Using Digital Assistants for Accessibility Developments

Steven Au

Dec. 7, 2022
CSE 185 Fall 22

General Audience

IEEE

Using Digital Assistants for Accessibility Developments

Steven Au

Abstract—Chatbots have existed for some time in the commercial space and have seen a surge in usage to due extended periods of social isolation from a global pandemic. As a branch of digital assistants, chatbots are becoming more sophisticated with advancements in machine learning that provide faster response times, cheaper computation power, and easy scalability. Chatbots generally serve as an all-purpose usage to direct the user across a platform. This paper will explore the usage of chatbots in healthcare, therapy, and disability learning.

Index Terms—Digital Assistant, chatbot, accessibility, healthcare, mental health, therapy, machine learning, artificial intelligence, natural language processing

1 Introduction

DIGITAL assistants are now incorporated into our everyday devices with Siri, Google Assistant, and Alexa. These general-purpose assistants allow us to navigate our respective devices better. Making calls to mom, navigating home, and adding items to a grocery list is one voice command away. The applications of digital assistants are branching out to more sophisticated tasks such as therapy chatbots and mental-health awareness agents.

The general question are using literature reviews from academic papers to understand the scope and extent of the question. The discussion question is looking at relevant academic papers to make an inference on how to better tailor digital assistants for accessibility in modern technology.

General Question 1: *What are the current research results in academia using machine learning techniques for mental health detection?*

Well-being practices and applications are known procedures to help improve mental health. Machine learning allows the system to better improve individuals' processes in recovery. Healthcare applications are applied to therapy [6] and anxiety and stress detection in IoT [7].

General Question 2: *What are the limitations of a predictive therapy chatbot?*

Therapy chatbots require an interdisciplinary analysis of computer science, human-computer interaction, and psychology to be implemented. What parts of computer science are unable to solve the problem of accessible therapy solutions?

Discussion Question 1: *What are the ways to adopt more accessible technology?*

Most digital assistants are commercial products to help promote brand image. Digital assistants can be better adopted for children, the elderly, disabilities, and chronic illnesses.

1.1 What are digital assistants?

Digital assistants are computer programs that use natural language processing and other technologies to understand and respond to user requests. They are often used in devices like smartphones and smart speakers, and they can assist with a variety of tasks, such as setting reminders, playing music, and providing information. The term ”digital assistant” is a broad umbrella that covers a variety of different technologies and applications, and it is often used interchangeably with other terms like virtual assistants, intelligent assistants, and voice assistants. They are also defined as predictive chatbots by their most common usage. Digital assistants weren't always "smart." When technology was still developing and handheld devices did not have internet access, digital cameras, tape recorders, or pagers could be considered digital assistants. However, that term will be outdated now even though it's an electronic device that helps perform a specific task. The term has been coined to be "smart" which is able to use artificial intelligence particularly machine learning to understand voice commands.

AI, or artificial intelligence, is a broad term that refers to the ability of a machine or computer program to perform tasks that typically require human intelligence, such as learning, problem-solving, and decision-making. AI can be classified into two main types: narrow AI, which is designed to perform a specific task, and general AI, which is capable of learning and adapting to new situations. On the other hand, a digital assistant is a specific type of AI that is designed to assist humans with various tasks. Digital assistants are often voice-activated, and they can perform a wide range of tasks, such as scheduling appointments, providing information, and managing various devices.In summary, AI is a broad term that refers to the ability of a machine or computer program to exhibit intelligent behavior, while a digital assistant is a specific type of AI that is designed to assist humans with various tasks.

Digital assistants, such as smart speakers, can be a valuable tool for managing daily tasks and activities. For example, instead of scheduling a meeting by hand, you can give a smart speaker the command to do it for you. This can save time and effort, and it can also improve the accuracy and consistency of your schedule. But, not all digital assistants are equally ”smart” or capable. Some digital assistants are limited in their abilities and only provide pre-determined outputs or have a singular use. For example, a digital assistant that is only capable of setting reminders or playing music is not as ”smart” as one that can transcribe audio into text and schedule meetings based on your availability.

In order to be considered truly ”smart,” a digital assistant should be able to go beyond simple commands and offer personalized recommendations and suggestions based on your past behavior and preferences. This type of digital assistant would be able to adapt and improve over time, becoming more useful and effective as it learns from its interactions with you.The term ”smart” has evolved to refer to digital assistants that are capable of more than just providing pre-determined outputs or performing singular tasks. In order to be truly ”smart,” a digital assistant should be able to learn and adapt to the user’s needs and preferences.

1.2 Types of Digital Assistant

General-purpose digital assistants are meant to solve a wide range of usage, usually, these allow a user to navigate a certain platform through their voice. Alexa, a smart speaker, allows users to navigate Amazon services, and similarly, Siri, Google Assistant, and Cortana help navigate their respective devices. Looking at Fig 1., the complexity of digital assistants is becoming more human-like as technology is developing.

The simplest digital assistant setup is a command and control system that allows a user to control a device with simple voice commands. The user is limited to a few commands to control a device and is less of a dialogue. Simple assistants are often used in environments where hands-free control improves efficiency, for example giving machine operators additional voice control on the factory floor.

In a step up from command and control systems, many of today’s assistants are task-oriented. The user and computer work together to achieve well-defined tasks like making a bank transfer or finding a mortgage recommendation. These assistants typically work in narrow domains like finance or customer service and require some dialogue back and forth with the user to complete the task.

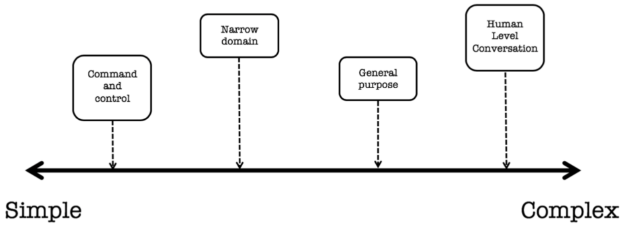

Fig. 1. Digital Assistants Complexity Line

Source: From [3]

1.3 History

The concept of artificial intelligence, or AI, has been around for a long time. The term was first coined in 1956 by John McCarthy, who is often referred to as the father of AI. He defined AI as ”the science and engineering of making intelligent machines,” and the field of AI research seeks to develop algorithms and technologies that enable machines to perform tasks that typically require human-like intelligence, such as visual perception, natural language processing, and problem-solving.

Alan Turing was a pioneering mathematician and computer scientist who also played a key role in the development of AI. In 1950, Turing published a paper called ”Computing Machinery and Intelligence” in which he proposed a test, now known as the Turing Test, to determine whether a machine is capable of intelligent behavior. The Turing Test is still used today as a benchmark for AI research [2]. Additionally, Turing’s work on the concept of a ”universal computing machine” laid the foundation for the development of the modern computer, which is an essential tool for AI research and development.

ELIZA is one of the earliest examples of natural language processing in a computer program. It was developed by Joseph Weizenbaum in the 1960s at the Massachusetts Institute of Technology (MIT). ELIZA was designed to simulate conversation with a human user by using a series of pre-programmed responses to certain keywords and phrases [1]. For example, if a user typed in the phrase ”I am feeling sad,” ELIZA might respond with ”Why do you feel sad?” ELIZA was not truly intelligent, as it did not have the ability to learn or adapt to new situations. Instead, it relied on a set of pre-defined rules and responses to simulate a conversation. Despite this, ELIZA was able to convincingly imitate human conversation and was seen as a significant step forward in the field of AI and natural language processing.

It wasn’t until 2011, Apple released its first commercial digital assistant, Siri, followed by Alexa in 2014, and Google Assistant in 2016.

1.4 Applications

Digital assistants are increasingly being used in a variety of settings and for a wide range of tasks. Some of the most common ways that digital assistants are being used today include:

- In personal devices, such as smartphones and smart speakers, where they can be used to manage schedules, set reminders, and answer questions.
- In homes, where they can be used to control smart home devices, such as lights, thermostats, and security systems.
- In offices and other work environments, where they can be used to schedule meetings, manage calendars and answer questions.
- In customer service, where they can be used to provide information and answer frequently asked questions.

2 Conversation AI Pipeline

In the last few years, deep learning has improved the state of the art in conversational AI and offered superhuman accuracy on certain tasks. Deep learning has also reduced the need for deep knowledge of linguistics and rule-based techniques for building language services, which has led to widespread adoption across industries like telecommunications, unified communications as a service (UCaaS), retail, healthcare, and finance.

When you present an application with a question, the audio waveform is converted to text during the automatic speech recognition (ASR) stage. The question is then interpreted, and the device generates a smart response during the natural language processing (NLP) stage. Finally, the text is converted into speech signals to generate audio for the user during the text-to-speech (TTS) stage. Several deep learning models are connected to a pipeline to build a conversational AI application.

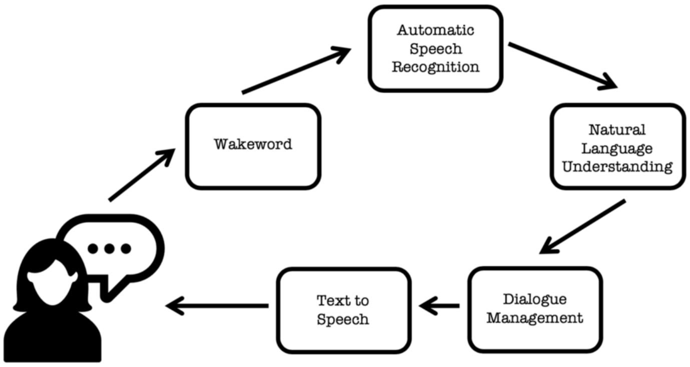

Fig. 2. Generic Digital Assistants Pipeine

Source: Altered from [3]

- A Wakeword (WW) detector runs on the device, listening for the user to say a particular word or phrase to activate the assistant. It's also possible to activate the assistant in other ways, like a push-to-talk button.
- Automatic Speech Recognition (ASR) converts spoken audio from the user into a text transcription.
- Natural Language Understanding (NLU) takes the transcription of what the user said and predicts their intention in a way that's actionable. This component understands that users can make the same request in a multitude of different ways that should all have the same outcome.
- The Dialogue Manager (DM) decides what to say back to the user, whether to take any action and handles any conversation.
- Text to Speech (TTS) is the output voice of the assistant.

A typical conversation AI pipeline process involves several steps for converting speech into text and back again. When a user speaks, the system first listens for a wakeword, which is a specific word or phrase that indicates the user wants to interact with the digital assistant. Once the wakeword is detected, the system uses an automatic speech recognition (ASR) system to convert the user's speech into text.

The resulting text is then input into a natural language understanding (NLU) system, which is responsible for interpreting the meaning of the user's words. The NLU system uses machine learning algorithms to analyze the text and understand the intent behind the user's request.

Once the NLU system has understood the user's intent, it passes this information to a dialogue management system, which is responsible for deciding how to respond to the user's request. The dialogue management system uses a set of rules and algorithms to determine the appropriate response, and it may also access additional information or data sources as needed.

Finally, the response generated by the dialogue management system is passed back to the ASR system, which converts the text into speech and outputs it to the user. This completes the conversation AI pipeline process, allowing the digital assistant to have a natural, human-like conversation with the user.

2.1 Automatic Speech Recognition

Traditional speech recognition algorithms take a generative approach, which involves modeling the full process of how speech sounds are produced in order to evaluate a speech sample. This process typically begins with a language model that captures the most likely and frequently occurring orderings of words. This language model is typically generated using an n-gram model [5], which is a statistical model that predicts the next word in a sequence based on the previous n words. The next step is to generate a pronunciation model for each word in the language model. This pronunciation model maps each word to the corresponding sequence of phonemes, which are the individual sounds that make up a word.

The final step is to translate the pronunciation model into an acoustic model, which maps the sequence of phonemes to the corresponding audio waveform. This acoustic model captures the way that speech sounds are produced, including factors such as pitch, volume, and timbre. Once these models have been generated, the goal of speech recognition is to take a spoken input and determine the most likely sequence of text that would result in that input according to the generative pipeline of models. This involves evaluating the input against the language, pronunciation, and acoustic models, and selecting the sequence of text that is most likely to have produced the given audio.

With advances in the capabilities of neural networks, each component of the traditional speech recognition model can be replaced by a neural model that has better performance and greater potential for generalization. For example, an n-gram model can be replaced by a neural language model, a pronunciation table can be replaced by a neural pronunciation model, and so on. Also, each of these neural network models needs to be trained individually on different tasks, and errors in any model in the pipeline can throw off the whole prediction.

More recently, deep learning has replaced these traditional statistical methods, such as Hidden Markov Models and Gaussian Mixture Models, as it offers better accuracy when identifying phonemes [5]. Recurrent Neural Networks (RNNs) are used to deal with this sequential information of audio data over time that corresponds to a sequence of letters.

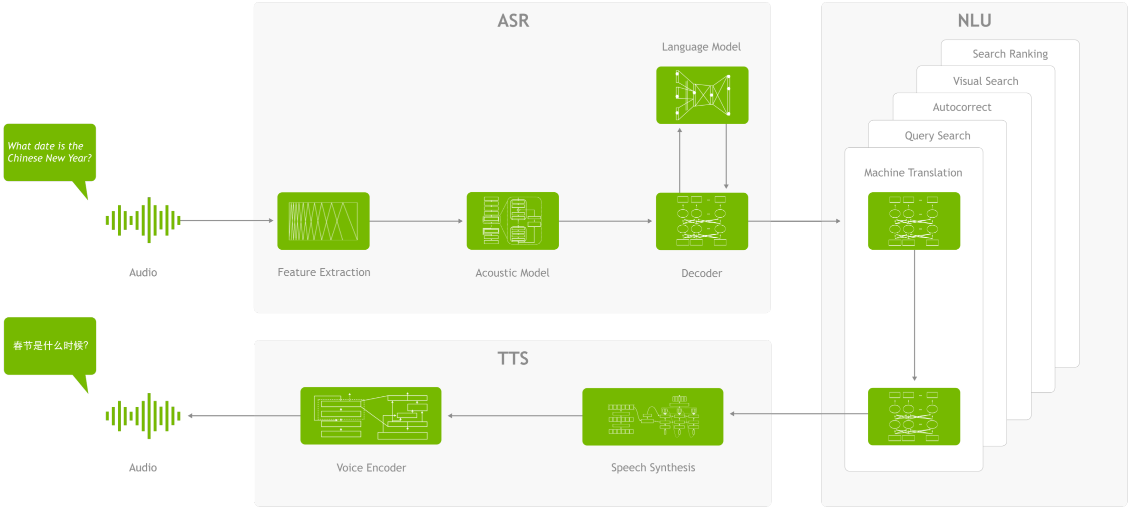

Fig. 3. Overview of a conversational NVIDIA AI pipeline

Source: From [4]

2.2 Text-To-Speech

This process involves several steps, starting with the output from the natural language understanding (NLU) stage, which is the text response generated by the AI system. The first step in the TTS process is linguistic analysis, which involves refining the pronunciations of words, calculating the duration of each word, deciphering the rhythm and intonation of speech, and understanding grammatical information. This step is important for ensuring that the resulting speech sounds natural and is easy to understand. The next step is to convert the text into a spectrogram, which is a visual representation of the frequencies of sound over time.

This is typically done using a synthesis network, which is a type of deep neural network that is trained on large datasets of speech data. The synthesis network is able to produce a spectrogram that captures the intonation and articulation of human speech. The final step in the TTS process is to convert the spectrogram into a waveform, which is a digital representation of an audio signal [4]. This is done using a vocoder network, which is another type of deep neural network that is trained on speech data. The vocoder network is able to produce a waveform that sounds natural and clear when played back as audio [4].

2.3 Natural Language Understanding

NLU is a branch of artificial intelligence that focuses on the ability of a machine or computer program to understand and interpret human language. NLU algorithms allow machines to understand the meaning of human language and respond to it in an appropriate way. This is a complex task, as human language is highly contextual, ambiguous, and constantly evolving. Deep learning models are often used for NLU because they are able to accurately generalize across a range of contexts and languages. Deep learning algorithms use multiple layers of interconnected nodes, or ”neurons,” to process data and identify patterns. This allows them to learn and adapt to new information and make predictions based on that information. /par

One of the most significant advancements in NLU has been the development of transformer-based models, such as BERT. BERT, which stands for ”Bidirectional Encoder Representations from Transformers,” is a type of deep learning model that is trained on large, unstructured datasets like books and Wikipedia articles. BERT is able to achieve human-like accuracy on benchmarks like the Stanford Question Answering Dataset (SQUAD) for tasks like question answering, entity recognition, intent recognition, and sentiment analysis [5]

Still, training language models like BERT are not without challenges. One of the biggest challenges is the lack of labeled data, which is data that has been labeled or annotated with the correct output. Most language models, including BERT, are trained on unsupervised tasks using large, unstructured datasets. This can make it difficult to train the model to produce accurate and reliable outputs for specific tasks. The process of encoding text into numerical vectors, which is a necessary step in many NLU applications, can also be challenging. Text encoding mechanisms like one-hot encoding and word embedding can make it difficult to capture the nuances of human language, and they can lose context when encoding long passages of text. BERT addresses these challenges by being deeply bidirectional, which enables it to better understand and retain context.

Therefore, NLU is a complex and rapidly-evolving field, and advances in deep learning and transformer-based models like BERT have greatly improved the ability of machines to understand and respond to human language. However, there are still many challenges to be addressed in the training and development of language models.

2.3.1 Sentiment Analysis

Sentiment analysis is the process of using natural language processing and other algorithms to automatically identify and extract subjective information from text data. This information can include the emotional tone of the text, such as whether it is positive, negative, or neutral, as well as the overall sentiment or attitude expressed by the text. It can be used for a variety of purposes, including determining the overall sentiment of a group of people based on their social media posts, analyzing customer feedback to identify trends and patterns, and providing input for other natural language processing tasks.

Overall, sentiment analysis is a powerful tool for automatically extracting and analyzing subjective information from text data. It can be used to gain insight into people's emotions and attitudes and to support a wide range of natural language processing tasks. This is an important component of therapy chatbots, as it allows the chatbot to automatically identify and respond to the emotional needs of the user. By using this technique, therapy chatbots can provide more personalized and effective support for individuals dealing with mental health issues.

3 Literature Review

The impact of anxiety and depression on individuals and society is significant and far-reaching. Anxiety and depression are common mental health conditions that can affect people of all ages, genders, and backgrounds. These conditions can cause a wide range of physical, emotional, and behavioral symptoms, including feelings of worry, fear, and sadness, difficulty sleeping, changes in appetite, and difficulty concentrating.

The impact of anxiety and depression on individuals can be significant. These conditions can interfere with a person's ability to function in their daily life, affecting their work, relationships, and overall well-being. Anxiety and depression can also lead to an increased risk of other health problems, such as heart disease and substance abuse.

The impact of anxiety and depression on society is also significant. These conditions can lead to decreased productivity and increased absenteeism from work, which can have a negative impact on the economy. Anxiety and depression can also strain relationships and put a burden on the healthcare system, as individuals with these conditions often require treatment and support.

3.1 Medical and Healthcare applications

In "A study about current digital assistants for healthcare and medical treatment monitoring" published in 2021 aims to give a broad-scope review of digital assistants and evaluate their targeted medical specialties [6]. The paper introduces the concept of digital assistant with the history of the Turing machine and ELIZA in order to garner its usage in healthcare. Healthcare requires a lot of workers to take care of patients and automating tasks is beneficial in improving patient response time, cost, and scalability. As Martins et al notes the benefits of them,

> "Digital assistants in healthcare have had an important role in healthcare, proving that their use could potentially improve the efficiency of health systems, as well as guaranteed access to high-quality medical care to everyone" [6].

The paper evaluates existing digital assistants in psychological therapy, symptom diagnosis and patient triage, and treatment monitoring through prior academic papers on digital assistants and healthcare [6]. Their main search term was "digital assistant" as they were specifically analyzing its usage in healthcare. Psychological therapy treats anxiety, depression, and other psychological problems. They utilize a chatbot in order to interact with the patient. This section is sparse as there are few commercial digital assistants on the market to test. The authors evaluated four different applications that serve as chatbots for mental health with minor variations in differences [6].

3.2 Anxiety and Stress Mapping

In "Analysing IoT Data for Anxiety and Stress Monitoring: A Systematic Mapping Study and Taxonomy" published in 2022 covers various techniques used to measure stress and anxiety [7]. They are focused on collecting data from the Internet of Things(IoT) devices that apply machine learning to detect levels of stress and anxiety. They analyze academic papers measuring stress or anxiety from the past 10 years from various databases [7]. The criteria for papers are studies published in a conference, workshop, or journal, presenting a computational approach to the problem, and a full paper [7].

After a thorough selection process, they reduced 260 papers to closely analyzing 56 papers that fulfilled their prerequisites of being pertinent to the research question. The paper answered five focus questions about sensors, identification of stress or anxiety, data analysis techniques, measurements of stress levels, and confirmation of said levels [7]. Generally, they focused on "presenting the data analysis techniques, the collection protocols, and the data collected" [7] into a visual to assess common trends. The paper used physiological data as concrete data on stress (or anxiety) and processed nonphysiological data into a more enriching dataset [7].

3.3 Depression Dectection

In "Ideal Construction of Chatbot Based on Intelligent Depression Detection Techniques" published in 2022 focuses on measuring levels of depression through sentiment analysis and facial recognition by evaluating respective papers [8]. It highlights the limitation of psychotherapy being contextual to society and certain aspects of psychology [8]. Chatbots are a way to alleviate issues of depression but do not solve them. This is hopeful as care is not universal for everyone. The pros and cons of technology show the gaps in our understanding of dealing with depression. Huang hopes that "From deep learning, it would be valuable to develop algorithms and instruments which are able to detect more featured variables and to process more data about the human's conditions in a short period in order to keep the fluent conservation" [8].

3.4 Examples

- Woebot is an automated conversation digital assistant designed around cognitive behavioral therapy with empathetic responses, goal setting, motivation, and reflection. It uses a survey to assess users' context and moods [6].
- Tess is a digital assistant aimed as a supplemental tool for therapy by providing support, psychoeducation, and reminders [6].
- Wysa is a mobile application chatbot designed to promote well-being and mental resilience [6].
- Mindful Moods is an application to assess depressive symptoms based on PHQ-9 (Patient Health Questionnaire-9 [6].

These commercial digital assistants all contribute to psychotherapy but have different functionality in terms of usage. They still aim to treat users in mental health, but only Wysa utilizes AI technology to better assess users. Perspective

3.5 Machine Learning Algorithms

The specific machine learning algorithm used by therapy chatbots will vary depending on the design and goals of the chatbot.

- Natural language processing algorithms, which are used to analyze and understand the meaning of the user's words and phrases. These algorithms can help the chatbot to identify the user's intent and respond appropriately.
- Sentiment analysis algorithms, which are used to automatically identify and extract emotional information from the user's words and actions. These algorithms can help the chatbot to understand the user's emotional state and provide appropriate support.
- Decision tree algorithms, which are used to model the decision-making process of the chatbot. These algorithms can help the chatbot to determine the most appropriate response to the user's request, based on a set of pre-defined rules and criteria.
- Reinforcement learning algorithms, which are used to improve the performance of the chatbot over time. These algorithms can help the chatbot to learn from its interactions with users and adapt its behavior accordingly.

3.6 Perspective

All papers are current and evaluate multiple academic papers to make a central conclusion that there is a growing interest in measuring stress, anxiety, and depression. These papers are new and not heavily cited, but the information is objective enough to provide a sufficient understanding of the development of quantifying emotion states with machine learning.

Each paper has its own perspective on the matter. Martins et al. take the stance of being prospective of emerging technology in the medical space. They are more optimistic in terms of the work done. This scope is beneficial for us to see how general digital assistants can be customized and tailored for specific uses. Paula et al. are a lot more thorough in their process to evaluate other academic papers with analysis [7]. This shows the extent of research that is being put into measuring anxiety and stress. Huang is explicitly targeting the elements of building a chatbot for the purpose of depression detection techniques [8]. Starting from a more general standpoint of digital assistants in healthcare [6] and working towards a particular treatment shows that predictive therapy chatbots with depression detection are a genuine concept [8].

These documents only show the computational approach of digital assistants and not the psychological perspective of psychoanalyzing users with technology. This proof of concept of measuring stress or anxiety [7], and depression [8] is not a machine learning solution, but merely a technological proof-of-concept [6]. Paula et al. show several tables of the various ways they used to record data such as wearable sensors from audio recording and heart-rate monitoring to mouse movements and gyroscopes [7]. The use case for anxiety and stress classification in varying academic papers shows the extent of research in modeling emotional behavior. Martins et al. and Narynov et al. show two different perspectives on the advent of well-being digital assistants. While Huang gives a more detailed approach to depression detection by providing past academic papers to display methods and techniques for a therapy chatbot.

3.7 Discussion

Depression is the 4th most common major disease [8] and affects 300 million people [7]. With the healthcare system overloaded [6] and the cost of psychologists prohibiting treatment [8], it makes sense to put automation and depression together to deal with this problem. Automation allows for care and access to be more widespread, but there are varying issues. Huang puts it together as

> "there is no such kind of chatbot that can independently make good counseling with the patients, and the best version of it is to act as an assistant to gather information for the human counselor [8]."

This summarizes the examples section of the digital assistant having limited usage within their respective scope. The current commercial products are simply band-aids for a cut. They do not replace the role of a therapist but serve as supporting [6] the injury. Even still a bandaid can greatly aid in healing an injury and promoting healthy bodily response. Digital assistants fulfill that role as a medium to provide more access to dealing with depression.

It would be great if we could combine all the current chatbots into one and have more coverage in use and treatment. Chatbots are successful since they are designed to do one thing well. Adding more features requires more data processing and training. More classification and categories will influence the accuracy of the machine learning model [7]. Paula et al. show that 45 papers out of 56 used machine learning as their main technique to analyze stress or anxiety [7]. Machine learning allows data to be sorted in high volumes by classifying information into categories. This in itself is a huge technological feat as thousands of academic papers on relevant algorithms [7] and thus need more development for an appropriate depression chatbot [8].

There are two main ways of effectively analyzing depression: sentiment analysis and facial recognition [8], [10], each with its own drawbacks. Word-based detection is good for recognizing depressing speech and facial recognition can find hidden emotions to better evaluate a patient [8]. However, both lack the judgment to treat a patient and the overarching issue with a therapy-based chatbot. According to a study, dialog agents (digital assistants) were found to be ineffective in providing prompt help when needed because they did not recognize that the questions being asked were symptoms of emotional or physical distress. When asked about issues such as suicide risk, some of these agents simply recommended that people seek professional help [9]. This runs the risk of the digital assistants now knowing what to say [11] and exacerbates the situation where the user might feel neglected. There need to be more deep-learning algorithms to detect more expansive variables and more psychological procedures for dialoguing depression treatment [8].

3.8 Limitations

- Lack of personalization: Like other predictive therapy digital assistants, therapy chatbots use algorithms and machine learning to provide advice and suggestions. However, they are not able to fully personalize their advice to the individual user, which can make it less effective for some people.
- Limited understanding: Therapy chatbots are limited in their understanding of human emotions and motivations, which can make it difficult for them to provide accurate and helpful advice.
- Ethical concerns: There are also ethical concerns surrounding the use of therapy chatbots, such as the potential for misuse or abuse of the technology.
- Dependence: Some people may become overly dependent on therapy chatbots, which can prevent them from developing the skills and knowledge they need to manage their own mental health.

Therapy chatbots can be a useful tool for some people, they should not be relied upon as a substitute for professional therapy or medical treatment. They can be useful for providing support and information to people who may not have access to traditional therapy or who prefer to use a digital platform.

4 Alternative Digital Assistants Uses

This juxtaposition of digital assistants and video games is important in improving human-computer interaction with digital assistance. Just as players enjoy video games for leisure, humans can utilize digital assistance in daily life. Using Fig. 1 as the guideline, video games can also fit this simple to complex scope.

4.1 Video Game Example

4.1.1 Command and Control

The first notorious digital assistant was Microsoft's mascot, Clippy. A paperclip with googly eyes and eyebrows that help navigated their Office software. It was known to be excessively redundant and more annoying than helpful because most of its users were adults who were already computer-proficient. The main audience did not fit its targeted audience showing how much market research context is important for digital assistants. Personally, I experienced using a dated copy of Microsoft Office when I was a kid in the early 2000s and felt Clippy was the most exciting thing in Office. Although I was 6 years old, I was rather curious about computers.

I grew up around television shows like *Blue's Clues*,*Dora the Explorer*, and *Sesame Street*. These shows utilized personified animals or animal mascots to promote early education platforms. These forms of media can be loosely interpreted as digital assistants with limited functionalities. As their whole appeal was centered around children. Command and control in early digital assistants simply revolved around a simple action for every input and output. With Blue Clues the actor pauses waiting for a response from the audience, the input does not determine the outcome as the video is prerecorded. This shows a rudimentary level of interaction between mediums.

4.1.2 Narrow Domain

I played a lot of kids-oriented learning games like *Reader Rabbit* and *Mavis Beacon Teaches Typing*. Moving down the line of figure 1, there is a digital assistant for a narrow domain. I would categorize these as mostly immersive games. With the previous example, there is no necessary end goal for the games. That was just inherently fun, but with early video games. These predetermined actions work well for building immersion in the player experience. Usually, a narrator or UI will help the player navigate around the video game in order to progress the story. The sense of progression only makes sense in a storyline and won't apply outside the game. The medium of video games shows the expanded capabilities of creating something purely enjoyable for most with all actions being predetermined beforehand.

4.1.3 General Purpose

The general purpose of video games is to provide players with the autonomy to explore and interact with the game world in their own way. Open-world games, in particular, offer players the freedom to roam and discover, allowing them to immerse themselves in a fantasy world. This type of game fulfills the need for players to have control over their own experience, as they are not forced to follow a specific story or set of rules. For example, in the game Red Dead Redemption 2, players can experience the Wild West as a member of a cowboy gang. The game world is expansive, allowing players to explore the Midwest from the 1860s-90s on horseback or on foot, hunting deer, engaging in gunfights, and completing various tasks and missions. The game provides a high level of autonomy and allows players to play at their own pace, without being forced to follow a specific story. The general purpose of video games is to provide players with a sense of freedom and control, allowing them to explore and interact with the game world in their own way. This can fulfill multiple needs for the target audience, such as the desire for adventure, exploration, and autonomy.

4.1.4 Human Level

Virtual reality and augmented reality games represent the next level of complexity for human-like conversation in video games. These types of games use player movements and inputs to allow players to interact with the game world and perform various tasks. While speech recognition can be used in these games, it is not commonly employed as a gameplay mechanic due to the fact that it can break the fourth wall, which is the invisible barrier between the game world and the real world.

In virtual reality games, the player's movements are often replicated in the game world, allowing them to move and interact with the game environment in a more realistic way. This can create a more immersive experience for the player, as they are able to interact with the game world using their own body and voice. However, there is still a degree of separation between the player and the game, as the player is viewing the game world from the perspective of the game character, rather than their own. They offer the potential for more realistic and immersive conversation in video games. Yet, the use of speech recognition as a gameplay mechanic is still relatively rare, as it can break the fourth wall and disrupt the player's immersion in the game world.

4.1.5 Discussion

Digital assistants with simple command and control exist already. There are mobility devices that are controlled by voice commands. Perhaps developing rudimentary digital assistance to monitor content learning for young children. For example, the child can say a show and watch it on a device. The assistance will limit his screen and answer to simple commands such as ″next″, ″stop″, ″time″ and ″watch.″ The next step is to cater the content for learning in the narrow domain category. The assistant can be customized with educational videos, games, and content. The assistant is able to keep track of his progression and able to ask related questions to mimic active learning. The next level of general purpose will implement a system where the child can ask general questions to the digital assistant and will respond back to them.

Nvidia recently showcase a demo showcasing their usage of digital assistance called Tokkio which responded to a mother taking food orders and responding to her child with food recommendations [12]. In the video, NVIDIA uses a digital avatar to give the kiosk a more friendlier appearance. Making a food ordering kiosk more life-like doesn't improve functionality except by passing the turning test. Similarly applying this digital representation of avatars is not necessary for digital assistance except for the emotional connection of facial gestures. This could be useful as a way for children on the autism spectrum to pick up social cues.

The next step is to make it better at handling requests and emulating emotional intelligence. Most digital assistants regurgitate information back, the information should be curated enough for the child to understand. The hope for human-level conversation is to display emotional intelligence to the user by socializing with the user, recognizing the user's emotions or intentions, regulating impulses, motivating or recommending the user solutions, and understanding the user's goals. These limitations carry over to therapy chatbots, as current digital assistants lack that sophistication.

4.2 Task Scheduling Example

Another example of the scaling of digital assistants is using them as productivity tools. The complexity scale can be illustrated using the metaphor of a watch, a calendar, an online calendar system, and a smart speaker.

A watch displays the time, date, and day, and its main function is to show the time accurately. The information provided by a watch is accurate and reliable.

A calendar can be inputted with important tasks and reminders, and an online calendar can be accessed from anywhere, making it more convenient to manage your schedule. The online calendar can also be shared with others, allowing for better planning and coordination of future events.

A digital assistant, such as a smart speaker, can function like a personal secretary, helping you schedule meetings, reschedule meetings, and manage your schedule. By talking to the digital assistant, you can review your plans and receive recommendations for avoiding scheduling conflicts.

To improve upon this model, another layer of machine learning can be added to enable the digital assistant to make more sophisticated and accurate recommendations. This would bring the digital assistant to the next level on the complexity scale, the autonomous level, where it would function like a personal secretary.

4.2.1 Possible Uses

By taking the immersion techniques from video games and the idea of personal companionship, therapy chatbots can utilize the different aspects of human-computer interaction to improve overall mental health. Digital assistants are moving towards a more accessible and inclusive approach. As conversation AI becomes increasingly indistinguishable from human interaction, digital assistants can act as a ″service robot″ for individuals with disabilities or chronic medical issues.

For example, self-assisting glasses with computer vision capabilities could help blind people navigate challenging environments, similar to how self-driving cars use computer vision. Additionally, digital assistants can provide support and companionship for individuals with mental health conditions, such as depression. While there are limitations and concerns about the medical efficacy of digital assistants, they can still provide a valuable coping mechanism for some people. By using media and AI as a way to express and process emotions, individuals can improve their overall well-being.

Predictive therapy chatbots are feasible and are a way to supplement important mental-health care for millions in need. Although there are all limitations, it could prove helpful in situations where therapy is restricted or unattainable. Chatbots for casual conversation and specific use are possible. since the "algorithms can understand what people want and what they say" [8]. Therapy chatbots can work but are blocked off by the AI needing to be able to understand the mental parts of humans to better serve them. Depression is not a simple treatment procedure for a chatbot to replicate, but serves as a stepping stone for solving real-world problems with ingenuity.

5 Conclusion

The use of digital assistants, such as voice-activated devices and AI-powered chatbots, has become increasingly common in recent years. These technologies have the potential to improve the lives of a wide range of individuals.

For the elderly, digital assistants can serve as a valuable tool for managing their daily lives. They can help with tasks such as setting reminders for medication, providing directions and offering information about local events and services. Digital assistants can also provide a sense of companionship and social connection, which is important for the mental well-being of older adults.

For children with learning disabilities, digital assistants can provide a convenient and engaging way to learn new skills and concepts. They can offer personalized feedback and support, helping children to build their confidence and motivation. Digital assistants can also provide a fun and interactive way for children to practice their reading and writing skills.

Individuals with chronic health illnesses and depression can also benefit from the use of digital assistants. These technologies can provide a convenient and accessible way to access information about health conditions and treatment options. They can also offer support and encouragement, helping individuals to manage their symptoms and improve their mental well-being.

Overall, the adoption of digital assistants has the potential to improve their quality of life and support their personal development. There are certain limitations and concerns about the use of these technologies, but they can provide a valuable resource for all of these groups.

References

[1] J. Epstein and W. . Klinkenberg, "From Eliza to Internet: a brief history of computerized assessment," Comput. Human Behav., vol. 17, no. 3, pp. 295–314, May 2001, doi: 10.1016/S0747- 5632(01)00004-8

[2] A. M. TURING, "I.—COMPUTING MACHINERY AND INTELLIGENCE," Mind, vol. LIX, no. 236, pp. 433–460, Oct. 1950, doi: 10.1093/mind/LIX.236.433.

[3] A. Mutchler, "What are Virtual Assistants? voicebot.ai. https://voicebot.ai/2019/10/05/what-are-virtual-assistants (accessed Nov. 11, 2022)

[4] Developer NVIDIA, "What is Conversation AI?" developer.nvidia.com. https://developer.nvidia.com/conversational-ai (accessed Nov. 20, 2022)

[5] NVIDIA Speech AI, "End-to-End Speech AI Pipelines" nvidia.com. https://resources.nvidia.com/en-us-speech-ai-ebooks-gated/speech-ai-using-asr-and-tts (accessed Nov. 11, 2022)

[6] P. M. Martins, J. L. Vilaca, and N. S. Dias, "A study about current digital assistants for healthcare and medical treatment monitoring," in 2021 *IEEE 9th International Conference on Serious Games and Applications for Health* (SeGAH), 4-6 Aug. 2021, Piscataway, NJ, USA, 2021, p. 7 pp. doi: 10.1109/SEGAH52098.2021.9551864.

[7] L. dos S. Paula, L. Pfeiffer Salomao Dias, R. Francisco, and J. L. V. Barbosa, "Analysing IoT Data for Anxiety and Stress Monitoring: A Systematic Mapping Study and Taxonomy," 2022, doi: 10.1080/10447318.2022.2132361.

[8] X. Huang, "Ideal Construction of Chatbot Based on Intelligent Depression Detection Techniques," in 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Feb. 2022, pp. 511–515. doi: 10.1109/EEBDA53927.2022.9744938.

[9] S. Narynov, Z. Zhumanov, A. Gumar, M. Khassanova, and B. Omarov, "Chatbots and Conversational Agents in Mental Health: A Literature Review," in 2021 21st International Conference on Control, Automation and Systems (ICCAS), Oct. 2021, pp. 353–358. doi: 10.23919/ICCAS52745.2021.9649855.

[10] A. Kartali, M. Roglić, M. Barjaktarović, M. Durić-Jovičić, and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," in 2018 14th Symposium on Neural Networks and Applications (NEUREL), Nov. 2018, pp. 1–4. doi: 10.1109/NEUREL.2018.8587011.

[11] J. Cho and E. Rader, "The Role of Conversational Grounding in Supporting Symbiosis Between People and Digital Assistants," Proc. ACM Hum.-Comput. Interact., vol. 4, no. CSCW1, p. 33:1-33:28, May 2020, doi: 10.1145/3392838.

[12] NVIDIA, "NVIDIA Tokkio Showcase" developer.nvidia.com. https://developer.nvidia.com/nvidia-omniverse-platform/ace/tokkio-showcase (accessed Dec. 5. 2022)