



Research Article

Personal Voice Assistant: from Inception to Everyday Application

Ilman Shazhaev ^{1,*}, Arbi Tularov ², Dmitry Mikhaylov ³, Abdulla Shafeeg ⁴, Islam Shazhaev ⁵

¹ Farcana, Dubai, United Arab Emirates, ilman@farcana.com

² Farcana, Dubai, United Arab Emirates, arbi@farcana.com

³ National University of Singapore, Singapore, dr.d.mikhaylov@nus.edu.sg

⁴ Farcana, Dubai, United Arab Emirates, abdulla.shafeeg@farcana.com

⁵ Farcana, Dubai, United Arab Emirates, -

Correspondence should be addressed to Ilman Shazhaev; ilman@farcana.com

Received 10 June 2023; Accepted 18 June 2023; Published 31 July 2023

© Authors 2023. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).

License: <https://creativecommons.org/licenses/by-nc/4.0/> — Published by Indonesian Journal of Data and Science.

Abstract:

The creation of automatic speech recognition systems is a popular trend in the development of information technology. These technologies are developing at a very fast pace, gradually covering more and more areas: already now we can say how firmly they have settled in our lives. The term "speech technologies" means a fairly large layer of information technologies, but one of the most advanced products in this area is a voice assistant, which includes the use of all types of speech technologies: speech recognition, speech synthesis, a system for developing and analyzing voice information, and voice biometrics. A voice assistant is software that allows you to control your device using voice commands. The range of possibilities does not end with the execution of commands; a modern assistant is even able to maintain a conversation with the user. Since the voice assistant is a complex innovation that consists of many different technologies, the task of a smart assistant is to ensure that they work smoothly with each other. Now there are many voice assistants on the market that can make life easier for a person. Now you don't have to manually enter a question into a search engine or search for a song, you just need to say what you want, and the voice assistant will find everything on its own. But every year there are more and more assistants and it becomes more and more difficult for the average user to choose, because each assistant has its own characteristics. Nevertheless, the daily use of this technology is becoming more widespread. Another area that is rapidly developing is gaming. Using more and more innovations, the next logical step is to implement a voice assistant, at least at the training stage. However, real progress is out of the question for the time being.

Keywords: Technological Advancement; Digitization; Innovation; Voice Assistan.

Dasataset link:

1. Introduction

An important aspect of development is information technology, as it forms a new Internet economy, which is based on knowledge, and not on the expanding consumption of non-reproducible resources. The use of voice assistants is a hot topic in the modern world, since it simplifies a person's life as much as possible. Voice assistants are bots that run on artificial intelligence through voice recognition and natural language processing to answer questions, conduct conversations, and run simple tasks. With the advent of this option, it has become much more convenient to perform many queries. Voice assistants are built on artificial intelligence (AI), machine learning and voice recognition technologies.

In 1962, the future computer giant IBM developed its speech recognition technology called Shoebox [15]. In the 90s, breakthrough solutions appeared when a voice assistant could be used in practice, e.g., in processing telephone requests. The first voice assistant to communicate with the user appeared in the new millennium. Development of neural networks, increasing the power of smartphones and cloud technologies. These combined factors led to the fact that the voice assistant could find a permanent place next to us: in laptops, smartphones, and even smartwatches [15].

Dialogue with a voice assistant is becoming utterly ordinary in our time. We are no longer surprised when a tourist asks his smartphone for the address of the nearest cafe on the street [3]. We calmly watch how the driver sets the navigator's course simply by saying his car's address. Children play educational games with intelligent assistants, and older people are interested in what the weather will be like tomorrow [3]. The whole industry has rapidly reached the point that users do not want to type a long search phrase into the browser line. They want to conduct a dialogue in ordinary human language. And the voice assistant of 2021 is already ready for it [15].

Already, "voice assistants" and intelligent assistants are clearly showing us what personal robots will be capable of soon. First, the voice interface will be the primary interface for interacting with the many functions of the personal robot [10]. Secondly, the robot will be connected to various online services, such as a calendar, a news portal, an online store, etc. [10].

The emergence of intelligent personal software agents, such as Siri, Cortana, and Google Now, is mainly because most people are not used to downloading and using individual applications to solve their problems. More and more programs will support searching for their content through external services. You don't even have to install them to do this. A critical factor in the growth of the popularity of intelligent assistants is the ability to search through the content of applications. In turn, those will use their search algorithms to pull content from apps without the user having to install them on their phone.

Voice assistants action algorithm

Since recurrent networks have a "memory" of past states, storing long-term relationships in recognizable speech, that is, the context, has become possible. Now the neural network can predict the result based on whole sentences, expanding the learning possibilities and making it possible to use ordinary human speech as a training database [14]. Since neural networks are accurate, the only serious problem is improving the recording quality in aggressive, noisy environments. After processing the speech from the sound signal into a format convenient for the program, the voice assistant can proceed to the request processing algorithms [14]. More simple helpers, which almost any user with special software can configure, are based on pretty simple algorithms. The user collects key phrases and indicates the reaction (program actions) to them, after which, in the active phase, the Assistant checks the voice commands for compliance with the keys [4]. The probability with which the team satisfies each of the keys is revealed. If the probability is satisfactory and maximum among all the others for a specific team, then the key activates the action attached to it. More interesting are advanced helpers that run on multiple connected neural networks [4]. The output for each previous network is the input for the next one.

Starting from the speech recognition stage, they can also analyze it, determine the subject of the request, highlight key points and independently synthesize the response to the request. Robust systems that use a considerable number of technologies. Since the result of the work of the neural network is the prediction of the most probable result and not an unambiguous answer, this system is flexible enough to support possible retraining [14]. Therefore, such assistants are left with the learning function even in operation mode, so their behavior is correctly adjusted to users' needs [14]. Unfortunately, these agents do not implement the ideals of machine understanding. The capabilities of these resources are pretty limited by the scope in which they were trained to operate. The more abstract and larger the framework, the more problems they can solve—however, the more specific the framework, the better and more complete the solution. Therefore, modularity is necessary, in which requests are analyzed and solved by the fragments of the system responsible for a specific problem. This principle is typical for programming in general. Perhaps the only obstacle at this stage is only the "reasonableness" of the resource. For speech agents at this stage, only support from developers is essential - the more modules are developed, the more opportunities the resource will have [12]. Accordingly, at this stage, the only possible breakthrough will be a full-fledged artificial intelligence introduction [6]. At the same time, artificial intelligence as a human assistant is likely the final stage in the evolution of speech agents.

How to choose a voice assistants

Before choosing a voice assistant, you need to consider a lot of nuances. To begin with, you should familiarize yourself with the most popular systems, as well as study the advantages and disadvantages of each of them. Critical selection criteria include the following:

A. Compatibility

Each operating system has its virtual assistants. For example, Cortana is in high demand for Windows. However, it is worth noting that this Assistant is also suitable for mobile devices [7].

The Siri system has been developed for Apple gadgets - it has a wide range of functions and a high speed of information retrieval.

Voice control

Voice control of intelligent devices is carried out through personal assistants. Voice control is a necessary feature for a modern Smart Home in 2018. Moreover, thanks to technological advances in voice control and artificial intelligence, the voice assistant is now the control center of an automated home.

An intelligent home controlled via an application on the phone, without a voice assistant - this is the best level of 2011. And to turn the light on or off, you need to find the phone, open the manufacturer's application, find the appropriate device, and only then perform the desired action. With a voice assistant, you have to say, "Turn on the lights" [1].

Voice assistants are already built into phones, watches, tablets, computers, TV boxes, and smart speakers. In the next few years, intelligent assistants will permeate every device we use, including extensive home appliances.

The most popular device with a voice assistant today is a smartphone. The voice assistant is available in almost all modern phones: the iPhone has a built-in Siri, and the Google Assistant app on Android smartphones. You can activate the Assistant by pressing a button or by starting a command if the Assistant on your phone is in the "Always listen" mode. Thus, if one has an iPhone, you need to say "Hey Siri" and then say a command [1].

The primary device for managing the Smart Home in 2018 is smart speakers with a built-in voice assistant. This is a fundamentally new interaction between a person and an apartment. Each IT giant has already released its smart speaker or a whole line of speakers: Amazon has Amazon Echo and Amazon Echo Dot, Google has Google Home and Google Home Mini, Apple has HomePod [2].

Controlling a house through a column has several advantages:

- The speaker is always connected to power. It cannot be discharged.
- The speaker always listens to the owner. No need to carry a phone/tablet/watch with you around the house.
- Each family member does not need an expensive phone.
- The speaker can be used as a high-quality home audio system, especially for HomePod or Amazon Echo.

Managing a modern smart home is not only an application on your phone with icons of connected devices but also a full-fledged dialogue with your virtual "butler," from whom you can find out about the weather and traffic jams, ask for the latest news, ask to fulfill some request and receive from answering him in the most familiar form for a person - in a voice, in one's native language [8]. At the same time, the personal Assistant does not have a bad mood. He will always be happy to see you and always ready to help as soon as he says "Siri" or "Hey, Google."

Overview

How voice assistants work is as simple as using them [2]:

1. Passive sound reading by activating the function with a built-in code word;
2. Signal filtering - the stage of eliminating noise and interference that occurs when recording a voice request;
3. Digitization of sound - the audio signal is converted into a digital form understandable to the computer;
4. Signal analysis - sections with speech are highlighted, and parameters are evaluated, such as part of speech, word form, and connection in one request;
5. Search for template data - artificial intelligence collects different pronunciations of the word, compares it with templates, and gives the result. A voice assistant can be written using almost any programming language, but Python was the most popular in 2020 [12]. To write program code, one will need knowledge of the language, the ability to work with the connection of various libraries, and demonstrate a precise formulation of input and output data.

The activity of the subsystem originates from the input of a voice command. The received signal is converted into a digital form and passes the procedure of filtering external noise [6]. Further, the converted

signal enters the identification subsystem. In this subsystem, a query is first sent to the signal database to recognize a command. If the input and stored signals match, the recognition is booming, and the command is transmitted to the executive device that performs a particular action [6]. If the voice command is not recognized, the system returns to the beginning - entering the voice command. The algorithm of actions is repeated until a positive result is achieved, until the voice command is recognized.

The functionality of using voice assistants is based on solving simple daily tasks (building a route, calling a person from contact information, setting an alert, maintaining a conversation, and others). As a result of frequent use, the voice assistant remembers the most frequently used functions of the user of this device and tries to facilitate further requests [4]. Since all voice assistants have artificial intelligence, when communicating, they take into account changes in location, time of day, days of the week, search query history, various previous orders, and much more [2].

For those who do not want to use peripheral devices (mouse, keyboard), the most popular assistants in the review are:

1. "Cortana" - created by Microsoft for the Windows operating system. Platforms it can also run on: Android, Xbox One, Microsoft Phone, and Microsoft Band. An intelligent assistant plans the day, reminds you of important things, plots the route for the car, and keeps up the conversation. Cortana is connected to Windows 10, so it manages some applications: it helps with emails and checks the contact list [7]. If one enters data, it will put the music according to the user's preference. Cortana syncs with other devices.
2. OK, Google - combines the functions of a voice assistant and a search program. "OK, Google" searches for music, finds any addresses, goes to sites, and reminds of events. It works for free. It was built into the Google Chrome browser, available for Android.
3. Siri is a virtual assistant that works on Apple devices: iOS, iPads, and iPods touch. On gadgets from Apple, Siri is built-in by default. It only requires activation in the settings. Executes simple commands.
4. Bixby Samsung is a proprietary digital assistant for Samsung and Galaxy S8. "Bixby" is activated by a button on the case. Feature - using the camera recognizes objects and finds them on the Internet.
5. Jarvis - voice assistant for home control. "Jarvis" is activated offline by a code word. Recognizes and executes commands and reports on the work done.

Google Assistant

Google's voice assistant is installed by default on all Android smartphones with GMS support - as it was developed by the same company that produces the operating system. Usually, the Assistant is prompted to set up the first time you turn on the device, saying "OK Google" several times (Pal et al., 2019). And then, the Assistant is launched either by voice command or by long pressing the "Home" button.

It is logical that when working, he accesses Google services - he searches for videos on YouTube, builds a route using Google maps, and so on. Many household appliances also support the Assistant - for example, appliances from the Mi Home ecosystem.

If to talk about which is better, Google Assistant or Siri, then one would better start off with user preferences: if one uses search, navigator, and other Google services more often, then one should opt for the Google Assistant. Or if one is faithful to Gmail and Google Maps and listen to YouTube Music, it would be more logical to use the native Assistant. By the way, Google Assistant also works on iPhones; however, its application will need to be installed separately.

Amazon Alexa

This voice assistant first appeared in Amazon intelligent speakers. They are not sold in Russia. Moreover, Alexa's program is not available in our app stores. One can only install it on the smartphone if the Apple or Google account is set to a different country .

This assistant is predominantly an English only version and does not support most domestic localized services. However, one can use it if you say the commands in English [11]. With Alexa, one can play music from Apple Music and Spotify, create to-do lists, edit one's calendar and search for information on the Internet..

Interestingly, the Amazon assistant supports Xiaomi intelligent home devices to control light bulbs, vacuum cleaners, and other appliances from the Mi Home ecosystem. The Amazon Echo Dot 3rd Gen is one such example. This model is compact and, if necessary, will fit in a jacket pocket. For the sake of mobility, the manufacturer "cut down" the sound characteristics of the Dot 3rd Gen. For example, there is one stereo speaker. But there are already four microphones, so everything is in order with the perception of voice and commands at the speaker. The device is powered only by the mains, so you can't take it with you on trips or outdoors. This is a severe drawback since the Dot 3rd Gen, with such a compact size, is designed for outdoor use.

Microsoft Cortana

This is another well-known old-timer among voice assistants. Cortana is one of the best options for PC and laptops. Of course, it is best to synchronize the Assistant with Microsoft technology because this brand is the Assistant's developer [1]. It is also possible to install an analog on a smartphone with the Android operating system. In addition, the utility is compatible with the Xbox One, a popular gaming console [1]. It is also worth considering that the utility is based on the Bing browser. This is a good browser, but users rarely resort to its help. However, this is not essential since the voice assistant works with the search engine.

The Assistant responds not only to the voice but can also remember the owner's needs over time and fulfill them automatically. Cortana makes shopping lists, creates reminders, sets alarms, and recognizes voice commands perfectly. A popular device for home appliances is the Google Home device. It has a large number of positive characteristics. The small speaker supports Google Assistant. With this device, you can search for information. It will be the perfect addition to your smart home. With the help of the application, you can always be aware of all the latest news and change settings on time. If desired, one can set individual functions and use daily reminders.

Another popular device is the LG WK7Y with powerful bass speakers. This is an option for those who love music of good quality. LG WK7Y has two speakers (tweeter and bass) with a total power of 30 watts. The device's audio system was developed jointly with the well-known acoustics manufacturer Meridian. There is a separate driver and a unique Enhanced Bass technology to enhance the bass.

SIRI

The voice assistant gave rise to the development of speech recognition technology. In 2011, Apple integrated a voice assistant into the iPhone 4s. Then Siri easily interacted with 12 applications, could create calendar reminders, sound the weather, navigate and send emails.

After more than 11 years, it has become much more intelligent. Now the volume of her knowledge has increased by 20 times [2]. She can send voice messages, translate text into foreign languages, and much more.

Assistant from Apple, which is by default on iPhones and iPads. Its main advantage is that Siri works with Apple's system applications - you can answer calls and messages, create notes and reminders, set alarms, get directions, and so on [5]. Siri is compatible with Apple Watch and will even help you find your lost AirPods.

When comparing which voice assistant is better, Alice or Siri, you need to consider which services you use more often. If you do not use Apple services and work, for example, with Yandex Maps, it will be more convenient for you to use Alice, not Siri. Also, Siri only supports Apple Music, which requires a paid subscription [5]. Theoretically, you can create quick commands for the Assistant to work with applications - but this is done at the development level using the SiriKit toolkit.

Technology flaws

Even the most ingenious device has problems. The voice assistant device, like the Google Home mentioned above, is always listening to the owners, waiting for their command. And no one still knows if the Assistant keeps a record and where all the information is stored [9]. Manufacturers have repeatedly reported that nothing is being recorded, but sometimes unpleasant incidents happen. For example, several times, Echo recorded and

sent the usual conversations of its owners to random numbers [9]. It turned out that she perceived random phrases as the command to "record the conversation."

Also, access to the conversations of household members can theoretically be obtained by hacking the email or local network attached to the column. A smart speaker won't protect your secrets if you have something to hide (Porcheron et al., 2018). Although developers have not yet understood what exactly needs to be hidden: passwords are rarely spoken out loud. Even if they are, attackers must go through hundreds of hours of conversations to get to valuable details [13].

Everyday Application

Apart from the smart home systems application and the smartphone implementation, the voice assistant did not received widespread acclaim. Nonetheless, there have been attempts to implement this feature in other spheres of human life. Gaming is one such attempt. Although, calling it an attempt is a little far-fetched. Sony has filed a patent that describes an assistant for gamers based on artificial intelligence. This assistant can suggest how to defeat a specific in-game boss and where to find specific items. It's called PlayStation Assist. The Assistant should help the player in difficult situations. For example, if the character has little health, he will tell you where the first-aid kit is. The Assistant fully understands the context at the time of the request: it analyzes the player's actions, equipment, and location (in the game world). It is reported that a mobile application can also be created for PlayStation Assist, which will provide information about the date and time of the start of the passage of a particular game, the amount of time spent in it, etc. Of course, this is all just a patent and not the fact that Sony will release an authentic product.

However, the very idea is very promising. Of course, we all play games for the thrill of the feeling to have passed it, or levelling up with the help of our mind and wit. Yet, with the help of a voice assistant this would be much easier, allowing the developers to include this feature as an element in the complexity setting.

Conclusion

Based on the foregoing, we can conclude that although voice assistants are developing quite quickly, they are not at the peak of popularity and development of their capabilities in 2020. It is also worth noting the fact that in the future, using voice assistants, it is planned to conquer space and perform simple medical operations. It turns out that the capabilities of the voice assistant will spread from solving simple user requirements to solving more complex technically costly tasks, the implementation of which will require long-term training of artificial intelligence.

Acknowledgments

Funding: This work has not received any funding.

Conflicts of Interests: All authors declare that they have no conflicts of interest.

Authors contribution: All authors contributed to manuscript revision, read, and approved the submitted version.

References

- [1] [Berdasco, A., López, G., Diaz, I., Quesada, L., & Guerrero, L. A. \(2019\). User experience comparison of intelligent personal assistants: Alexa, Google Assistant, Siri and Cortana. *Multidisciplinary Digital Publishing Institute Proceedings*, 31\(1\), 51.](#)
- [2] [Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Valdez, A. C. \(2019, July\). "Hey, Siri", "Ok, Google", "Alexa". *Acceptance-Relevant Factors of Virtual Voice-Assistants*. In *2019 IEEE International Professional Communication Conference \(ProComm\)* \(pp. 101-111\). IEEE.](#)
- [3] [Devi, S., Merchant, Z. A. M., Siddiqui, M. S., & Lobo, M. \(2019\). Artificial intelligence based personal assistant. *Asian Journal For Convergence In Technology \(AJCT\)* ISSN-2350-1146.](#)
- [4] [Erol, B. A., Majumdar, A., Lwowski, J., Benavidez, P., Rad, P., & Jamshidi, M. \(2018\). Improved deep neural network object tracking system for applications in home robotics. In *Computational Intelligence for Pattern Recognition* \(pp. 369-395\). Springer, Cham.](#)
- [5] [Ewers, K., Baier, D., & Höhn, N. \(2020\). Siri, Do I Like You? Digital voice assistants and their acceptance by consumers. *SMR-Journal of Service Management Research*, 4\(1\), 52-68.](#)

- [6] [Faiz, Z., Srivastava, V., & Khoje, S. \(2022\). Virtual Voice Assistant for Smart Devices. ECS Transactions, 107\(1\), 4315.](#)
- [7] [Hong, G., Folcarelli, A., Less, J., Wang, C., Erbası, N., & Lin, S. \(2021\). Voice Assistants and Cancer Screening: A Comparison of Alexa, Siri, Google Assistant, and Cortana. The Annals of Family Medicine, 19\(5\), 447-449.](#)
- [8] [Kaye, J. J., Fischer, J., Hong, J., Bentley, F. R., Munteanu, C., Hiniker, A., & Ammari, T. \(2018, April\). Panel: voice assistants, UX design and research. In Extended Abstracts of the 2018 CHI Conference on human factors in computing systems \(pp. 1-5\).](#)
- [9] [Lei, X., Tu, G. H., Liu, A. X., Li, C. Y., & Xie, T. \(2018, May\). The insecurity of home digital voice assistants-vulnerabilities, attacks and countermeasures. In 2018 IEEE Conference on Communications and Network Security \(CNS\) \(pp. 1-9\). IEEE.](#)
- [10] [Mishra, A., Makula, P., Kumar, A., Karan, K., & Mittal, V. K. \(2015, May\). A voice-controlled personal assistant robot. In 2015 International Conference on Industrial Instrumentation and Control \(ICIC\) \(pp. 523-528\). IEEE.](#)
- [11] [Pal, D., Arpnikanondt, C., Funilkul, S., & Varadarajan, V. \(2019, July\). User experience with smart voice assistants: the accent perspective. In 2019 10th International Conference on Computing, Communication and Networking Technologies \(ICCCNT\) \(pp. 1-6\). IEEE.](#)
- [12] [Panarese, P., Meraglia, E., Vergallo, R., & Mainetti, L. \(2021, September\). Enhancing Voice Assistants: A Proactive Approach. In 2021 6th International Conference on Smart and Sustainable Technologies \(SpliTech\) \(pp. 1-4\). IEEE.](#)
- [13] [Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. \(2018, April\). Voice interfaces in everyday life. In proceedings of the 2018 CHI conference on human factors in computing systems \(pp. 1-12\).](#)
- [14] [Rabassa, V., Sabri, O., & Spaletta, C. \(2022\). Conversational commerce: Do biased choices offered by voice assistants' technology constrain its appropriation?. Technological Forecasting and Social Change, 174, 121292.](#)
- [15] [Soofastaei, A. \(2021\). Introductory Chapter: Virtual Assistants. Virtual Assistant, 1.](#)