Prayagraj, Allahabad, Uttar Pradesh, India

H. Jain
ABES Engineering College, Ghaziabad, Uttar Pradesh, India
e-mail: himani.jain@abes.ac.in

V. Yadav (✉)
Department of Technical Education, Lucknow, Uttar Pradesh, India

P. R. Pal
SAGE University, Bhopal, Madhya Pradesh, India

Rishabh
Galgotias College of Engineering and Technology, Greater Noida, India

X.-Z. Gao et al. (eds.), *Advances in Computational Intelligence and Communication Technology*, Lecture Notes in Networks and Systems 399,
https://doi.org/10.1007/978-981-16-9756-2_1

Keywords Alexa · Voice assistant · False authentication · Dolphin attack · AI · Amazon

1 Introduction

Amazon Alexa is a well-known digital voice assistant, enabled with AI and used as smart speakers launched in 2014 by Amazon. Various enhanced versions came into market till now and reached to large households for various tasks. Alexa can be engaged in our daily life routine works, viz. controlling the home appliances, getting infotainment, scheduling our daily tasks, shopping, and many more. Alexa can play an effective role in home automation and to control various smart devices and merely by using voice. The Alexa system consists of Echo products identified as—the hardware and software that directly communicate with consumers and the cloud modules—that have most of the "smarts": automatic speech recognition, understanding natural language, and response. The working architecture of Alexa has been shown in Fig. 1. Third-party services through applications can also provide some responses via "skills." The third parties who write and publish those skills are responsible for the behavior of their skill. To demonstrate the working of Alexa, we are presenting an example to obtain weather information by simply requesting Alexa by speaking "Alexa, what is the weather." We can see how request flows which are caught by Echo and how it is interpreted, acted, and finally responded by Alexa. Amazon Alexa works on natural language processing which records the words and interprets the sound by using computational power. It then identifies words with matching in the database, makes sense of the tasks, and carries out corresponding function.

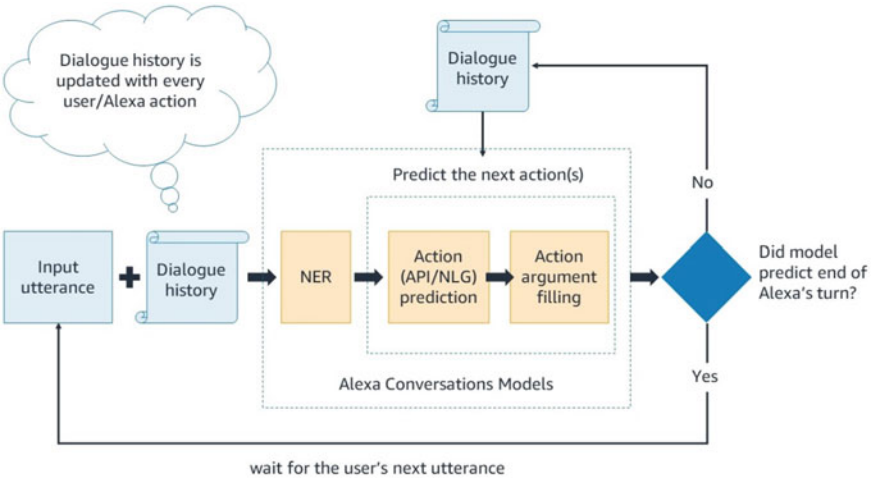


Fig. 1 Alexa working architecture [14]

Amazon Alexa was launched in 2014, and it is available for general public in 2015. Alexa's API availability to developers helps in expanding its market due to integrating it with non-Amazon products. This feature is amazing and exciting other brands to inculcate Alexa in their products. But at the same time, it also questions the quality of these third-party products. It may impact on the brand image of Alexa. But experts say that unless and until there are some products with good quality available there is no adverse impact on Alexa's brand image.

Why is Amazon Alexa so important? Alexa is successful in setting up a benchmark up to some extent for inaction machine learning with artificial intelligence as an effective consumer-based implementations.

Systems such as Alexa are an early sign of a significant shift of communication between humans and computers.

Who has affected by Amazon Alexa? Customers are primarily affected by Amazon Alexa who are using it in terms of positive as well as negative aspects. If we say positive, they ease their life by exploiting the services of it. If we see the dark side of it, then definitely security concerns and data ownership are the primary anxieties of it. Alexa also affects its rivals like Microsoft, Google, and Apple; all have digital assistants pushing Amazon to play catch-up.

After discovering that Amazon stores your audio recordings for Amazon employees to review, Alexa users grew even more alarmed. All this has made people a little bit cautious of using Alexa, and it is facing a major challenge of reaching different households. These privacy concerns are keeping people hesitating from using Amazon Alexa.

2 Evolution of Alexa

With the enhancement of digitalization, the market is seen to be demanding a more convenient and on the touch lifestyle. Amazon Alexa is one such digitally revolutionized AI-based product introduced by Amazon. As per January 2019, over 100 million Alexa devices have been sold [1].

Initially, Alexa was introduced with few utilities in the market but with the time these utilities are increasing. Alexa can perform various daily life tasks such as setting alarm, informing about weather, preparing list, scheduling task, accessing informative articles, etc. One of the interesting skills of Alexa is to listen live sporting events, National Geographic quiz, big sky as well as news stations [2]. Alexa also has built in support for Spotify and Pandora accounts.

As far as concern to technological advancements, in April 2019, Alexa-enabled devices could support the download of over 90,000 functions which was merely 1000 functions in 2016 [3]. The home automation feature was launched in April 8, 2015 [4]. Developers are able to create their own smart home skills using the Alexa Skills Kit. Alexa also emerged into the field of food delivery. Starbucks used Alexa for

placing pick up orders, but very few Alexa owners used these services for placing the orders [5, 6].

According to research conducted by Consumer Intelligence Research Partners (CIRP), in 2018, 69% of the US smart speaker market is captured by Amazon Alexa. And it is increased by 1% in 2019. As far as concern to virtual market, Alexa and the other competitors like Siri, Google Assistant, and Cortana had their market shares—25%, 36%, and 17%, respectively [7].

However, we see due to poor digital knowledge in India as well as a few other countries, Amazon Alexa is finding it difficult to reach out to the majority of the crowd. As of November 2018, Alexa is available in 41 countries. Also, Amazon Alexa is seen to have been facing strong competitions over the years which is one of the major hurdles faced by Amazon Alexa.

3 Challenges to Retain Household and New Market Expansion

3.1 Investigation of Security Vulnerabilities in Voice Assistants like Alexa

The reason for selecting these HDVA devices like Amazon Alexa is manifold. Amazon Alexa is flagship device in HDVA devices. There are 5 million devices being sold since its launch in 2014 within 2 years. Secondly, there are 10,000 skills (Alexa voice services) available for Amazon Alexa. Moreover, every smart device is equipped with voice assistant system, right from smart speaker to connected cars. Hence, exploring the security aspects of these devices will be helpful in finding out problems and advancement of these HDVA devices. Currently, Amazon Alexa takes voice commands from humans and performs some action. The actions include playing music, shopping online, checking weather, scheduling appointments, making payments, and controlling smart devices like garage doors, plugs, and thermostats. Basically, these HDVA devices have a feature like (always listening mechanism) which helps users to talk to the devices continuously so that users need not hold some button and pass commands but this will expose them to security vulnerability. In past there were several incidents came across which breached the security agreements.

3.2 Design of Voice Support Systems

Before understanding the security breaches of Alexa, let us understand the voice service model of Alexa. To control a smart device, a user can speak voice command to Alexa after waking up with a voice called “Alexa.” Alexa then sends a command to the remote cloud for authentication purpose using Wi-Fi. Once the command

gets authenticated, then the command is forwarded to the server called smart home skill adapter, which is maintained by Alexa in cooperation with third party. Later, the command is sent to another cloud which is used to control the corresponding devices.

3.3 Security Vulnerabilities in Alexa—False Authentication

The voice service facility in Alexa enables it to listen to users' commands [8]. However, it can respond and act to the commands during the absence of humans which is a security breach. When Alexa was designed, its design considerations were limited within the room and only home members can access it, but there can be some exceptions like the sounds accepted by Alexa can come from outside surroundings also which can lead to the compromising of security. Since there is no second factor authentication in these devices, it will lead to trust issues. After careful study, we observed that these kinds of issues are occurring when victims are not at home. If HDVA devices stop taking commands when users are not in the surroundings, then the fraudulent commands can be restricted. Alexa works on the scenario that when a user stands nearby, it takes commands from him/her and provides services. Now, we will exploit this feature by sending commands to Alexa using a Bluetooth speaker. We have used a Bluetooth speaker controlled by a smartphone at significant distance (within 8 m) from the Alexa. Interestingly, Alexa responded to such commands from the speaker.

Validation: Even though when a false command is passed by an unknown person to Alexa, it can be overcome by the Alexa authentication system, but the authentication system can be easily bypassed by generating commands using text to sound (TTS). There are many Web sites present that can be used to create mp3 sound of a speech; thereby, any device (mobile, Bluetooth, laptop) can be used to play audio and get access to Alexa.

3.4 Dolphin Attack Through Inaudible Voice Commands

Till now, we have seen cases where the HDVA devices can be activated without the human presence. In this regard, another attack has been explored on HDVA devices called Dolphin attack. Dolphin attack is basically injecting a sequence of inaudible voice commands (modulating human voice on ultrasonic carriers) that can lead to unnoticed security breaches to the voice-controlled systems. This is a technique to exploit ultrasonic sound channel (ultrasonic channel is the frequency channel which is greater than 20 kHz). Human's voice system is designed to listen voices up to 20 kHz. Mobile phones adopt audio sampling techniques lower than 44 kHz in which they deploy low-pass filter so that any signal greater than 20 kHz will

not be passed through it. Even though ultrasonic sounds are received and correctly sampled, they cannot pass through the sound recognition system of HDVA as they do not belong to the human tone. Dolphin attack overcomes all these issues and performs attacks. Dolphin attack may be imposed on Google, Siri, Cortana, and Alexa through sequence of inaudible voice commands. But all the hurdles like accessing malicious Web sites, spying, getting fake information, denial of service, concealing the attacks, listening the inaudible voice (>20 kHz), greater intelligence of inaudible sounds than speech recognition system, etc., can be overcome by dolphin attack and activate audio hardware of the device.

Any voice assistants will be activated using 2 commands: 1. activation commands like "Alexa," "hey Siri," etc., and 2. recognition commands like "Call 123456," etc.; Alexa uses speaker-independent algorithm that means it accepts "Alexa" word spoken by any individual as long as the word is clear. Whereas Siri gets trained only by a human, viz., it uses speaker-dependent algorithm. The speaker-dependent algorithm works on a local server, whereas the speaker-independent algorithm works on cloud server.

3.4.1 Attack Design

Dolphin attack is carried in 3 steps

1. Voice command generation
2. Voice command modulation
3. Voice command transmission

Voice command generation: The voice command generation is basically generating the activation words. Creating activation commands is different unless the user speaks "Siri." Here, the experiment will generate activation commands by two methods: 1. Attacker cannot find owner (stolen cases). 2. Attacker can obtain few recordings of owner. When attacker cannot find owner, TTS-based brute force technique is used which means we get recordings of human voices from TTS Web sites and then use them for activating. There are different TTS sources that provide human voices like Selvy Speech, Baidu, Neo speech, acapella, etc. Also the case when attacker obtained few recordings of victim. In this case we will use concatenative synthesis technique to extract the activation words. Let us say that the attacker got the recordings of victim like "City and Carry." So, sampling can be done to extract the word Siri like CIty + ca RRY = Siri. So, like this we can get activation commands in the first step.

Voice command modulation: Here, the malicious commands can be modulated over suitable carriers, while modulating select carriers of ultrasonic range so that they are inaudible. Then, amplitude modulation of baseband signal and ultrasonic signal has been obtained. The amplitude modulation parameters are like 1. modulation dept., 2. carrier frequency, 3. voice selection. Firstly, the modulation depth is set to 0.5 here which means the carrier amplitude varies 50% above and below its unmodulated level. Secondly, the carrier frequency is depending on many factors like frequency range

of ultrasounds, the cutoff frequency of low-pass filter, and the frequency response of the microphone. The lowest modulated frequency has to be 20 kHz to ensure inaudibility.

To impose successful attack for base signal of 6 kHz carrier signals, frequency must be of 26 kHz to ensure the lowest frequency is larger than 20 kHz. Also, different voices vary in terms of baseband frequency ranges. Let us say female voices have wideband than male voices due to which there is a probability of frequency leakage. So, it is preferred to select narrowband voices for the attack.

Voice command transmission: The voice command transmission consists of signal source, modulator, and speaker. Powerful transmitter is used with signal generator and portable transmitter with a smartphone. The first one is we are used to validate the dolphin attack and the second one is for walk-by purpose. The powerful transmitter uses all powerful equipment to generate, modulate, and send. The portable transmitter uses a smartphone to transmit the modulated signals. In the smartphones, we have low-pass filter that attenuates the higher frequencies in order to overcome these problems; narrowband ultrasonic transducers are used as speakers and amplifier to amplify them. In order to find how much the TTS-generated voice differs from a voice played by a smartphone, a method called Mel Cepstral dispersion may be used. It is observed that the MCD factor is less than 8 and hence is preferred for the attack.

Recognition versus activation: Various devices exhibit different results in terms of attack distance considered. In addition, if we append the controlling command to the recognition command, the chance of success is higher due to the fact that the activation commands are trained by speech recognition systems that are always on mode. The length of command matters when it comes to the activation of voice assistants. For example, "Call 12345/open abc.com" is harder to recognize than "Turn airplane mode" or "What's new today." In first case the word "open" and "call", has to correctly recognize for execution and in the second case only "turn" and "What" has to be recognized for execution. So, the attack can be successful if the attack word is short and commonly recognized by speech recognition system.

4 Implementation Issues

With the rise of voice-controlled AI-led voice assistants in the ICT ecosystem, there are still serious issues regarding the inhibitions about the same. Organizations are also looking into the business aspect of this technology and are considering an AI-led voice assistant ecosystem into their enterprise management to increase the efficiency and reduce the turnaround time (TAT) of the system. We will now look at some of the challenges [9, 10] for Amazon Alexa and its rivals which pose to business organizations for implementing the same.

- **Against Popular Opinion**

Fig. 2 Consumer Intelligence series voice assistant survey, 2018 [13]

It is a well-established fact among the general workforce that whenever anything automated is introduced into the system, there is always a fear of loss of jobs among the masses. Voice assistants are no exception. There is widespread fear that the introduction of AI will eventually cut our jobs.

- **Knowing the Real Areas of Action**

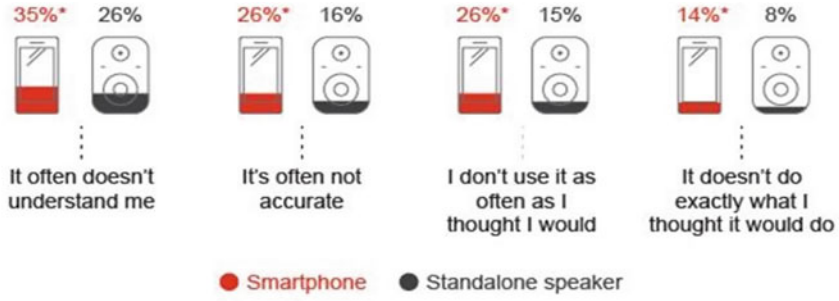
Even after your organization has finalized to go with an AI-led voice assistant (VA), the real challenge becomes to choose which vertical to implement and direct your R&D department to work upon and increase your productivity and efficiency. The verticals vary from HR, finance, technology, etc. Since there is very little information available about any known use cases of AI-led AI, it becomes a challenge for organizations to come up with one. A consumer survey was conducted in 2018 to check the usefulness of voice assistant in human's life as shown in Fig. 2.

- **On boarding and Integration of your Voice Assistant into the System**

Just like with any integration, your voice assistant (VA) will also have to be onboarded and integrated into the system which is not just loading them with information or giving them access to all your resources. To be really efficient, your AI-led VA must know your entire flow-of-work as well as access to your databases and resources. This is a continuous and gradual task and does not happen overnight. Moreover, the real challenge [10] is to determine how much access you are going to give your VA so that you are not putting your organization at a business risk. Figure 3 shows a survey conducted among several consumers in 2018 to check the satisfactory rate of voice assistant device.

- **Handling Data**

The major unique selling proposition of an AI-led voice assistant is to process and analyze huge chunks of data and store as well as to make sense out of this data for later use. The main challenge is what to do with this huge database of data and how your



Q. Here are some reasons people claimed to be unsatisfied with voice assistants. Please select which one(s) apply to you. *Statistically significant at a 90% confidence level
Source: PwC Consumer Intelligence Series voice assistants survey, 2018

Fig. 3 Consumer Intelligence series voice assistant survey, 2018 [13]

AI engine handles this data without being a security risk. There is no standardized data handling framework existing as of now and are puzzled organizations to use voice assistants [8].

Amazon Alexa has practically 59% of the piece of pie in India followed by Google Home to 39%. Yet, it is very much deprived of arriving at each house in India. As indicated by the International Data Corporation (IDC)'s India Monthly Smart Speaker Device Tracker, a sum of 753,000 units was dispatched in 2018 in India. This number is very less if we consider the total Indian population. The consumer segment added to most of the keen speaker deals in the nation. In any case, the acquisition of brilliant speakers by ventures likewise saw a flood during the celebration or festive time, basically for gifting purposes, while the real endeavor reception for vertical explicit use cases stays at an incipient stage. With regard to the favored channel for savvy speakers, online channels comprehensive of e-trailers and merchant's own sites commanded brilliant speaker deals with 55% of the shrewd speakers sold in this channel. Current Web-based business infiltration in India stands at 28%, so we can clearly estimate the room for improvement.

At a public interview in New Delhi, the Web-based business mammoth said Alexa now coming with Hindi, a language verbally expressed by generally a large portion of a billion people in India, as the organization hopes to grow its compass in the country. To bring support for Hindi in Alexa, the company was working on it over a year. Users can now be able to ask Alexa their inquiries in Hindi, and the digital assistant will have the option to react in a similar language. The component, which will start turning out through a product update to Alexa gadgets beginning today, as of now just backings one voice type for Hindi. (For English, Alexa offers numerous voice types.) In the months to come, Amazon said it intends to include support for multilingual families, which will empower individuals from the family to associate with Alexa in the language they each like [9].

In view of its most current review, the firm said: "the US introduced base of smart speaker gadgets is 76 million units, up from 70 million units in the March 2019 quarter and 50 million units in the June 2018 quarter." Other studies contend there are in excess of 100 million units in American homes [11]. So, from these stats we can easily estimate that America being the less populated country is getting a huge number of Alexa shipped than India almost 10 times more than India.

5 Evaluation and Conclusion

5.1 Defense Mechanisms

The primary solution is concerns about today's microphones' sensing capability which can sense acoustic sounds, i.e., >20 kHz. Thus, a microphone has to be designed in such a way that it should suppress frequencies that are of ultrasonic range. This is called microphone enhancement.

Secondly, we can add a module before low-pass filter (LPF) to detect modulated baseband signal and demodulate them to obtain baseband signal. This process does not spoil the normal operation of microphone since there is no relation between captured voice signals and noises of ultrasonic range. This is called inaudible voice command cancelation technique. The voice recognition should not be passed through a random voice command.

Thirdly, we can use software-based defense mechanisms. This feature uniquely identifies modulated voice commands to the genuine ones. In this, we take a demodulated voice and genuine voice which both indicate a difference of 500–1000 Hz that means if we detect a signal in the range of 500–1000 Hz then we can detect the dolphin attack. To be specific, a machine learning-based model shall classify the voices. This method is called support vector-based technique which can be used as a defense mechanism against dolphin attacks.

Support vector machine (SVM) is a supervised machine learning algorithm mostly used in classification issues in which each element of data is plotted as a point in n-dimensional space (where n is the number of features you have) with the value of each function being the value of a particular coordinate.

5.2 Discussion and Analysis

In [12], authors discussed dolphin attack and they also presented hardware and software defense solutions. Our paper presented and discussed the challenges for Alexa which might be hurdle to get expansion in the market for voice assistant products. The study also discusses the impact of different attacks and the solutions presented for them. The research study enables to critique the security vulnerability in Amazon

Alexa AI voice assistant and other similar product which helps to have a profound understanding of key market and industry-based challenges in the current markets.

6 Summary and Recommendations

Voice assistants are becoming an important feature in the current smart device world. Taking instructions from human and providing services to them has increased the data traffic in the IOT systems due to which telecom players are getting benefitted due to heavy traffic. However, these devices are connected to the network all the time and without proper security measures will cause security breaches in the voice assistant systems like Alexa, Siri, Google Now, Cortana, etc.; hence in this case study, we have highlighted major cybersecurity attacks on the voice assistants. First one is false authentication in which the Alexa system is activated without sending human voice commands rather than sending the same through a recorded voice which indicates that a voice assistant system can be activated without the presence of humans. Secondly, we have discussed dolphin attack in which the ultrasonic frequency range is exploited by VA which is inaudible to humans. The modulation process of the ultrasonic sound on a high frequency carrier wave is discussed which can be used as an attack channel on victims' voice assistant system. We have also discussed how baseband signal can be generated in the form of several TTS systems and synthesis techniques, thereby generating activation commands from them. Several methods have been discussed, viz., how to prevent such attacks by modification of a voice assistant system like support vector machine algorithm which is a software-based firewall that can avoid dolphin attack and other physical changes of the system. We did analysis of voice assistant markets in the current position and the challenges like natural language processing, unable to process commands at every time, AI-based challenges, knowing the real areas of action, and handling data which became latest market challenges now.

Many other competitors in the market are into this same segment of the voice assistant. As discussed above in the case study, only 753,000 units of Alexa sold in the year 2018 which is very small concerning the total population of India. Here, the biggest challenge comes that how to increase this number in the coming years. Alexa should look at how to modify itself with the growing technology, and only voice assistant service is not enough to stand out in front of the competitors. Alexa can be integrated to control the IOT-enabled devices with the help of voice assistants run by mobile data provided by telecom operator, and it can increase the mobility of Alexa. This feature will make Alexa a more modified device which the Indian consumer can find worth buying.

References

1. Bohn D (2019) Amazon Says 100 million Alexa devices have been sold—what's Next? The Verge, 4 January 2019
2. Erickson S (2018) "Alexa, Make Me Money": conversational AI prepares for the enterprise, 22 September 2018
3. Press release (2019) Amazon.com announces first quarter sales up 17% to $59.7 Billion, 25 April 2019
4. Callaham J (2015) Amazon Echo owners can now control WeMo and Philips Hue devices with their voice, 8 April 2015
5. Kell J (2017) Starbucks adds voice ordering to iPhone, Amazon Alexa, 31 January 2017
6. Anand P (2018) The reality behind voice shopping hype, 6 August 2018
7. Olson C, Kemery K (2019) Voice report: consumer adoption of voice technology and digital assistants. Technical Report. Microsoft, 2019
8. Yadav V, Mishra D (2020) Home automation system using raspberry pi zero W. Int J Adv Intell Paradigms 16(2):216–226. https://doi.org/10.1504/IJAIP.2018.10017087
9. Healy P (2018) 9 challenges in implementing an AI virtual assistant into your organization, 30 April 2018
10. Chapman K (2018) The challenge of AI voice assistants in customer service, 7 June 2018
11. Sterling G (2019) Alexa devices maintain 70% market share in U.S. according to survey, October 2019
12. Zhang G, Yan C, Ji X, Zhang T, Zhang T, Xu W (2017) Dolphinattack: inaudible voice commands. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security, pp 103–117
13. PricewaterhouseCoopers (2018) Consumer intelligence series: prepare for the voice revolution. Retrieved from PricewaterhouseCoopers (PwC) Website: https://www.pwc.com/us/en/advisory-services/publications/consumerintelligence-series/pwc-voice-assistants.pdf
14. https://www.amazon.science/blog/science-innovations-power-alexa-conversations-dialogue-management