Chapter 11
AI Chatbots in Mental Health: Are We There Yet?

Raquel Simões de Almeida
Santa Maria Health School, Portugal & School of Health, Polytechnic Institute of Porto, Portugal

Tiago da Silva
Independent Researcher, Portugal

ABSTRACT

People with mental health problems often struggle in getting the suitable treatment regarding not only the type of interventions available but also the conditions required for a proper treatment, mainly cost, locality, and frequency. The use of AI chatbots for this population is a new trend and can reduce the gap between the need for mental health care making them accessible in a cost-effective way. Although chatbots are not a substitute for formal treatments, they are sometimes used in tandem with other treatments with positive results. This chapter provides a review on the subject, presenting several chatbots for mental health problems and also addressing some concerns such as privacy, data security, AI limitations, and ethical implications. Future research directions are also discussed.

INTRODUCTION

In 2017, the tech giant IBM stated that Artificial Intelligence (AI) will transform the delivery of mental health care over the next five years by helping clinicians better predict, monitor and track conditions, and that "what we say and write will be used as indicators of our mental health and physical wellbeing" (IBM, 2017). In 2021, we are already seeing some of those promised transformations and positive impacts.

Chatbots, as part of AI devices, are natural language processing systems acting as a virtual conversational agent, mimicking human interactions. While this technology is still in its developmental phase, health chatbots could potentially increase access to healthcare, improve doctor–patient and services–patient communication, or help to manage the increasing demand for health services such as remote testing, medication adherence monitoring or teleconsultations. The chatbot technology allows for activities as specific as health surveys, setting up personal health-related reminders, communication with clinical

DOI: 10.4018/978-1-7998-8634-1.ch011

teams, booking appointments, retrieving and analysing health data or the translation of diagnostic patterns considering behavioural indicators like physical activity, sleep or nutrition. Such technology could potentially alter the delivery of healthcare systems, increasing uptake, equity and cost-effectiveness of health services while narrowing the health and well-being gap, but these assumptions require further research.

The development of Cognitive Behavioural Therapy (CBT) chatbots, which mimic normal conversational style to deliver CBT interventions (Kirkpatrick et al., 2017; Inkster et al., 2018) are being developed, however the outcomes are still unclear given the initial stage of research. These advanced chatbots rely on AI techniques to implement the conversational style that mimics a normal conversation as if it were a human being on the other side interacting with the user.

The use of AI chatbots may also increase accessibility by overcoming some barriers associated with stigma in the demand for services. Due to stigma, individuals with psychopathology tend to have reduced social support, being mostly supported by family members. However, users are more prone to perceive chatbots as non-judgmental (Lovejoy et al., 2019). AI chatbots are increasingly being seen by psychiatrists, psychologists, therapists, politicians and tech companies as having a significant role in future mental health treatment and care, with developments in the field being driven by their particular agendas and goals. Nonetheless, it appears that key stakeholders are currently excluded from the discussions about AI in mental health – service users, carers, and families. If rights-based guidelines for ethical AI (AI HLEG, 2019) are to be implemented in mental health, then the implications of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) needs to be considered. The UNCRPD have said that it is essential to involve disabled people (including those with psychosocial disabilities (Szmukler, Daw, & Callard, 2014) and their representative organisations in developments and decision-making that will affect their lives. It is therefore time to assess the situation, to question those who are driving this transformative agenda forward and to listen to excluded experts – those whose lives will be ultimately impacted by these technologies (Carr, 2020).

On the other hand, the current limitations of AI and chatbot technology and lack of knowledge about the real capabilities of the available tools may lead to an inappropriate usage of these technologies as a serious approach to solve mental health issues. Most of the tools and applications available focus on specific subjects or problems e.g. applications to cope with stress, anxiety, addictions are publicly available in the market, but consolidated applications that aim for mental health on its broader spectrum do not exist. This problem is directly related to the maturity level of conversational AI and chatbots, a metric that is described by different authors using different methods and perspectives. Bosek (2018) describes it as a 4-level pyramid based on the features the system delivers, Gadiyar (2020) describes it as a 3-level schema based on the automation and communication skills of the system, Garga (2020) uses a multi-layered ellipse to describe the technology in three major vectors: interaction, intelligence and integration. These three analyses help to demystify a broader romanticized understanding of AI and chatbots that only exist in science fiction and that were popularized by Hollywood movies like "Her" (2013) and "I, Robot" (2004), describing the real limitations and scope of conversational AI.

Data privacy is also a main concern when it comes to healthcare applications. Most of these tools collect, process and store data, in most cases sensitive data such as mental health status that raises all kinds of ethical, legal and moral questions. What software companies do with this data should be a subject of analysis. Even if the main goal is to provide a service with the aim of helping the user, more and more frequently the collected data is used for different purposes, sometimes with little or no knowledge and lack of explicit consent of the user. Moreover, news of data leaks that expose sensitive data of the same users that put trust in an application to collect and store their own personal data, are becoming increas-

ingly common. From 2005 to 2019, roughly 250 million individuals were affected by healthcare data breaches, being 157 million in the last five years alone (Seh et al., 2020). This is clearly an issue and a risk when using chatbots designed to help users to overcome or cope with mental disorders - what happens to user data right after being collected by the chatbot should be a major subject of concern. What are the major threats and opportunities, what is the right balance between the data that is asked from the user and the benefit produced by these companies and organizations, how can this process of asking, storing and treating personal and health data can be as transparent as possible for all stakeholders, and what mechanisms exist to protect data, these are all questions worth to explore.

Although there are plenty of studies that describe the benefits of using technology in mental health, there is an aspect that can be a potential issue while using chatbots and technology in general: the well-studied effects of addiction and social isolation (Pontes, 2017), which can be especially dangerous in this population - some of these individuals go potentially undiagnosed, others embraced self-diagnose and self-treatment (Bauer et al., 2017). There is apparently contradictory evidence that points in different directions when it comes to usage of technological devices and applications by users with some form of mental illness.

Throughout this chapter, the authors intend to address all these themes with depth and reflection beginning with a comprehensive and comparative presentation of the state of art of AI chatbots relevant for mental health, including a critical analysis of weak and strong points of each application. Next, the authors provide an analysis of how these applications and tools can profit from the inclusion of key stakeholders and potential users right from the design phase and overcome some of the most common obstacles and flaws faced by the most used chatbots. Then, the authors explore the technical boundaries and the current technical limitations of AI and chatbots, how those limitations are an obstacle to better applications and what the immediate landscape in terms of new developments in conversational AI looks like. The authors also explore the benefits and the dangers of data handling by applications of this sort, where data can be provided by the user without the same kind of awareness as if the user is wilfully filling a form. To conclude, the authors make some observations about the pros and cons of the use of AI chatbots by mental health patients, exploring why some studies point to social isolation and proneness to anxiety while others present positive results and progress while using technological applications by this population, especially chatbots.

BACKGROUND

Mental health problems are a growing concern worldwide because they impair quality of life, cause disability and represent an issue to the economy (GBD Disease and Injury Incidence and Prevalence Collaborators, 2018).

The demand for better mental health services has increased, and meeting these demands has become increasingly difficult and costly due to a lack of resources. Therefore, new solutions are needed to compensate for the deficiency of resources and promote patient self-care. Distance can delay the reach of traditional mental health services to populations in remote areas in both high-income and low-income countries. Technology-based treatment, such as mobile apps, can overcome most of these barriers and engage hard-to-reach populations.

One technology that offers a partial solution to the lack of capacity within the global mental health workforce is mobile apps. Due to factors like cost, portability, ubiquity and ease of use, mobile apps are

becoming powerful tools that contribute to making mental health more accessible and affordable. One of the main mobile apps used for mental health are conversational chatbots or simply chatbots, computer programs able to maintain a conversation and interact with human users. These systems use spoken, written and visual languages to interact with human users. The number of available chatbots and the number of users of these systems is growing year after year with special focus on the last decade. This is also true for chatbots aimed for mental health. More recently, the Covid-19 pandemic has exposed even more the need for these kinds of services, because of the limitations in the access of healthcare services and also because of the psychological consequences of the lockdowns (Miner, Laranjo, & Kocaballi, 2020). Some authors argue that chatbots will address the lack of mental health care services available and also, they can facilitate interactions with people that do not search for mental health care services due to stigmatization and allow more conversational flexibility (Abd-Alrazaq, Rababeh, Alajlani, Bewick, & Househ, 2020).

ARTIFICIAL INTELLIGENCE AND THE EMERGENCE OF CHATBOTS

Chatbots are becoming more prevalent in our daily lives, as we can now use them to book flights, manage savings, and check the weather. Chatbots are also increasingly being used in mental health care, with the emergence of "virtual therapists". Chatbots are programs "that use machine learning and artificial intelligence methods to mimic human-like behaviours and provide a task-oriented framework with evolving dialogue able to participate in conversation" (Vaidyam et al., 2019); some include psychotherapeutic interventions (like cognitive behavioural therapy techniques) offered in real-time and may have a role to play in patient care — in a sense, therapy without the (human) therapist. Chatbots may offer certain advantages: unlike a human therapist, a chatbot is always available when the patient chooses to make contact, never distracted by thinking about what to cook for dinner (or anything else), and "remembers" everything a patient told it through its data repository, using that information to develop a more data-informed understanding of the patient.

A clear and recent example of the development of a chatbot with positive and direct impact in a population that experience mental health and wellbeing challenges has been made by Christine Grove. In this case, a chatbot using AI technology to respond to anxiety, stress and depression in youth was developed by different stakeholders that included healthcare experts, technological experts and the target population. The result of this collaborative work was the design and development of a chatbot that answers to the needs and habits of this specific population having the input and support of healthcare professionals (Grove, 2021).

AI is being widely used by different applications and services in order to deliver a better and more personalised user experience, offering to each user a different experience based on the user needs and preferences. AI is also used to create models that can predict or detect mental health conditions. A good example is the detection or prediction of "digital exhaust" of a given individual based on data gathered from the different systems that the user interacts with. The data collected can be analysed to produce conclusions and insights. And this data can be something from simple metrics like the number of hours spent interacting with digital systems to more advanced techniques like natural language processing. AI algorithms are able to read and interpret natural language and extract insights as powerful as inferences of the current mental health status of an individual but also able to provide complex answers, which is a great feature to be used by chatbots for therapeutic intervention (D'Alfonso, 2020).

Natural Language Processing (NLP) is a field of Artificial Intelligence that interprets what the user says or writes using its own words as one would usually do in an ordinary situation. NLP then converts the sound or the text into machine language, which is binary. It is not a simple word recognition technique that translates word by word, it is more than that as NLP tries to get context or the intention of what the user is trying to say, explain or ask using AI techniques like pattern recognition. This is a powerful approach to human-machine interaction as it removes all the complexity of sending commands to the system to execute tasks or set the environment, NLP tries to do exactly that just by analysing the natural language of the user.

Machine Learning is also a field of Artificial Intelligence that uses data analysis to train and produce models that are able to categorize, segment and predict values based on past observations. With a model or multiple models running in parallel and connected between them imitating neuronal connections (Deep Learning), it is possible to provide different but correct answers to different requests from different users with a degree of probability based on past data. Thus, NLP and Machine Learning used in tandem are the basis for an AI powered chatbot that is able to understand common language used by its users and provide responses that are based on past learnings of the current user or of other users.

As AI techniques continue to be refined and improved, it will be possible to help mental health practitioners re-define mental illnesses more objectively than currently done in the DSM-5. Identify these illnesses at an earlier or prodromal stage when interventions may be more effective, and personalize treatments based on an individual's unique characteristics. However, caution is necessary in order to avoid over-interpreting preliminary results, and more work is required to bridge the gap between AI in mental health research and clinical care (Graham et al., 2019).

The rapid integration of AI into the healthcare field has occurred with little communication between computer scientists and doctors. The impact of AI on health outcomes and inequalities calls for health professionals and data scientists to make a collaborative effort to ensure historic health disparities are not encoded into the future. There are studies that evaluate bias in existing Natural Language Processing (NLP) models used in Psychiatry and discuss how these biases may widen health inequalities (Straw & Callison-Burch, 2020).

Abd-Alrazaq and colleagues have done a scoping review (2019) and a systematic review with meta-analysis (2020) to compare the different levels of safety and effectiveness across mental health chatbots using results of previous studies and research. The aim was to assess the effectiveness and safety of using chatbots for improving mental health. In the scoping review, the authors checked 53 chatbot studies - 17 of them aimed for therapy, 12 for training and 10 for screening. Forty-nine of the analysed chatbots were rule-based and implemented in stand-alone software. In forty-six studies, chatbots controlled and led the conversations. The most common form of input is written language (seen in 26 studies) and the most common output form is a combination of written, spoken and visual languages, 28 in total. In the majority of studies chatbots have some form of virtual representation. The most common scope of these chatbots are depression or autism. This review reveals a great offer of chatbots with different focus and approaches, which means that healthcare providers have a variety of tools they can use or promote to help their patients and their mental health needs. However, in the systematic review they conclude that there was not sufficient evidence to draw solid conclusions about chatbots safety and effectiveness at this moment.

According to a scoping review on perceptions and opinions of patients about mental health chatbots, the results showed positive feedback, although it is necessary to invest more in the good quality

and variability of responses to unexpected questions from users (Abd-Alrazaq, Alajlani, Ali, Denecke, Bewick, & Househ, 2021).

EXAMPLES OF CHATBOTS IN THE MENTAL HEALTH FIELD

There are several chatbots in the mental health field currently available and some of them enjoy a high degree of popularity and success. This often happens because of a multiplicity of reasons.

The first one is the scalability aspect of these systems. Unlike human professionals, chatbots can be replicated fast, easily and at a low cost and be available at any time anywhere via the internet. They are not also affected by fatigue or cognitive errors and they do not have any personal bias. Because chatbots are depleted of these human factors, some users may see in chatbots a way of being more transparent when discussing private matters, concerns or acts such as diseases or risk behaviours. Users can also feel less anxiety exposing their private lives to a chatbot than to a human being. Some chatbots allow the configuration of parameters to be closer to the user and offer a more personalised experience, these parameters can include the physical appearance of the avatar, language, accent, mannerisms, race, ethnicity or socioeconomic status. This is an important feature that aims to establish a stronger relationship with the user hoping to contribute to a higher impact and engagement of the treatments and ultimately a better health outcome.

- WYSA

Wysa is a chatbot with focus on the management of anxiety, loss, worries, energy, sleep and other issues. According to the authors "Wysa is an AI-based emotionally intelligent mobile chatbot app aimed at building mental resilience and promoting mental well-being using a text-based conversational interface". The Wysa app assists users to develop positive self-expression by using AI to create an external and responsive self-reflection environment." The app is free however if the user wants to interact with a human coach that is a paid service. Using Facebook Messenger as its own interface and chatbot capabilities, "the app responds to emotions that a user expresses over written conversations and, in its conversation, uses evidence-based self-help practices such as CBT, dialectical behaviour therapy, motivational interviewing, positive behaviour support, behavioural reinforcement, mindfulness, and guided micro actions and tools to encourage users to build emotional resilience skills" (Inkster, Sarda, & Subramanian, 2018).

- TESS

Tess is a psychological artificial intelligence chatbot, which delivers emotional wellness coping strategies. This mental health chatbot coaches patients and also caregivers to create resilience by having text message conversations using a mix of machine learning techniques and supervised intervention by psychologists, Tess provides a variety of interventions that responds to the users' needs. The users can access this app through Facebook Messenger, SMS texting, web browsers, and also as smartphone apps (Joerin, Rauws, & Ackerman, 2019).

- WOEBOT

Woebot is a fully automated conversational agent developed by Woebot Labs in San Francisco. It treats depression and anxiety using a digital version of time-tested cognitive behaviour therapy. NLP techniques are used by Woebot to build a human-like conversation interface. According to a 2019 comparative study, it was found that Woebot was able to have a positive effect in people that suffer from depression (Singh, 2019).

- ELLIE

Ellie almost serves as a virtual therapist as it can detect subtleties in facial expressions, rates of speech, or length of pauses and responds accordingly. It also provides an option to meet an actual therapist (Fitzpatrick, Darcy, & Vierhile, 2017). Ellie is part of a larger project developed by the University of Southern California that identifies and tracks multimodal signals like body posture, language patterns or facial expressions to detect signs of stress or anxiety in patients.

- YOUPER

Youper is a mental health app with a chatbot it calls an "emotional health assistant". For users who have never consulted with a clinical professional, Youper introduces the types of questions and exercises they might experience in therapy. The questions and exercises given by Youper's chatbot are meant to help users achieve a better understanding of their emotions, thoughts and behaviour (using CTP techniques and strategies). Youper's chatbot asks users to focus on their thoughts and identify how they are feeling from a list of descriptive words. Then a scale lets them rate the strength of that emotion from "slightly" to "extremely." More questions help them narrow down what is causing those feelings and track their mood. As it learns more about the user, it fine-tunes the experience to better fit users' needs. Users are also given options for mindfulness exercises and journaling prompts (Shu, 2019).

- REPLIKA

A companion chatbot that is "an AI companion who cares" and was created to provide a place for people to express themselves in a "safe, judgement-free space" and engage in meaningful conversations. Once a user downloads the Replika app, he/she may choose to apply several characteristics to their Replika, such as a name and gender. Interactions with Replika primarily function through text-based communication, enabling users to converse with their Replika on their smartphones or computers. Like other chatbots, increased interactions with Replika allow it to learn more about the user, and it is built to resemble natural human communication as much as possible (Ta et al. 2020). Replika makes use of Open AI's GPT-3 language model that runs deep learning techniques to generate text and conversations indistinguishable from humans.

The ability of chatbots to provide companionship, support, and therapy can lessen the load on therapists. It emerges as an option for people who have problems with accessibility and affordability both in terms of time, distance, and finances. However, several concerns are being raised in this matter. Confidentiality is the foremost concern. Other concerns are universality of application, lack of standardization and monitoring, overdependence on the bots, and lack of severe mental disorders. We need to develop chatbots more "suited" to our culture and have a regulatory and evaluating process in place to enjoy the benefit of this technological advancement (Singh, 2019).

NEW TRENDS

The impact of the various presentation modalities currently used by chatbots (text, verbal, or embodied as a 3D avatar) and the preference therein remain largely unknown. While some groups have claimed that voice, and not animation of a 3D avatar, is the primary determinant of a positive experience with a chatbot, it remains difficult to conclude today as no studies compared adherence or engagement measures between chatbots of identical functionality but different modalities.

Vaidyam and colleagues (2019) highlight the need of establishing appropriate rapport or therapeutic alliance on patient interactions, because an early alliance establishment predicts more favourable outcomes. However, more research is needed to know exactly how patients feel supported by chatbots.

Creating chatbots with empathic behaviours is an important research area. Exhibiting humanlike filler language such as "humm"s and "ah"s may allow patients to feel more socially connected, and studies focusing on adding these behaviours into chatbots suggest that such simple and subtle changes may more effectively build rapport. With today's technology, patients must be explicit about their emotions while communicating with a chatbot since they cannot reliably understand the subtleties or context-dependent nature of language. However, since such explicit dialogue would be unnatural between humans, it may break an established illusion with the chatbot. In addition, chatbots that ask scaffolding-based questions with open-ended "why" or "how" prompts, subsequently leading to irrelevant and non-contextual conversation, risk losing the interest and alliance of the patient. Another challenge regarding empathy is that patients know chatbots cannot empathize with "lived experiences" so phrases such as "I've also struggled with depression" will likely fracture the patient-chatbot relationship (Vaidyam et al., 2019).

OLD CONCERNS

It is obvious the potential of chatbots and other AI powered services to help address some health care services, however it is always difficult to provide an e-health service that safeguards aspects that are crucial in a relationship between professionals and patients that include dignity, respect and ethics. This is a two-way challenge. On one hand it is desirable that developers include healthcare professionals during the software development life cycle in order to ensure that from a technical standpoint the product delivers the most accurate response to the user needs. On the other hand, ethics codes and practice guidelines of healthcare professionals should include the use of technologies as part of common practice.

There is the need and the space for global organizations like the World Health Organization to lead and promote a cooperative environment to address these questions, create guidelines and promote the safe and effective use of technology in the healthcare space. Moreover, the cooperation between software developers and healthcare professionals should also include the input of patients and potential users of these technologies, especially the ones that are underserved and most affected by health-care disparities as valid and important stakeholders. This approach would also potentially accelerate the adoption of new technologies with evident benefits for all stakeholders (Luxton, 2020).

Ethical Implications

Much of the impact of the usage of chatbots by patients is still unclear. Some aspects that are taken for granted in in-person relationships between healthcare professionals and patients like privacy and confi-

dentiality are yet a discussion topic when analysing chatbots and other services that deal with personal and healthcare data. Although there is specific legislation for data handling aimed for technological services (more of that in a topic below) in the United States, most chatbots are not currently covered under the Health Insurance and Portability and Accountability Act (HIPAA), meaning that there is the potential of users' personal data to be handled freely and possibly traded by companies the owners of these services bypassing the usual confidentiality rules followed by classic in-person consultation.

It is also important to consider the relationships that may be formed between users and chatbots. Due to the constant availability of these chatbots (24 hours per day, every day of the year), there is the risk of addiction and isolation by these users that seek responses for their mental issues making the case worse. In these cases, there is not a clear responsibility that can be pointed to chatbot developers as laws and regulations are practically non-existent.

Currently we are experiencing a boom of chatbots and other conversational services aimed for mental health, however most of the research seems to be happening in the engineering and technological side, leaving behind aspects that should be a concern for research in the mental health scope. For instance, there is a lack of tools and frameworks that can evaluate the outcomes and impacts of these services as well as transparency of how these systems work as many of these companies work in a market logic and that see intellectual property as part of their business. Until such evaluation tools are created, it will be difficult to compare and understand in depth what chatbots do under the hood, how they react to different scenarios, what are their outcomes and what are the real impact they have to their users.

Although we are living in a world where there are virtual assistants that are able to perform tasks such as booking restaurant tables and interact with humans sounding indistinguishable from a human, we are still to see these assistants do the same kind of interaction to create diagnoses and implement treatments and therapies (Vaidyam et al., 2019). Nevertheless, we can already speculate that different ethical issues will surge when conversational chatbots reach to a fairly advanced state that makes them indistinguishable from human beings and proactively integrate with other services or data sources in order to provide an answer to the user (Gratzer and Goldbloom, 2020).

Risk of Harm

Due to the level of autonomy of chatbots, there is a real risk of harm for their users. This can be especially problematic if the technology does not address or identify scenarios of risk. For example, a person conversing with a chatbot could reveal that they are experiencing suicidal thoughts, this should be a cause for alert and immediately addressed. Also, patients that may suffer from some kind of psychotic symptoms or cognitive deficits, may not be suitable candidates for the use of chatbots. To help address these concerns, the stakeholders of these technologies should put in place processes or rules to identify who are the suitable users for each application and what are the potential risks. It is also desirable that these systems are able to monitor and alert risk scenarios. Ideally, these chatbots should provide immediate help or the resources to search for help, such as a phone number or a contact with the responses for each scenario. It is also desirable to have human intervention and validation; a review of the information is key to monitor the functioning of the system and the prevention of risk.

Currently, most companies that develop mental health chatbots, describe them to the public as information providers or training tools, and not as replacements for health professionals. This is a simple way of avoiding the same level of responsibility that is demanded for healthcare professionals. Making clear what are the scopes, limitations and potential risks of each chatbot, platform, application or system

is essential to set expectations, both for users and healthcare professionals. Moreover, these chatbots developers, owners and other stakeholders should have a clear picture of what are the services and resources available in each region or country and helping users to contact them in case of need is also a key requirement to mitigate risk.

Artificial Intelligence Limitations

The romanticized narrative of AI and chatbots that only exist in science fiction and that were popularized by Hollywood movies like "Her" (2013) and "I, Robot" (2004) may lead to inappropriate usage of mental health applications and set unrealistic expectations of these technologies as a serious approach to solve mental health issues. Many of these misleading ideas are in part caused by unrealistic marketing campaigns around the topic of AI and popular beliefs influenced by science-fiction. In reality, most of the tools and applications available that claim to have AI technology make a limited use of AI resources and most of them focus on very specific subjects or problems. There are in the market applications to cope with stress, anxiety and addictions among other specific topics, but consolidated applications that aim for mental health on its broader spectrum and that are able to handle a vast set of mental conditions do not exist.

This problem can be linked to the current overall maturity level of conversational AI and chatbots, a scale that is described by different authors using different methods and perspectives. Bosek (2019) describes it as a 4-level pyramid based on the features the system delivers, Gadiyar (2020) describes it as a 3-level schema based on the automation and communication skills of the system and Garga (2020) uses a multi-layered ellipse to describe the technology in three major vectors: interaction, intelligence and integration. Each perspective makes clear that AI technology is yet being perfected and matured and that the current capabilities are just a fraction of what could be achieved in theory in the near future - maybe closer to the science fiction movies cited above.

The last level of Bosek's pyramid is what the author defines as Turing Bot, an AI application indistinguishable from a human-being capable of understanding the user and providing responses considering the situational conditions. An application with this kind of capability would be as good as a trained human, however we are far from this reality and the author forecasts that we are years away from this level of AI (if possible).

Gadiyar's perspective of the near future of bots is what the author defines as a master bot or a bot of bots. In his point of view, bots are becoming more and more specialized, opening the need for the orchestration of multiple bots in one single point, the master bot, that would work as a meta bot that calls the adequate bot to solve a specific user question or request.

In Garga's ellipse, the outermost layer (the most advanced form of chatbot) defines a system able to start a conversation based on a context, recognize the user's mood and build or have access to a personalized knowledge base, all requirements for an advanced chatbot. Like in the former two researches, Garga also did not find an example of a bot in any context that fulfils the most advanced requirements that define an advanced chatbot capable of replacing a trained human.

Until technology reaches an evolutionary stage of AI, human-machine interactions have limitations that should be considered when using and developing conversational applications and services in healthcare, especially in mental health. Users should be aware of these limitations as well as other stakeholders that encourage the use of chatbots and AI-powered applications as a substitute or complement for a healthcare professional.

Privacy and Data Issues

In Information Technology, and in chatbots in particular, user privacy is a concern that needs to be taken seriously. If poorly addressed the potential for user harm is enormous. Chatbots have the capability of collecting and storing large sets of private and sensitive information of their users. Laws and regulations about data collection vary from country to country and from region to region. The developers of these technologies should be aware of the legislations that apply, cope with them and inform users about privacy, data protection and terms of service.

There are many regulations that protect users against the misuse of personal data and enforce developers and administrators of these systems to comply with policies and standards. Although most of the time users are willing to give some details of their own personal information to companies in exchange for products or services, there is always the risk of this data being used within a different scope or with different purposes than the initially agreed by the user. Aggressive legislation exists to fight this kind of conduct, the European Data Protection Regulation (General Data Protection Regulation (GDPR) – Official Legal Text, 2019) and California Consumer Privacy Act (California Consumer Privacy Act (CCPA), 2021) are two of the most advanced regulations that cover these and other misuses of personal information with heavy penalties for companies that do not comply with the rules.

A common practice of personal data misuse is seen in applications that offer a service in exchange of the user's personal information. The user is lead to believe that the personal information that he/she is about to provide is fundamental for the good execution of the service and that the shared information will be used only for that purpose, however the information now on the hands of a third party is used for other scopes that can include direct advertising of unrelated products or services, data trading with other companies or data inference. Famous examples of data misuse by tech companies include for instance the 2014 Uber's God View incident where both drivers' and riders' GPS position were used for other purposes than Uber's goal of moving people from point A to point B (Hill, 2014), or the Cambridge Analytica scandal during the US elections in 2016 (Confessore, 2018). Big tech companies like Google and Facebook are often in the news because of the suspect of data misuse and they were even called to public hearings in Europe and in the United States to explain their methods and goals around the collection and handling of personal information, but smaller companies are also referenced in the news from time to time due the abusive use of their users' personal data. This reveals that data misuse is a transversal issue across all kinds of companies and that the general public is more and more aware of the importance of data management.

Electronic health records are one type of personal information that has a great value for tech companies working in the healthcare area. An individual's healthcare data can have many useful and positive usages, for instance, it can be used on all sorts of algorithms or compared against databases of healthcare data in order to find patterns that can alert for possible diseases or health issues that can be treated in an early stage. On a negative note, knowing healthcare details of an individual may have an undesirable impact in scenarios like the enrolment for a health insurance plan or in the hiring process for a new job. Today, these kinds of data points are collected seamlessly by mobile devices and wearables that continuously track users' heartbeat, physical activity, sleeping patterns and other healthcare related metrics. Often the user gives an initial consent for data collection but not knowing exactly what will happen with it once collected. Chatbots are also a clever way of collecting data, they use a conversational style that leads the user to slowly giving details of its personal information in opposition to the more traditional way of data collection using forms. In the healthcare chatbot's scope, these personal details are often informa-

tion about the current health status that may include for instance an overall mood, physical wellbeing details or mental issues.

Even when coping with all standards and best practices of data handling, a serious issue that tech companies face is the chance of data breaches or data leaks. An event like this expose publicly or to unauthorized third parties personal data that is intended to be secured by the companies who collected and stored the data. Most of these events happen because of an attack intended to steal data or because of deficient implementation of data security, in any of these scenarios the security of any personal data, including healthcare data, is compromised and exposed to people or organizations that should not have access to this information. Current regulations include penalties for organizations that expose their own users' personal information in data breaches and data leaks or that do not comply with the security measures to protect data. For instance, in 2013, Adobe suffered one of the biggest attacks ever which exposed personal data, including credit card information, of 150 million users. In August 2015 Adobe had to pay $1 million USD in legal fees and an undisclosed amount to users to settle claims of violating the Customer Records Act and unfair business practices (Hern, 2017).

There is a long road to be done in order to educate common users how to effectively handle and provide their personal and healthcare data to third parties, organizations and services. Cybersecurity awareness is a topic that is not top of mind and a priority for common users of IT products and services (Koyuncu, 2019). Having informed and educated users of these applications and services help to have a correct handling of data and prevent catastrophic events like data leaks. On the other hand, there is also a long road in order to force tech companies to implement best practices and proven standards around data privacy and protection. Although security is a main aspect of software development, sometimes shortcuts are taken when building an application or a service that jeopardises the security of applications and postponing or deprioritizing security implementation is something that often happens when deadlines to deliver a product are approaching (Hala, 2019). Both parties - users and tech companies - are responsible for their ends of data management, a poor handling of data security can lead to disastrous consequences for both users and organizations undermining the confidence on technological solutions to solve healthcare issues.

DESIGNING AI CHATBOTS IN MENTAL HEALTH

To increase usability, engagement and security, the focus should be on developing short, simple, and consistent modules and testing them with small iterative studies. Then, developers can move toward expanding the content (or modules) of the chatbot. As with most digital interventions, the attrition rates are significantly high; therefore, developing an extensive set of modules that users do not end up engaging with is not a good use of resources. Research on frameworks for developing engaging and effective chatbots offers the opportunity to create and test scalable interventions. Data from large studies on chatbots could lead to effective personalized interventions that could eventually answer the question of which intervention works for which individual (Dosovitsky et al., 2020).

Safety and trust are factors to be considered when developing new AI-based technologies. This is the opinion of different experts in AI, healthcare professionals, software developers, system owners and administrators and other stakeholders. That need was clearly stated in a recent Lancet commission on global mental health - "technology-based approaches might improve the reach of mental health services

but could lose key human ingredients and, possibly, lower effectiveness of mental health care" (Patel et al., 2018). Human supervision will always be needed at some point.

AI powered chatbots may augment the work of psychotherapy, therefore product designers, healthcare professionals and researchers must evaluate the impact of these new approaches and practices on mental health patients and professionals. The changes of processes and workflows must be considered when AI-powered treatments are made available and used. The deployment of these technologies must be followed up by training and awareness campaigns to inform clinicals of the possible impacts, limitations and scopes of these technologies. And because this requires the expertise of two different areas, healthcare and IT, the discussions about the needs of each end should be constant. Having a healthcare professional understanding the capabilities and limitations of a technology is as important as an IT professional to understand the needs of a healthcare professional.

Therefore, these are Kretzschmar and colleagues (2019) recommendations for chatbot development around three major vectors: efficacy, privacy and safety.

Efficacy:

- The technology provided should be evidence-based;
- Platforms should be tested empirically;
- Users should be informed about the extent to which the service is backed up by evidence;
- Users should be informed about what the chatbot targets and what effects to expect.

Privacy and Confidentiality:

- Personal information, if collected, should be kept confidential;
- Content of conversations, if shared, should be de-identified;
- Privacy arrangements and limitations should be made transparent to users;
- Users should have the option of being reminded of privacy arrangements and limitations at any stage.

Safety:

- Users should be informed that they are talking to a robot;
- Automated chatbots should encourage people to seek human support;
- Automated chatbots should have systems in place to prevent over-reliance;
- Automated chatbots should have systems in place to deal with emergency situations.

FUTURE RESEARCH DIRECTIONS

There are risks when using conversational assistants for mental health provision purposes (Bickmore et al., 2018), so more research is required into the design of conversational assistants for safety-critical dialogue that allows the flexibility and expressivity of natural language while ensuring the validity of any recommendations provided. Given the state-of-the-art in Natural Language Understanding (NLU), conversational assistants for health counselling should not be designed to use unconstrained natural language input, even if it is in response to a seemingly narrow prompt. Also, users should be advised

that medical recommendations from any non-authoritative source should be confirmed with health care professionals before they are acted on. More reviews are needed to summarise the evidence regarding the effectiveness and acceptability of chatbots in mental health as well as the impacts in the mid and long term of the usage of these technologies.

Chatbots, especially the ones that are text based, open new forms of communication through the use of emojis, avatars and other visual elements. What is the impact of the use of non-textual elements on chatbot users and how does that relate with patients that do not use chatbots as part of their therapy is as of now an open question and worthwhile of exploration.

Regarding usability best practices and evaluation, a lot has been done lately regarding mobile applications and web applications in the scope of e-health, however little to none was published under the scope of e-health chatbots. What kind of interaction works better, what approaches have the greatest impact for the users and how to optimize the communication between humans and chatbots are questions yet to be answered.

Ultimately, the great goal of any mental e-health tool, application or service is to promote the patient's recovery. Despite of the variety of tools that are emerging in this area, and all the new players involved, there is the lack of solid evidence that this goal is effectively achieved by any of the available services, especially chatbots, either used as a stand-alone alternative by the users or in tandem with the supervision of a healthcare professional. Even if these tools are capable of offering diagnoses and therapies, the capability of delivering a satisfactory personalized recovery plan for each user is yet to be assessed (Meadows 2020).

CONCLUSION

AI chatbots are programmed with therapeutic techniques to help people with mental health problems, but the promise of this technology is softened by concerns about the apps' efficacy, privacy, safety and security.

As AI techniques continue to be refined and improved, it will be possible to help mental health practitioners re-define mental illnesses more objectively than currently done in the DSM-5, identify these illnesses at an earlier or prodromal stage when interventions may be more effective, and personalize treatments based on an individual's unique characteristics. However, caution is necessary in order to avoid over-interpreting preliminary results, and more work is required to bridge the gap between AI in mental health research and clinical care. Conversational chatbots are seen across the literature as a low-cost and highly customizable, a technology capable of addressing an ever-growing need of mental healthcare services. However, at one and the same time, it is made clear that conversational agents should complement rather than replace traditional therapeutic options.

The Covid-19 pandemic produced a major impact on mental health services and chatbots could be useful in this matter (Miner, Laranjo, & Kocaballi, 2020).

Nevertheless, the field of AI has shown improvement in leaps and bounds in the last few decades. And one can only hope that mental health treatment protocols using AI in a growing digital world would be effectively bridging the gap between the patient and the treatment.

REFERENCES

Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A., Bewick, B., Gardner, P. H., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, 103978. doi:10.1016/j.ijmedinf.2019.103978 PMID:31622850

Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. *Journal of Medical Internet Research*, *23*(1), e17828. doi:10.2196/17828 PMID:33439133

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *22*(7), e16021. doi:10.2196/16021 PMID:32673216

Assal, H., & Chiasson, S. (2019, May). 'Think secure from the beginning' A Survey with Software Developers. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-13). ACM.

Bauer, M., Glenn, T., Monteith, S., Bauer, R., Whybrow, P. C., & Geddes, J. (2017). Ethical perspectives on recommending digital technology for patients with mental illness. *International Journal of Bipolar Disorders*, *5*(1), 1–14. doi:10.118640345-017-0073-9 PMID:28155206

Bickman, L. (2020). Improving Mental Health Services: A 50-Year Journey from Randomized Experiments to Artificial Intelligence and Precision Mental Health. *Administration and Policy in Mental Health*, *47*(5), 795–843. doi:10.100710488-020-01065-8 PMID:32715427

Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M., & Cruz, R. (2018). Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *Journal of Medical Internet Research*, *20*(9), e11510. doi:10.2196/11510 PMID:30181110

Bosek, P. (2019). *A Chatbot Maturity Model.* EasyDITA. Retrieved from https://easydita.com/a-chatbot-maturity-model/

California Consumer Privacy Act (CCPA). (2021, March 4). *State of California - Department of Justice - Office of the Attorney General*. https://oag.ca.gov/privacy/ccpa

Confessore, N. (2018, November 15). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Dosovitsky, G., Pineda, B. S., Jacobson, N. C., Chang, C., Escoredo, M., & Bunge, E. L. (2020). Artificial Intelligence Chatbot for Depression: Descriptive Study of Usage. *JMIR Formative Research*, *4*(11), e17065. doi:10.2196/17065 PMID:33185563

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2), e19. doi:10.2196/mental.7785 PMID:28588005

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2), e19. doi:10.2196/mental.7785 PMID:28588005

Gadiyar, A. (2020). The Chatbot Imperative: Intelligence, Personalization and Utilitarian Design. *Cognizant - Digital Business*. Retrieved from https://www.cognizant.com/whitepapers/the-chatbot-imperative-intelligence-personalization-and-utilitarian-design-codex2469.pdf

Garga, S. (2020). A Conversational UI Maturity Model: a guide to take your bot to the next level. *Medium*. Retrieved from https://chatbotslife.com/a-conversational-ui- maturity-model-a-guide-to-take-your-bot-to-the-next-level-4552d16724a2

General Data Protection Regulation (GDPR) – Official Legal Text. (2019, September 2). *General Data Protection Regulation (GDPR).* https://gdpr-info.eu/

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. *Current Psychiatry Reports*, *21*(11), 116. doi:10.100711920-019-1094-0 PMID:31701320

Gratzer, D., & Goldbloom, D. (2020). Therapy and E-therapy—Preparing Future Psychiatrists in the Era of Apps and Chatbots. *Academic Psychiatry*, *44*(2), 231–234. doi:10.100740596-019-01170-3 PMID:31898301

Grove, C. (2021). Co-developing a mental health and wellbeing Chatbot with and for young people. *Frontiers in Psychiatry*, *11*(606041). Advance online publication. doi:10.3389/fpsyt.2020.606041 PMID:33597898

Hern, A. (2017, February 21). Did your Adobe password leak? Now you and 150m others can check. *The Guardian*. Available at https://www.theguardian.com/technology /2013/nov/07/adobe-password-leak-can-check

Hill, K. (2014, October 6). "God View": Uber Allegedly Stalked Users For Party-Goers' Viewing Pleasure (Updated). *Forbes*. https://www.forbes.com/sites/ kashmirhill/2014/10/03/god-view-uber-allegedly-stalked-users-for-party-goers-viewing-pleasure/?sh=39989f431411

IBM Research Editorial Staff. (2017). *IBM 5 in 5: With AI, our words will be a window into our mental health*. IBM Research Blog. Retrieved from https://www.ibm.com/blogs /research/2017/01/ ibm-5-in-5-our-words-will-be-the-windows-to-our-mental-health/

Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, *6*(11), e12106. doi:10.2196/12106 PMID:30470676

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., ... Murray, C. J. L. (2018, November). GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, *392*(10159), 1789–1858. Advance online publication. doi:10.1016/S0140-6736(18)32279-7

Joerin, A., Rauws, M., & Ackerman, M. L. (2019). Psychological Artificial Intelligence Service, Tess: Delivering On-demand Support to Patients and Their Caregivers: Technical Report. *Cureus*, *11*(1), e3972. doi:10.7759/cureus.3972 PMID:30956924

Koyuncu, M., & Pusatli, T. (2019). Security awareness level of smartphone users: An exploratory case study. *Mobile Information Systems*.

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019). Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights*, *11*. Advance online publication. doi:10.1177/1178222619829083 PMID:30858710

Lovejoy, C. (2019). Technology and mental health: The role of artificial intelligence. *European Psychiatry*, *55*, 1–3. doi:10.1016/j.eurpsy.2018.08.004 PMID:30384105

Luxton, D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, *98*(4), 285–287. doi:10.2471/BLT.19.237636 PMID:32284654

Meadows, R., Hine, C., & Suddaby, E. (2020). Conversational agents and the making of mental health recovery. *Digital Health*, *6*. doi:10.1177/2055207620966170 PMID:33282335

Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. npj. *Digital Medicine*, *3*(1), 65. doi:10.103841746-020-0280-0 PMID:32377576

Miner, A. S., Shah, N., Bullock, K. D., Arnow, B. A., Bailenson, J., & Hancock, J. (2019). Key Considerations for Incorporating Conversational AI in Psychotherapy. *Frontiers in Psychiatry*, *10*, 746. doi:10.3389/fpsyt.2019.00746 PMID:31681047

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., ... UnÜtzer, J. Ü. (2018). The Lancet Commission on global mental health and sustainable development. *Lancet*, *392*(10157), 1553–1598. doi:10.1016/S0140-6736(18)31612-X PMID:30314863

Pontes, H. M. (2017). Investigating the differential effects of social networking site addiction and Internet gaming disorder on psychological health. *Journal of Behavioral Addictions*, *6*(4), 601–610. doi:10.1556/2006.6.2017.075 PMID:29130329

Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Khan, R. A. (2020). Healthcare Data Breaches: Insights and Implications. *Health Care*, *8*(2), 133. doi:10.3390/healthcare8020133 PMID:32414183

Shu, C. (2019, June 18). *Youper, a chatbot that helps users navigate their emotions, raises $3 million in seed funding*. Tech Crunch. Available at https://techcrunch.com/2019/06/18/youper-a-chatbot-that-helps-users-navigate-their-emotions-raises-3-million-in-seed-funding/

Singh, O. P. (2019). Chatbots in psychiatry: Can treatment gap be lessened for psychiatric disorders in India. *Indian Journal of Psychiatry*, *61*(3), 225. doi:10.4103/0019-5545.258323 PMID:31142896

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research*, *22*(3), e16235. doi:10.2196/16235 PMID:32141837

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Canadian Journal of Psychiatry*, *64*(7), 456–464. doi:10.1177/0706743719828977 PMID:30897957