AFRICAN JOURNAL OF BIOETHICS
ISSN: 2951-9853 | Volume 2, Issue 1 | Year: 2024
Published by the African Bioethics Network
Article Type: Review Article

(De) Stigmatising Mental Health: Could Chatbots Be the Solution?

Fikile Muriel Mnisi[1]
[1] *Tshwane University of Technology*

Corresponding Author Email: mnisimf@gmail.com

DOI; https://doi.org/10.58177/ajb230009

ABSTRACT

Recently, there has been an increase in the use of computer programs that use textual or vocal interfaces to replicate human communication, namely, chatbots. Chatbots have changed the provision of healthcare services, especially within the context of mental health; this field has seen an introduction of various chatbots designed to offer mental healthcare services that would be conventionally offered by humans. Despite the numerous advantages brought forth by chatbots within the context of mental health. However, the use of chatbots within the context of the provisioning of mental healthcare services raises certain ethical issues that have not been comprehensively addressed, especially those linked to the (de)stigmatisation of mental health disorders. Granted, these are not new ethical issues but, in this case, they may be exacerbated by many factors including the stigmas surrounding mental healthcare. Stigmas that often prevent individuals from seeking the necessary clinical help that they need. As a result, innovative solutions have had to be put in place in order to address some of these stigmas Therefore, the aim of this paper is to discuss and unpacks some of the ethical issues and considerations associated with the use of chatbots in mental healthcare and their potential impact on the (de)stigmatisation of mental health illnesses or disorders.

ARTICLE HISTORY
Submitted: November 16, 2023
Published: September 6, 2024

Introduction

In ancient medicine, known to us primarily through the ample corpus of Greek and Latin medical writing that extends from the earliest Hippocratic treaties to Byzantine medical complications, there is no established term for 'mental illness' as we understand it today. However, there are several diseases and kinds of illness recognised and described in ancient medicine which had prominent symptoms affecting, for example, mood, judgment, or memory, and we find reference to the mind, or the soul being affected by them (Ahonen, 2019). Moreover, at this point in human history, the modern conception of psychiatrists did not exist; this role was occupied by philosophers (e.g., Democritus, Socrates, etc.), who were considered, that is, doctors of the soul (psyche). Alternatively, from Democritus and Socrates onwards, as well as numerous other scholars' ancient philosophers were considered to be 'doctors of the soul,' that is, taking care of the ills of the soul in the same way medical doctors attended to ills of the body (Ahonen, 2019). Considering the rich and complex history outlined above, which is only a snippet of general human history, it is clear why mental illnesses are linked to deeply embedded stigmas and why mental healthcare, as it is presently practiced or may be understood to be practiced in modern days or times, is associated with ethical issues and dilemmas that require a lot of consideration.

Hem et al. (2018) highlights the many ethical dilemmas and challenges faced by professionals within the mental healthcare industry. Some of these challenges may contribute towards the stigmas associated with mental illness, particularly if they are not understood well and addressed in an 'appropriate' manner. There is somehow a link between these ethical challenges and the stigmas surrounding mental illnesses. For instant, some of the mental health challenges could arise in situations where a professional has to prioritise between patients or when trying to ensure cooperation between patients and their families, and where this priority could create an opportunity for the patient to be coerced into some type of therapeutic means or hospital administration (Ham, 2018). These ethical dilemmas and challenges may drive major changes in modern medicine, particularly concerning patient autonomy and/or the prioritisation thereof. Specifically, these challenges could result in healthcare systems in which patient's self-determining judgment prevails and the clinician primarily acts on a patient's wishes rather than solely relying on his or her judgement to make decisions that would be in the best interest of a patient (Igel and Lerner, 2016).

In my opinion, such a change could be problematic, particularly in countries or communities where individual autonomy is interlinked with their communities' autonomy and shared responsibilities are regarded as having an important moral value. Under such communities, the shared decision-making involves not only the patient and clinician, but also the patient's family or caregiver. This may exacerbate some of the ethical challenges that are already prevalent in mental healthcare services, with the doctor-patient relationship having significantly changed. In recent years, there has been an increase in the use of audio and textual chatbots (or digital intermediaries). This trend has further changed doctor-patient engagement and, therefore, their relationship, thus contributing to some of the ethical challenges in this discipline.

Chatbots, particularly those applied towards the provisioning of mental healthcare services, raise many ethical issues and concerns, many of which are not new but hold (among other things) an 'increase' in risk or harm. Whereas some of these ethical (and legal) concerns encompass user group suitability, data collection, security of data storage, privacy, and confidentiality, later linking of data, repurposing, as well as broader questions of accountability (Leins et al., 2019). It is clear from my perspective that as digitalisation continues to permeate the mental health industry, there will be an increase in the development and distribution of chatbot-based mental health applications (APPs). These APPs will greatly improve the ease at which mental health services can be accessed (e.g., this could benefit groups that may not normally afford to access such services).

Some of the questions underlying the use of chatbots in the mental health industry are as follows: How reliable will these chatbots be? How safe will their use be? What are the key ethical considerations that will need to be made? What impact will chatbot have on decision- making (i.e., ethical, and clinical)? How will they impact society (i.e., socially, legally, psychologically, culturally, etc.)? Answering these questions and maybe more of such questions, will assist in understanding the impact that these chatbots will have in (de)stigmatisation.It may be so that technological innovations (such as chatbots) tend to be accompanied by a variety of ethical concerns. However, such concerns are often less likely to be considered (if considered at all) compared to the technical challenges affecting the development and use of a

1 By mental health industry I will refer to everything that compass mental health such as: mental healthcare service, mental health systems, mental health research, mental health tools, etc.[2]

new innovation (Ruane, Birhane, and Ventresque, 2019). This is because ethical matters and processes are often seen as hindering and/or delaying innovation, even in instances when that is not the case. Developers need to equally consider technical and ethical issues associated with a certain innovation; this is especially important to ensure that the development does not negatively impact human rights and dignity. Specifically, if ethics are not carefully considered, this may result in unforeseen socioeconomic issues (clinical and moral, etc.).

Therefore, ethical considerations that encompass the development, implementation, and the use of an innovation are important for ensuring the success of an innovation. This paper will unpack and discuss the ethical issues and considerations concerning the use of chatbots for the provisioning of mental healthcare services. Specifically, these considerations will be studied within the context of their effect on the (de)stigmatisation of mental health illnesses.

OVERVIEW OF THE CHATBOT: HISTORY, CLASSIFICATION, AND CONCEPT

The year 2016 was dubbed the "Year of the Bot" after Microsoft Chief Executive Officer (CEO) Satya Nadella, who at the time described bots as the new "APPs" (Applications); this year saw the launch of more than 30,000 chatbots on the Facebook Messenger platform alone. By 2018, there were more than 30,000 active bots and 8 billion messages exchanged every month on the platform, with much of this growth driven by commercial interest. Chatbots are computer programs that respond as smart entities when conversed with via text or audio input. Chatbots understand one or more human languages via Natural Language Processing (NLP). Chatbots are also known as smart bots, interactive agents, digital assistants, or artificial conversation entities (Adamopoulou and Moussiades, 2020). In 1950 Alan Turing proposed the Turing Test, which tests whether a computer (or any other machine) has the ability to display human characteristics and intelligence (Birkenstock, n.d.; Adamopoulou and Moussiades, 2020). In other words, 'can the machine think?' It was at this time that the idea of a chatbot was popularised. The first chatbot was developed at MIT (Massachusetts Institute of Technology), by Joseph Weizenbaum; specifically, this chatbot, named, ELIZA was developed in 1966, and its primary role was to simulate a Rogerian psychotherapist by demonstrating "the response of a non-directional psychotherapist in an initial psychiatric interview". ELIZA used simple pattern matching and template-based response machine, thus, its conversational ability was not good. However, it was enough to confuse people at a time when they were not used to interacting with computers and give them the impetus to start developing other chatbots (Adamopoulou and Moussiades, 2020).

But ELIZA was plagued by certain limitation, and its only limitation, as noted by Joseph Weizenbaum, was that the "communication between man and the machine was superficial" (Lein et al. 2019). It was clear that ELIZA would need to be improved drastically for it to reliably and accurately mimic the qualities that drive human interactions. Since then, several chatbots have been developed to improve on ELIZA, with the chatbot ALICE having won the Loeber Prize, an annual Turing Test, in 1995 (Adamopoulou and Moussiades, 2020). ALICE was the very first computer to gain the rank of "most human computer". Overall, these timelines highlight the still-present complexity that is inherent in the development of chatbots. Building chatbots that come increasingly close to passing the Turing Test is a challenging task. Even though engineers have used the Turing Test to create better user experiences and drive significant value for a diverse range of companies and goals. Yet chatbots still have a long way to completely pass the Turic Test (Birkenstock, n.d.).

Lein et al. (2019) show that although chatbots have continued to be developed, advancing to the point where they facilitate human-machine interactions across a wide range of industries,

including the mental healthcare industry the term 'chatbot' has now come to encompass a much broader range of capabilities, including textual and voice recognition. Using the input processing and response generation method used by different chatbots, they can be classified as follows:

(a) Rule-based: This type uses some predefined rules or a decision tree to manage its response and dialog (Abd-Alrazaq et al., 2021); many of the earliest chatbots were built using this model, such as numerous online chatbots (Adamopoulou and Moussiades, 2020). Under this system, companies choose the responses of a chatbot to be based on a fixed, predefined set of rules which recognise the lexical form of the input text. The knowledge of such a chatbot is human-hand-coded, organised, and presented in a conversational pattern with a more comprehensive rule database, allowing the chatbot to reply to a wide range of user input (Abd-Alrazaq et al., 2021).

(b) Smart, also called artificial intelligence (AI): This chatbot is based on artificial intelligence; it collects information through conversations with customers and uses this information to improve its dialog (Sara 2021; Abd-Alrazaq et al., 2021). Through this technology, chatbots can learn and provide their own answers.

(c) Hybrid: This is a combination of the two previous systems. (Illescas-Manzano, 2021). An example of a hybrid system is outlined in a study by Bhattacharyya et al., (2021), which brings forth a driver model developed using a rule-based model and a data-driven model; these researchers conclude that while rule-based models provide interpretable driving behaviour, the data-driven parameter estimation provides fidelity to real-world driving demonstrations. This hybrid driver model approach is one of the few examples that represents the use of a rule-based model and a smart or AI model. Specifically, the rules under this approach include determining the colour of a traffic light and enforcing an appropriate response. However, this approach is not enough to adapt to the many dynamic patterns that characterise driving on a real-world road (e.g., drivers have different styles of driving), therefore, this is where the inclusion of the highly adaptable AI systems is important, and thus yielding a hybrid system. It is worth noting that such a system comes with its own set of challenges, howbeit, to improve a system's autonomous decision-making, a hybrid model is employed.

In addition to the classification criteria above, chatbots are further classified according to the permission provided by their development platform, which could either be open-source or proprietary code. An open-source platform provides the chatbot designer with the ability to intervene in most aspects of a chatbot's implementation while closed platforms typically act as "black boxes, which may be a significant disadvantage depending on a project's requirements. Therefore, based on the chosen type and classification, chatbots can further work using a variety of fundamental concepts, as shown in Table 1, which summarises the basic technologies that drive different chatbots' essential concept of the chatbot technology.

CONCEPT	EXPLANATION	EXAMPLE
Pattern Matching	Chatbots use pattern matches to cluster the text in order to find a related pattern and thereby generate an appropriate response for to a user's query (Sara, 2021). Thus, since this is predicated on the stimulus-response block, a sentence (stimuli) is entered, and an output (response) that is consistent with the input that the user has created.	ELIZA and ALICE
ARTIFICIAL INTELLIGENCE	AIML is based on a basic unit of dialogue called categories (tag categories›), which are formed by user input patterns (tag pattern›) and chatbot responses (tag template›). This	Aistifsar and MegaHAL

CONCEPT	EXPLANATION	EXAMPLE
MARKUP LANGUAGE (AIML)	approach applies a natural language modelling approach for characterising dialogue between humans and chatbots that follows a stimulus-response paradigm (Adamopoulou and Moussiades, 2020).	
LATENT SEMANTIC ANALYSIS (LSA)	LSA may be used together with AIML to develop chatbots; this approach is also used to find likeness between words as a vector representation (Adamopoulou and Moussiades, 2020). Furthermore, LSA answers an unanswered question.	Amazon5 and WienBot
CHATSCRIPT	Chatscript is the successor to the AIML language; it is an expert system. It consists of an open-source scripting language and an engine that runs it. It also consists of rules that are associated with particular topics and finds the best item that matches the user query string of a user, then executes a rule for that topic (Adamopoulou and Moussiades, 2020).	Outfit7's mobile app, Tom Loves Angela, and ESL chatbots
RIVERSCRIPT	It is a plain text, line-based scripting language for the development of chatbots and other conversational entities. It is an open-source with available interface for Go, Java, JavaScript, Perl, and Python (Adamopoulou and Moussiades, 2020).	Scarecrow, Admiral, I.R.I.S., and Alexa-Rivescript-Chatbot
Natural Language Processing (NLP)	NLP is an area of artificial intelligence that explores the manipulation of natural language text or speech by computers (Adamopoulou and Moussiades, 2020). Chatbots developed using this approach convert query text into structured data for entity recognition, sentiment analysis, and dependency passing (Sara, 2021). Thus, the knowledge of the understanding and use of human language is gathered to develop techniques that will optimise how computers understand and manipulate natural expressions to perform desired tasks. Most NLP techniques are based on machine learning (Adamopoulou and Moussiades, 2020).	Florence, Your.MD, Babylon Health, and Ada
Natural Language Understanding (NLU)	NLU is at the core of NLP tasks. It is a technique used to implement natural user interfaces, such as those for a chatbot. NLU aims to extract context and meanings from natural language user input, which may be unstructured, and respond appropriately according to user intention (Adamopoulou and Moussiades, 2020).	Amazon, Google, and Netflix – uses NLU software system

Table 1: Overview of the essential concepts of chatbots and examples

Therefore, being able to understand chatbots from a developmental level and how these chatbots could potentially drive innovation, it is crucial when developing a chatbot for a specific goals and purposes for a particular user(s). Thus, the growing global concerns about mental illnesses and the challenges of accessing mental healthcare services due to the shortage of mental healthcare providers, the lack of healthcare insurance coverage, and the perceived stigma linked with mental health illnesses may increase the risk of self-harm and suicide across the globe (Abd-Alrazaq et al., 2021). Therefore, chatbots may be a feasible solution to addressing these problems in the mental health industry, as seen within a "systematic review of 12 studies by Abd-Alrazaq et al., (2021). This review shows that chatbots are effective in improving some mental disorders, such as depression, stress, and acrophobia". Additionally, this study also reports that, not all mental health illnesses can be treated by using a chatbot. This issue highlights the following question: How effective will mental healthcare chatbots be?

Chatbot as a Tool for Mental Healthcare:

In mental healthcare, chatbots are most often a conduit to a real person, whom one can professionally consult (Lein et al., 2019). Additionally, chatbots are now slowly emerging as a viable complementary service that can act as to provide a person with assistance and offer some sort of companionship also known as 'virtual therapist' (Tewari et al., 2021). Over the years, numerous chatbots have been designed, developed, and applied to address a range of mental health issues. Examples of chatbots in the field of mental health include: Wysa: Anxiety, therapy chatbot; Mental Health App and Community; 7 Cups: Online Therapy for Anxiety and Stress; Mental Health Check Up; Woebot: self-care expert for depression and anxiety and one of the most popular chatbot for mental healthcare, SERMO: doctors' global social platform, etc, The growing number of these mental health chatbots is reflective of their ever-increasing roles in the context of mental health, as shown in Table 2.

APPLICATION USE OF CHATBOT USE CASES
Support for Cognitive or behavioural therapy; Example is the Woebot
Assessment of risk, suicide prevention; example is the Vickybot
Rendering social support for loneliness and isolation; Example is ChatPal
Coaching for behavioural and lifestyle change, self-help, and well-being; Example is Ginger
Providing digital counselling; Example is SmartBot360
Monitoring of mental health; Example is Wysa
Providing mental health information; Example is Wysa
Preventive mental health support for adolescent and young adults; Example is the University of Johannesburg (UJ) Chatbot known as Bolt-Impilo
Diagnostic screening for mental health; Example is Limbic

[3]

Table 2: The application of chatbots in mental health (Damij and Bhattacharya, 2022) These applications, as mentioned in Table 2, indicate how advantageous chatbots are or can be in the context of mental healthcare. Furthermore, these advantages are linked to the great accessibility that is exhibited by chatbot-based systems; such systems offer the following: immediate screening or routine check-ups; self-care; they may be free and/or affordable or cost-effective (i.e., some have in-app purchases); and they offer real-time feedback and a weekly summary that may assist individuals to gain an insight into their own patterns and continue data collection through diaries (Abd-Alrazaq et al., 2021). These could be of benefit in low- and middle-income countries due to the lack of resources and funding, particularly for mental healthcare services and the increase overburden by the number of individuals who may seek the care. This is

[3] Virtual therapies as suppose to a face-to-face therapy (therapist or psychologist) and/ or when one uses an online therapy (actual human being), and in the same way a virtual therapist will not include family and/ or friends of the patient. [I also want to note that it is also possible that chatbot can include an online therapist (human being)] However, it is advisable that these should be used and monitored by a mental healthcare professional, since these are only regarded as complimentary tools and not replacement tools for a mental healthcare professional.

supported by what Rathod et al. (2017) says in their report that, "globally, the expenditure on mental health is less than US$2 per year per capita across all countries and less than 25 cents in low-income countries. Many LMICs, including 15 of 19 African countries, allocate less than 1% of their health budgets to addressing mental illness". Furthermore, low-income countries tend to have a low mental health experts and providers as reported by the World Health Organisation (2018) that, "In low-income countries, the rate of mental health workers can be as low as 2 per 100 000 population, compared with more than 70 in high-income countries. This is in stark contrast with needs, given that 1 in every 10 person is estimated to need mental health care at any one time". Therefore, chatbots' advantages indicate how this technology can contribute particularly in low-income countries and improve quality of care and maybe also quality of life, respectively.

Ethical Matters Posed by Mental Healthcare Chatbot:

Despite the numerous benefits of chatbots in the mental healthcare industry, their use could be linked to a few issues. For instance, the operation of chatbots without users' knowledge raises multiple issues around consent and human dignity; and the collection of data from an individual that is potentially mentally ill and has a mental health problem, which may be used for certain purposes (i.e., commercial purposes), may raise ethical, legal, and social concerns and issues. The nature of chatbot interaction with certain individuals may give rise to potential risks and benefits. The information provided for privacy may be misunderstood or not fully comprehended and that type of 'deception' raises legal and ethical concerns regarding privacy and confidentiality, as well as concerns about human dignity. Ultimately, issues surrounding trust and safety of individuals while using these chatbots in the mental healthcare sector are of utmost importance, since privacy and confidentiality may not always be possible or could be limited. Such concerns could further exacerbate the stigmas surrounding mental health and cause more harm than anticipated.

It is therefore important to highlight the potential key ethical issues and concerns relating to chatbots in mental healthcare; additionally, it is also necessary to delve into the justifications for the use of chatbots, and outline what may ought to be considered 'proper' use within such a context. Analysing these ethical issues will allow for a thorough and comprehensive understanding of how chatbots may contribute towards the (de)stigmatisation of mental health illnesses.

The Principle of Privacy and Confidentiality in Chatbots:

The use of chatbots can be a discreet option for the provisioning of mental health support, and those who may be afraid of seeking treatment may find chatbots to be a suitable option, even if only temporarily. (This could mean that the individual may have some sort of awareness for their mental health problem and hence the use of a mental health chatbot). Moreover, those who may fear judgement linked to stigmatisation may also feel more comfortable communicating remotely with a chatbot (Sepahpour, 2020) in comparison to a face-to-face session. Howbeit, the use of chatbots in mental healthcare raises issues linked to privacy and confidentiality, and these issues are of paramount importance as the user may have to be aware of how privacy and confidentiality will be maintained. It may be easier to know the terms and conditions of maintaining privacy and confidentiality in face-to-face sessions than when using a digital therapy, i.e., a chatbot. As a result, it is important that users of chatbots understand how their privacy and confidentiality will be maintained and kept, and what could lead to a violation of this principle upon using the chatbot; this is crucial, as the violation of privacy and confidentiality could affect users' trust of chatbots, resulting in this tool falling into a grey area, making the

violation of privacy and confidentiality linked to the violation of trust, in which the use of chatbots as a health technology falls into a grey area (Sepahpour, 2020). Overall, these trust-linked issues need to be carefully considered, as it may severely cause a lot of harm to users, not only in the present but also in the future, resulting in a negative public perception of mental health chatbots.

Privacy is often tied to information shared by an individual (i.e., a user or patient), whereas confidentiality is linked to how a service provider or healthcare professional will secure the information. This requires trust, as the individual must trust that their information will be kept confidential and not shared with other parties without their consent. While there may be measures in place to ensure privacy and confidentiality, these measures are sometimes not enough, resulting in the data being lost through a leak or breach or hacking. Privacy and confidentiality are not always guaranteed even in traditional settings; how do we ensure that this is clearly conveyed to chatbot users? Additionally, how do we also keep ensuring that individual users are constantly aware of the terms and conditions associated with the use of chatbots within the context of mental healthcare? We have seen many of such cases; an example is Facebook, where the users' information has been used without their knowledge because of the terms and conditions where most of them unknowingly agreed to upon using the application. Even though such users did not fully read the terms and conditions (i.e., due to their great length) nor fully understand the terms and conditions. But because they agreed, the APP has a right to use their information based on what is outlined in its terms and conditions. The problem is that these terms and conditions are often too long, and users do not have the time to read them sufficiently to make an informed decision. Imagine a person who may be going through a mental health problem. Is this person expected to read and comprehend the terms and conditions associated with the use of a certain chatbot? Hence, it will be advisable that mental health chatbots be linked to a mental healthcare professional and used as his/ her prescription so they are able to monitor the process and as well as the type of follow-ups a particular individual will require. Additionally, if the chatbot service is suggested by such a person's healthcare provider, are they expected to 'blindly' agree to the chatbot's terms and conditions or is the provider supposed to go through these terms and conditions with the patient prior to their consent in using that chatbot?

It is not clear how chatbots should be applied within the context of mental healthcare while ensuring that the users' privacy and confidentiality is protected. However, answering some of these questions could greatly contribute towards guiding chatbot developers in the right direction. Who would be responsible for maintaining the confidentiality of users? Will it be the responsibility of the service provider or that of the healthcare provider, or both? Damij and Bhattacharya (2022, p.155) says that "lack of confidentiality obligations in commercial chatbots have been raised as a concern that demands government regulations and ethical guidelines to protect against the misuse of information, especially for mental health interventions such as suicide prevention situations". I agree with what Damij and Bhattacharya are saying and suggest that this be considered not only with suicide intervention and commercial chatbots but also with mental health chatbots in general. Furthermore, to consider methods and ways to safeguard the privacy of the user, as well as the overall security measures and issues around anonymity, Thus, the reduction and elimination of breaches of privacy and confidentiality are not only necessary

[4] See: Kozlowska, I. (2018). 'Facebook and Data Privacy in the Age of Cambridge Analytica'. From: https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/

in effecting competent therapeutic solutions, but also in upholding public (and user) trust in chatbot technology as an effective mental health intervention (Sepahpour, 2020).

Trust As an Ethical Concern Raised by The Principle of Privacy and Confidentiality:

Trust is a very important factor in chatbot development and usage; users and stakeholders should be able to have justified trust in the chatbot they use. Without trust, users will not fully incorporate chatbots in their daily routines (Srivastava et al., 2020). According to Pesonen (2021), people have the capacity to trust computers and assign them human characteristics. This phenomenon is similar to how individuals relate to television characters and how may come to associate a character's life and attributes with the actors. Pesonen (2021) further states that literature on the differences between human-human interactions and human-computer interactions, especially in the context of sensitive topics, shows mixed findings. They (2021) then define trust as the willingness of a truster to be vulnerable to a trustee's actions based on the expectation that the trustee will perform a particular action important to the truster, irrespective of the ability to monitor or control the trustee. The truster in this case would be the human user, and the trustee (i.e., or maybe better understood as trusted) will be the chatbot. Some studies, as indicated by Pesonen (2021), found that human users tend to be more open and self-disclosing when interacting with machines (chatbots) than with humans. Meaning that such individuals feel 'safe' without the fear of being judged or facing retaliation; this trend may greatly contribute towards combating stigmas surrounding mental illnesses. Howbeit, They (2021) further report on studies that show how 'chatbots' may be mistrusted, raising concerns about the increase in mental health stigmas that may arise from chatbots being mistrusted within the context of mental healthcare. They (2021) note that in certain studies there is evidence that users have voiced concerns regarding chatbots mishandling their sensitive data (i.e., privacy) and were afraid of possible leakage of data (i.e., confidentiality). This evidence indicative of the importance of prioritising trust when it comes to the usage of chatbots, especially in a mental healthcare context, where individuals are likely to be vulnerable. Trust is a concern, not only for breaching privacy and confidentiality but also for the overall experience of using the chatbot as a digital therapist.

Chatbots can become abusive, leak user information to other users and developers, and have incomprehensible conversational styles. On the other hand, users may want to use chatbots that are not bias, that do not use abusive language, and that do not leak information (Srivastava et al., 2020). In cases where an individual was getting help from a face-to-face healthcare provider, their information would remain private and confidential, albeit absolute privacy is not always guaranteed even under such scenarios. Howbeit, in the case of chatbots, breaches in privacy and confidentiality may be higher than in a face-to-face session. Apart from the issue of leaking information, there are also security issues linked to cybersecurity. How can chatbots prevent hacking as part of their security measures? In addition, as we move into more personalised medicine, there may be a need for developers to start personalising users' information, which means that users may have to continue sharing information of ever-increasing sensitivity.

Therefore, trust needs to be considered when designing and developing chatbots, with factors to consider to when aiming to improve trust including robustness, reliability, transparency, explainability, and fairness (Srivastava et al., 2020). For example, trust can be violated if: conversations were recorded and then shared; if personal information that the user offered to the chatbot was collected and stored for distribution; or if a user becomes confused about whether they are indeed speaking with a chatbot or not, as opposed to a human agent in a time of absolute distress. Even if the conversation is not recorded for the purpose of either teaching the chatbot or incorporating it into edited conversations for the purpose of supervised learning

or personalising them for the user's needs, other violations of trust may occur (Sepahpour, 2020). Even though there may be measures put in place to try and maintain the safety of users' while they use the chatbot, these measures could also result in a violation of trust. Overall, a balance must be struck to maintain users' trust and their safety while using the chatbot. This is going to be vital, considering that studies conducted in South Africa indicate that 70– 84% of patients with different ailments have consulted traditional healers (THs) at some point. The South African Stress and Health study (SASH) reported that only 5.7% of persons with a mental illness had received any conventional mental health care in the preceding 12 months. In contrast, 5.8% of the general population had accessed a TH or alternative/ complementary medicine practitioner, and a further 6.6% had used services from social or religion-based practitioners during the preceding 12 months (Zingela, van Wyk, and Pietesen, 2019). In addition, many South Africans have always exhibited a level of mistrust in "Western" medicines, driving patients to seek religious help or consult traditional healers. Therefore, finding ways to enhance the level of trust in innovations such as chatbots, will be of paramount in countries such as South Africa.

There are protective measures that can be implemented within a chatbot's structure to protect the user, but these measures may negatively impact trust. One protection against harm that might occur in moments of crisis is to switch the user over to the human agent who can make an informed and rapid decision regarding resources to offer the user in distress. In this case, the user may feel deceived, which will result in confusion; the user may feel as if their entire digital therapy session had been monitored or that they had been speaking to a human agent the entire time. Under such a protective measure, if a user shares information with a chatbot only because they feel more comfortable talking with a chatbot than a human agent, the switch to a human agent in a moment of absolute crisis could be unwelcomed and a violation of trust (Sepahpour, 2020), as well as the perceived safety. This may be considered a typical method of violating privacy and confidentiality while trying to safeguard the human user. Therefore, to safeguard against such a violation of trust, it would be necessary to notify the user when this switch is about to occur (Sepahpour, 2020) and possibly give them an option to consent the switch and further assistance. However, feature should form the basis of the privacy settings that drive chatbots; and chatbots should be designed in such a way that this message is not just a tick in a box. In addition, chatbots should also have systems that ensure that users have a clear understanding of the terms and conditions associated with the use of a specific chatbot. For instance, there could be messages that are often sent through the chat regarding these terms and conditions or

a specific one at a time as a reminder. Consequently, legal provisions will be necessary for safeguarding chatbot users, to ensure that their safety and security concerns are adhered to; such a measure would need to be implemented in addition to clinical safety, where chatbots must indicate, through the intervention offered, how they would work through a justifiable clinical trial and the evidence from those clinical trials should be used to approve the chatbot for clinical use (Sepahpour, 2020). Maybe further studies into the perceived safety of chatbots while they are being used by individuals afflicted by mental health illnesses could provide valuable insights into the value of the using multiple perspectives (i.e., clinical and legal) to when addressing safety and security concerns are not only taken from a legal and clinical perspective. But also, from a societal or attitude and behavioural perspectives, even though there is evidence

[5]Trust can be impacted by a variety of factors, such as language, the chatbot's perceived caring and comfort, the chatbot's 'gender', cultural differences, etc. (Pesonen, 2021).
In my mind, this was seen during the COVID-19 vaccine hesitation; some of this hesitation

of feeling secure and safe while using chatbots. One thing is for sure: we must take all necessary precautions before we can be ethically confident of the safe use of chatbots.

Therefore, careful, and continuous considerations will be necessary when it comes to using chatbots in mental healthcare; these considerations should include their impact on the principle of autonomy, including their application through the processes of informed consent (i.e., for decision-making), privacy, and confidentiality, and how this can further contribute to the (de)stigmatisation of people living and/ or diagnosed with mental illness, potentially exacerbating ethical issues already-plaguing the mental healthcare sector.

Considerations of Stigmatisation in Using Mental Health Chatbots:

In professional health and social work roles, a fiduciary duty, with both ethical and legal dimensions, binds practitioners to act in the interests of their patients or clients. This includes duty to access the benefits and understand the risks of certain interventions and to carefully weigh them out with the aim of promoting the client's well-being (i.e., within the limits of respect for personal autonomy) (Lein et al., 2019). However, where mental healthcare services are concerned, it may be challenging to weigh out what may be in the best interest of one's patient, and this can be exacerbated by the stigmas surrounding mental health illness. These stigmas are difficult to address, which makes it challenging to weigh up the benefits and risks involved and provide the necessary care when it comes to mental healthcare services. Moreover, these stigmas often make it 'difficult' for individuals who may be suffering from mental health illnesses to seek the necessary help. While in other instances it could make it difficult for healthcare professionals to offer a dignified service due to these stigmas.

It is important, particularly in the context of mental healthcare services, to address stigmas surrounding mental health illness, to reduce some of the biases that may hinder individuals from

[6]seeking the necessary help or intervention. This is where chatbots can help alleviate the impact of these stigmas by greatly improving the ease at which mental healthcare services can be accessed. Research has shown that embarrassing situations or situations that carry stigma may

facilitate a greater acceptance of chatbots in the broader community; specifically, users might be as likely, if not more likely, to disclose emotional and personal information to a chatbot than they are to a human agent (Lein et al., 2019). Although this might raise some psychological concerns, which could trigger a variety of social issues, as the use of chatbots has the potential to cause harm. These issues may be compounded by a scenario in which a severely mentally ill person (i.e., in need of human intervention) may become reliant on a chatbot, to the detriment of their health.

[6] In my mind, this was seen during the COVID-19 vaccine hesitation; some of this hesitation was driven by citizens believing that the vaccines were a means of inserting a 5G-based chip within them, with this chip being used to monitor and control them. Apart from that, there are many instances, throughout (South Africa's) history, that have broken the trust of citizens in the medical fraternity.
Some of the legal provisions to safeguard privacy and confidentiality issues include the Protection of Personal Information (POIP) Act (South Africa), General Data Protection Regulation (GDPR) (Europe), and HIPAA (Health Insurance Portability and Accountability Act) (USA).

Stigmatisation And Its Relation to The Application of Chatbots in Mental Healthcare:

How could chatbots (de)stigmatise mental health illnesses? Being able to find a comprehensive answer to this question may facilitate an optimal application of chatbots not only as a tool to alleviate the societal impact of mental health illnesses, but also as a tool to resolve those stigmas that still surround such illnesses (i.e., since chatbots are perceived as less stigmatising). People with serious mental illnesses are doubly challenged. On the one hand, they struggle with the symptoms and disabilities that result from their illness. On the other hand, they are challenged by the misconception-driven stereotypes and prejudices about mental illness. Due to these challenges, people with mental illness are robbed of the opportunities that define a decent quality of life: good career opportunities, accessing of jobs, safe housing, satisfactory healthcare, and an affiliation with a diverse group of people (Corrigan and Watson, 2002). Overall, these stereotypes and prejudices in mental health have led to the stigmatisation of mental health and illnesses. Therefore, the application of an innovation (i.e., chatbots) that could potentially exacerbate the stigmatisation already experienced by people that are mentally ill could further lower their quality of life. Thus, stigmatisation can be understood to mean that someone views another person in a negative way because that person has a distinguishing characteristic or personal trait that is thought to be or is a disadvantage (e.g., a negative stereotype) (Mayo Clinic, n.d.). Although when looking at this definition one can ask how a chatbot can view a user in a negative way since it is not a person in a sense of a having personhood, and thus cannot contribute (truly in this sense) into causing stigma. Howbeit, there is a way in which mental health chatbots can contribute to the stigmas around mental health, thereby either increasing them or reducing them and it is vital to figure out how this can be so.

In the case of mental health, stigmatisation would be anything to do with another's mental health illness(es), the complexity of (sometimes) being able to relate to them or form relationships with them, their perception of the illness(es) or disorder(s), etc. These could be anything that may be viewed as being different and 'negative' and does not necessarily have to be a physical difference (as with mental health) but can also be a psychological or behavioural ones. Corrigan and Watson (2002) explain that stigmatisation is twofold; it includes public- and self-stigmatisation.

7

Public-stigmatisation characterises the reaction that the general population has towards people with mental illness, whereas self-stigmatisation is the prejudice that people diagnosed with mental illness have against themselves. Public- and self-stigmatisation can be further understood in terms of three components: stereotype, prejudice, and discrimination. Due to the stigma surrounding

mental health, chatbots can be used to offer an alternative means to address some of the stereotypes, prejudices, and discrimination concerning both public- and self-stigmatisation.

[7] Personhood as defined by the Oxford Reference (www.oxfordreference.com) is a philosophical concept designed to determine which individuals have human rights and responsibilities. Personhood may be distinguished by possession of defining characteristics, such as consciousness and rationality, or terms of relationships with other. In my understanding and in a nutshell, personhood speaks of intrinsic and extrinsic values. For further reading on AI and Personhood see a paper by Robert K. Garcia, titled Artificial Intelligence and Personhood, published in 2002, in Cutting Edge Bioethics: A Christian Exploration of Technologies and Trends.

Stigma is one of the fundamental reasons why many individuals do not seek help, and sometimes this stigma is exacerbated by the complications and issues that are based on certain traditional, cultural, and/ or religious perceptions of a particular country. In my opinion, chatbots could be used to manage and monitor some of these issues and assist in creating an environment to also address them. Specifically, the development, promotion, and the use of chatbots could grant individuals access to mental healthcare services via a platform that is neutral enough to induce shame or fear of being stigmatised. Howbeit, despite being perceived as less stigmatising, chatbots might pose a threat to users due to their limited capacity to re-create human interactions and to provide tailored treatment, especially when it comes to interacting with people when they are at their most vulnerable state. Subsequently, could issues surrounding a chatbot's programming, or language, or accessibility, or evaluation and monitoring, exacerbate an increase in the stigmatisation already associated with mental health issues? Presently, ongoing evaluations of harm and benefit, which are essential for ethical and responsible practice, are absent in many digital platforms and APPs used for the provisioning of mental healthcare services. Yet, such evaluations are important, they will allow for a deeper understanding of the role of chatbots in the (de)stigmatisation of mental health issues.

Conversely, due to the dual effect of chatbots, they have the potential to cause harm, particularly to individuals who may suffer from self-stigmatisation; when faced with issues beyond the scope of current digital means, such individuals may avoid human intervention, opting for a heavy reliance on a chatbot,, to the detriment of their mental health; instead of going to seek human intervention at a point that they need or should, they may hide behind the chatbot's interventions. Additionally, in a case where the chatbot is built in such a way that it can suggest human intervention, the user may still refuse help, which is their right, but as a result, it would become even more challenging for them to be channelled towards the most suitable assistance. Furthermore, some of these chatbot-reliant individuals may be driven by self-discrimination, denial, and fear of rejection by others, thus contributing to their inability to seek appropriate clinical and psychological interventions. Research suggests that self-stigma and the fear of rejection by others lead many people to not pursue life opportunities for themselves (Corrigan and Watson, 2002).

On the other hand, if an individual that had been using a chatbot for their mental health issues does indeed go and seek the necessary help as suggested by the chatbot, this could (eventually) lead to a form of public stigmatisation . The consequences for the individual seeking the

appropriate help could result in them being avoided (ostracised) by either them close friends and family members and/or their communities, and further result in them possibly having work-related problems leading them to losing their job. Albeit this may often be caused by several other factors related to mental illness. In line with these possibilities, Corrigan and Watson (2002) state that, prejudice caused by public stigma leading to fear may result in avoidance. For example, employers tend to avoid individuals with mental illness, so they try not to hire them for many reasons, such as missing work because of the illness. Subsequently, if a mentally ill individual seeks professional help, they may, in hindsight, have to deal with the stigma that

[8] Of course, I am aware that this may not always be the case, but since there are negative stereotypes, resulting in many individuals not seeking the necessary help [I am speaking based on experiences and observations I have made as a South African]. Howbeit, there is a possibility that these chatbots may produce positive results, via chatbot screening and diagnostic efforts.

comes with having to be or having been diagnosed with a mental illness or disorder. Either way, they are doomed. The question is, will they be prepared for all these potential consequences? Therefore, how can we then ensure that chatbots applied within the context of mental health are equipped to arm individuals with insights intended to prepare them for such outcomes?? Overall, if chatbot users seek human intervention, as suggested by a chatbot, they will have to seek the help fully informed of all possible consequences, including those linked to stigmatisation. On the other hand, chatbots could facilitate authoritarianism, depending on the service provided, which could lead to users of chatbots being coerced into taking treatment or being institutionalised.

Therefore, it is crucial to carefully evaluate and understand the impact and outcomes (be they benefits or risks) that a mental health chatbot may pose prior to use. Moreover, it is important to understand the role of autonomy, with its relationship to trust and safety, as well as decision-making while using a chatbot as a mental healthcare tool. The next section will unpack the subject of autonomy as an ethical factor to be considered in the process of (de)stigmatising mental health through the use chatbots. The subject of autonomy plays a crucial role in the evaluation of ethical considerations and evaluations, especially when one is also examining stigma as an ethical barrier.

Ethical Factors Associated with Mental Health Stigmas

The Role of Autonomy When Evaluating Mental Health Stigma:

Respect for autonomy implies acknowledging that autonomous agents are entitled to hold their own viewpoints, are free to make choices, and act voluntarily according to their values, beliefs and preferences (Motloba and Makwakwa, 2018). This is often challenging in the space of mental health as the individual who may be struggling with mental health problems may not always be able to make an informed and voluntary choice or decision. This is further exacerbated by self-stigma and / or public-stigma that the individual may likely be exposed to. Furthermore, for autonomy to be realised these individuals will have to be able to consent. Motloba and Makwakwa (2018) write that, according to the Oxford Dictionary, consent means "permission for something to happen or agreement to do something". Seeking informed consent means that the doctor requests authorisation for a medical procedure that the patient fully understands and agrees with. Anything beyond the contracted agreement is a violation of informed consent. Informed consent is a core cornerstone of ethics in human subject research (Motloba and Makwakwa, 2018).

Through the informed consent process, participants learn about the study procedure, benefits, risks, and more to make an informed decision. However, recent studies showed that current practices might lead to uninformed decisions and expose participants to unknown risks (Xiao et al., 2023). Unkown risk in this instance such as; how using a particular chatbot could lead being stigmatised. There is currently no guarantee even with reports that chatbots are perceived as less stigmatising. Should these individuals base their decision on a perception or should the decision be based on concrete facts? I know that I would not allow myself or any of my loved one to use a chatbot if there are chances that they may (even by default) end up with self-stigma and/ or public-stigma due to the use of that chatbot. I assert that stigmatisation is a consequence that many would like to avoid and would not consent to. Therefore, the process of an informed consent can be sensitive particularly for those struggling with mental health as they are vulnerable and may feel judged or unduly influenced to make a decision that they may feel is not entirely what they would have decided. Even if the individual went to seek for help voluntarily on their own, being able to consent may still be challenging as they may not entirely be aware of what their consenting too, especially with the use of chatbot. Yet, chatbots will still be required

to ensure that these 'traditional' understanding (also legal requirement) are adhered too. But how would this procedure be maintained with chatbots?

Broadly speaking, autonomy includes (i) independence of thought, inclusive of the ability to "think for oneself", make decisions, determine preferences and moral assessment for oneself; (ii) autonomy of will or intention which is regarded as the ability of a moral agent to decide on his/ her plans of actions and activities; and (iii) lastly, autonomy of action, which involves doing what the agent thinks and intends or wills to do (Motloba and Makwakwa, 2018). Respecting autonomy means the mental healthcare professional acknowledges the agency of the patient to exercise the right over all mental healthcare processes undertaken on them, without undue interference or influence from the attending healthcare professional (Motloba and Makwakwa, 2018). Then how can we ensure that the individual is able to exercise their right without any undue influence while using the chatbot? This will require continuous monitoring and evaluation of these chatbots and to ensure that algorithms are not set so as to cause undue influence or coercions. Moreover, some of the influences may be caused by biases. Sometimes, fear of being discriminated against and the individual chooses to use a Chatbot even when this is not in their best interest. Moreover, how will the process of informed consent be designed? Will these chatbots have their own informed consent procedures? Now imagine again a scenario where the chatbot has to switch from digital therapist to face-to-face and how the user of the chatbot will feel 'disrespected' due to the bridge of privacy, confidentiality, and trust. This will not only cause emotional and moral harm, but also the individual may suddenly feel a threat of stigma through shame as the chatbot moves from the digital therapist to face-to-face.

As result of the challenges of the process for respecting autonomy through an informed consent, certain individuals wanting to use a chatbot for mental health may portray some form of stigma due to certain reasons that may hinder them from giving a fully informed consent. Thus, the process of informed consent will need to be carefully developed for chatbots in mental health and well managed to ensure that indirect and/ or unforeseen stigmas are addressed. This will also help to identify some of the chatbot biases and discrimination and how they could pose harm and contribute to stigmatisation.

Bias and Discrimination as Ethical Factor in (de)stigmatization:

Due to chatbot bias, certain groups of people could end up not receiving the help they require because they belong to a particular group that may be deemed to portray some form of corresponding stigma (Corrigan and Watson, 2002). Thus, Bias is another issue that may cause further stigmatisation with the use of chatbot. Bias can result in unfair treatment for certain groups compared to others (Srivastava et al., 2020). Therefore, existing racial, gender, and classism biases (i.e., together with the 90/10 rule in research), may intensify biases exhibited by chatbots. These biases could then perpetuate stigmatisation. Of course, such biases originate from the input data used to develop a specific chatbot, resulting in a variation in the biases exhibited by different chatbots. and this may differ from one place or society to another depending on the data used. However, this is an important ethical factor to take into consideration when analysing the relationship between stigma and chatbots.

While discussing bias especially as one of the factors to be taken into consideration in understanding (de)stigimatisation is its relation to discrimination. Although, discrimination is one of the themes that contributes to the misconception about mental illness and often result in stigmatisation (Corrigan and Watson, 2002). Discrimination can appear in public opinion about how to treat people with mental illness. For example, studies have been unable to demonstrate the effectiveness of mandatory treatment, more than 40% of the 1996 GSS sample agreed that

people with schizophrenia should be forced into treatment. While, the public endorses segregation in institutions as the best service for people with serious psychiatric disorders (Corrigan and Watson, 2002). This may have changed over the years but it gives some idea of the type of data that could be used to develop a chatbot, data that may be based on individual 'false' perceptions since they do not fully comprehend mental health illnesses. Thus, further contributing to increasing the stigmas related to mental illnesses through bias and discrimination points of view. Discrimination is an important factor that could also result from the biases that a chatbot may contribute too. Therefore, an important element to consider on how discrimination may be caused by these biases and further perpetuate to unfairness and the limitation to accessing mental healthcare services.

Accessibility As an Ethical Factor For (De)stigmatisation:

Ideally, all those who seek mental healthcare services and support should be able to access a highly trained mental healthcare provider, although studies show that this ideal scenario is far from our reality. To take seriously the reality of the world, in which significant injustice and imperfection exist, we must entertain solutions to the barriers we face in this arena (Sepahpour, 2020). Since chatbots are widely accessible to anyone with a smartphone and internet access and may be perceived as less stigmatising than formal mental health support (Kretzschmar et al., 2019), they should be considered as a serious option for granting a wide range of people access to mental healthcare services. Moreover, because not everyone can access a mental healthcare provider, chatbots are likely to be an improvement to our world by offering aid where we lack it (Sepahpour, 2020). Thereby, bridging the gap where there are limited resources within the healthcare system. Hence, Kretzschmar et al., (2019) claims that, chatbots can become the first step towards getting help. I agree with this statement, as not only chatbots can be the solution to the shortage of mental healthcare providers, but they can also bring dignity to many individuals who, in many ways, could not manage to have access to mental health services or support; such individuals include people, like those in rural communities and/or remote areas, and hopefully make this the first point of reference when seeking mental healthcare support. Thus, chatbots may be used as one of the mental healthcare resources to greatly improve people's access to healthcare services, resulting in wide-scale improvements in the quality of care and life, and finding solutions that can assist governments to increase and offer better quality healthcare and resources, thereby, maintaining citizens claimed rights and dignity.

Final Thoughts on the Use of Chatbots as a Tool for (De)Stigmatisation:

[9]

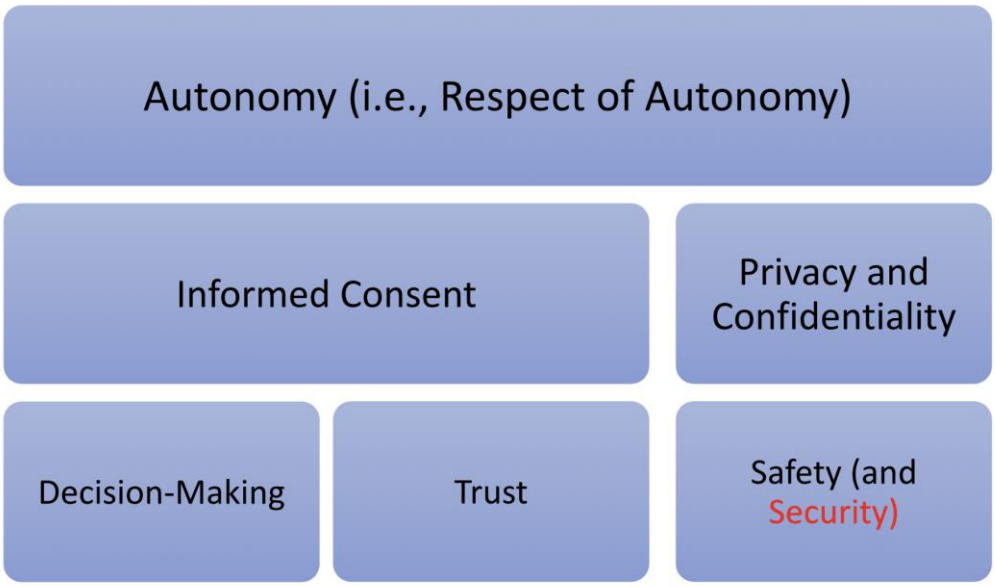
When an individual with an ongoing mental illness continues to obtain mental health services via a chatbot, potentially due to the fear of stigmatisation, their condition may continue to worsen.

On the other hand, other chatbot users may end up being misdiagnosed due to a lack of 'enough' data or "biased" data being used. This could be data from the user and/ or algorithm data used to assist the chatbot in decision making. This is even more relevant for countries where there is a low literacy level or English is not the first language. The use of chatbots in such contexts for the provisioning of mental health services could intensify stigmas surrounding mental health illnesses, causing more harm than good. This possibility is supported by Lein et al., (2019), who shows (i.e., through a survey) that the current mental health APP landscape, of which chatbots

[9] For example, vulnerability due to mental health problems, illiteracy, English being a second language, inability to comprehend the information to give consent, economic vulnerability, etc.

form a significant part, tends to over-medicalise states of distress and may over-emphasise 'individual responsibility for mental well-being'. Claims that users suffer from 'poor mental health literacy', which poses a barrier to the increased use of chatbots, therefore, should be regarded with caution (Lein et al., 2019). These issues may exacerbate the stigma surrounding mental health illnesses. Therefore, it is crucial that, prior to the wide-scale use of chatbots, certain stigmas are dealt with and the relationship between these stigmas and chatbots are well understood. This may be possible if these stigmas are also regarded as ethical dilemmas that result in social issues, and not seen as separate matter. Therefore, the only way to achieve this goal is to prioritise the consideration of ethical dilemmas in the design and development mental health chatbots; additionally, 'traditional' (i.e., autonomy, privacy and confidentiality, trust, etc.) ethical issues and their potential role in the (de)stigmatisation of mental health illnesses should also be considered.

Key ethical concerns linked to autonomy, decision-making, privacy, and confidentiality (leading to a breach of trust, safety, and security in the technology as well as in the mental healthcare fraternity) need to be considered when designing and developing mental health chatbots and channelling them towards the provisioning of a specific mental healthcare service. Graph 1 summarises the ethical concerns (i.e., and their relationship) discussed in this paper. These issues have the potential to intensify the stigma surrounding mental health, making it more challenging to address them. Therefore, critical assessments of how these chatbots either increase or decrease stigmas are going to be crucial for their overall ethical assessment. The stigmas surrounding mental illness should be considered as part of an ethical exercise aimed at evaluating the harms or risks versus benefits of using chatbots within the context of mental health; such an approach would allow for the systematic reduction of harm, making effective mental health treatments accessible to more people.



Graph 1: Summary of key ethical concerned associated with the use of chatbots in the context of mental healthcare with a potential to contribute to mental health stigmas.

Howbeit, chatbots do pose certain ethical challenges (some are more technical and/ or social), which will need to be prioritised and worked upon alongside other ethical issues in order to make the use of chatbots effective and ethically justifiable. Most importantly, alongside efforts to improve digital mental health resources, it is extremely important that we continue advocating

for research funding and professional services aimed at lessening the stigmas surrounding mental illnesses (Kretzschmar et al., 2019). Consequently, if these ethical concerns are not addressed and incorporated into the chatbot lifecycle, from design to development to clinical trials to market and usage, they could translate to chatbots causing more harm than good, thus, exacerbating the stigmas associated with mental health.

Conclusion

To conclude, it is crucial to consider the use of chatbots for the provisioning of mental healthcare services and for improving quality of care. Although chatbots may offer many great benefits there are some downsides that need to be taken seriously, especially if we are to fight against the stigmas surrounding mental illness. However, whether mental healthcare chatbots should be used or not will depend on the overall careful ethical evaluation of these chatbots. Their contributions to the overall health of the patients and the lessening of stigma and discrimination associated with mental health illnesses. Conversely, increased care is required in mental healthcare since individuals may be susceptible to vulnerability, and this vulnerability can also 'increase' with chatbot use. Additionally, apart from the technical aspects of the chatbot, what is equally important will be the consideration of ethical challenges and judgments to justify the permission and use of that chatbot for mental healthcare. This will require an overall evaluation of beneficence versus non-maleficence against justice, human rights, and autonomy. Autonomy may not only include (possible) patient autonomy, but considerations should be made on the effect or impact of chatbots on patient autonomy versus family or caregiver autonomy, especially in a shared decision-making model. Yet these ethical issues may also influence how chatbots may contribute to increasing or decreasing mental health stigma, and ultimately stigma should be considered as one of the ethical issues to be evaluated when making an ethical analysis on mental health chatbots.

With that said, mental health chatbots indicate the potential to mitigate stigma. However, careful ethical (clinical and legal) evaluations will be necessary if chatbots are permitted as one of the tools for mental healthcare. These ethical issues and dilemmas do not mean that chatbots cannot be implemented as one of the tools for mental healthcare and that they do not have the potential to reduce some of the stigma. Despite these ethical issues what ought to be prioritised is breaking the stigmas surrounding mental health illness, and hopefully being able to use chatbots to also realise this priority.

Acknowledgments

The author(s) would like to acknowledge the team at the Inkanyeti Foundation for their support in the work of raising awareness, educating, and promoting mental health. The author(s) would like to further acknowledge Dr Janeen Prinsloo and Mr Keith. M. Dube for their comments and edits of the paper.

Conflict Of Interest

The author(s) declare no conflict of interest.

Funding

The author(s) discloses that the was no receipt of funding for this project.

Ethics

The study did not require any ethical clearance as there were no human subjects involved in the study.

Reference

1. Abd-Alrazaq, A.A., Alajlani, M., Ali, N., Denecke, K., Bewick, B.M., and Househ, M. (2021). "Perception and Opinions of Patients about Mental Health Chatbots: Scoping Review". J. Med. Internet Res. Vol 23(1): pp. 1-15. Accessed Date: 3 January 2023. From: http://www.jmir.org/2021/1/e17828/
2. Adomopoulou, E., and Moussiades, L. (2020). "An Overview of Chatbot Technology". IFIP AICT. Vol 584: pp. 373-383. Accessed Date: 3 January 2023. From: https://doi.org/10.1007/978-3-030-49186-4_31
3. Ahomen, M. (2019). "Ancient Philosophy on Mental Illness". History of Psychiatry. Vol 30 (1): pp. 3-18. Accessed Date: 20 November 2021. From: https://doi.org/10.1177/0957154X18803508
4. Bhattacharyya, R., Jung, S., Kruse, L., Senanayake, R., and Kochenderfer, M. J. (2021). "A Hybrid Rule-Based and Data-Driven Approach to Driver Modelling through Particle Filtering". pp. 1-13. Accessed Date: 31 January 2023. From: https://arxiv.org/pdf/2108.12820.pdf
5. Birkenstock, R. (n.d.). "Chatbot Technology: Past, Present, and Future". Accessed Date: 11 July 2022. From: https://www.toptal.com/insights/innovation/chatbot-technology-past-present-future
6. Corrigan, P.W., and Watson, A.C. (2002). "Understanding the Impact of Stigma on People with Mental Illness". Accessed Date: 2 December 2021. From: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489832/pdf/wpa010016.pdf
7. Damij, N., and Bhattacharya, S. (2022). "The Role of AI Chatbots in Mental Health Related Public Service in a (Post) Pandemic World: A Review and Future Research Agenda". IEEC-TEMSCO EUROPE. ISBN: 9781665483, 9781665483131. pp. 152-159. Accessed Date: 3 January 2023. From: https://doi.org/10.1109/TEMSCONEUROPE54743.2022.9801962
8. Hem, M.H., Molewijl, B., Gjerberg, E., Lillemoen, L., and Pedersen, R. (2018). "The Significance of Ethics Reflection Groups in Mental Health Care Professionals". BMC Medical Ethics. Vol 19(54): pp. 1-14. Accessed Date: 30 Nov 2021. From: https://doi.org/10.1186/512910-018-0297-y
9. Igel, L.H., and Lerner, B. (2016). "Moving Past Individual and Pure Autonomy: The Rise of Family-Centred Patient Care". American Medical Association Journal of Ethics. Vol. 18(1): pp. 56-62. Accessed Date: 02 Dec 2022. From: https://www.amajournalofethics,org
10. Illescas-Manzano, M.D., López, N.V., González, N.A., and Rodríguez, C.C. (2021). "Implementation of Chatbot in Online Commerce, and Open Innovation". J. Open Technol. Mark. Complex. Innov. Vol. 7 (125): pp. 1-20. Accessed Date: 11 July 2022. From: https://www.mdpi.com/journal/joitmc
11. Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., and NeurOx Young People's Advisory Group. (2019). "Can Your Phone Be Your Therapist? Young People's Ethical Perspective on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support". Biomedical Information Insight. Vol 11: pp. 1-9. Accessed Date: 3 January 2023. DOI: 10.1177/1178222619829083. From: https://pubmed.ncbi.nlm.nih.gov/30858710/
12. Lein, K., Cheong. M., Coghlan, S., D'Alfonso, S., Gooding, P., Loderman. R., and Paterson, J. (2019). "To Chat, or Bot to Chat, Just the First Question: Potential Legal and Ethical Issues Arising from a Chatbot Case Study". Accessed Date: 02 Dec 2022. From: https://www.researchgate.net/publication/344637804_To_Chat_or_Bot_to_Chat_Just_the_First_Question_Potential_legal_and_ethical_issues_arising_from_a_chatbot_case_study/link/5f8818f5299bf1b53e28f274/download

13. Mayo Staff Clinic. (n.d.). "Mental Health: Overcoming the Stigma of Mental Illness". Mayo Clinic. Accessed Date: 3 Feb 2022. From: www.mayoclinic.org
14. Motloba, P.D., and Makwakwa, N.L. (2018). "Respecting Autonomy (Part 2)" SADJ. Vol.73(7): pp. 460 – 462. Accessed Date: 26 Dec 2023. From: http://dx.doi.org/10.17159/2519-0105/2018/v73no7a7.
15. Pesonen, J.A. 2021. "'Are You Okay?' Students' Trust in a Chatbot Providing Support Opportunities". HCII 2021, LNCS 12785: pp. 199-215. Accessed Date: 3 January 2023. From: https://doi.org/10.1007/978-3-030-77943-6_13
16. Rathod, S., Pinninti, N., Irfan, M., Gorczynski, P., Rathod, P., Gega, L., and Naeem, F. (2017). "Mental Health Service Provision in Low and Middle – Income Countries. Health Service Insight. Vol 10: pp. 1-7. Accssed Date: 24 Dec 2023. DOI: 10.1177/1178632917694350. From: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5398308/#:-:text=Globally%2C%20the%20expenditure%20on%20mental,budgets%20to%20addressing%20mental%20illness.
17. Ruane, E., Birhare, A., and Ventresque, A. (2019). "Conversational AI: Social and Ethical Consideration". CEUR-WS. Vol 2563: no page. Accessed Date: 30 Nov 2021. From: www.CEUR-WS.org/Vol-2563/aics_12.pdf
18. Sara. (2021) "Chatbot Development: Step-By-Step Guide in 2022". Accessed Date: 11 July 2022. From: https://www.notifyvisitors.com/blog/chatbot-development/
19. Sepahpour, T. (2020). "Ethical Considerations of Chatbot Use for Mental Support". DISSERTATION, John Hopkins, Baltimore, Maryland. pp. 1-33. Accessed Date: 3 January 2023. From: https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/63294/SEPAHPOUR-THESIS-2020.pdf?sequence=1
20. Srivastava, B., Rossi, F., Usmani, S., and Bernogozzi, M. (2020). "Personalized Chatbot Trustworthiness Ratings". 2020-01-0004-OA10-TTS: pp. 1-9. Accessed Date: 3 January 2023. From: https://arxiv.org/abs/2005.10067
21. Tewari, A., Khalsa, A.S., and Kanal, H. (2021). "A Survey of Mental Health Chatbot Using NLP". ICICC-2021: pp. 1-6. Accessed Date: 3 January 2023. From: https://ssrn.com/abstract=3833914
22. World Health Organization. (2018). "Mental Health: massive Scale-up of Resources Needed if Global Targets Are to Be Met". No page. Accessed Date: 24 Dec 2023. From: https://www.who.int/news/item/06-06-2018-mental-health-massive-scale-up-of-resources-needed-if-global-targets-are-to-be-met.
23. Xiao, Z., Li, T. W., Karchalias, K., and Sundarum, H. (2023). "Inform the Uninformed: Improving Online Informed Consent Reading with an AI-Powered Chatbot". CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Article No 12: pp 1-17. Accessed Date: 26 Dec 2023. From: https://doi.org/10.1145/3544548.3581252.
24. Zingela, Z., van Wyk, S., and Pietersen, J. (2019). "Use of Traditional and Alternative Healers by Psychiatric Patients: A Descriptive Study in Urban South Africa". Transcultural Psychiatry. Vol 56 (1): pp. 146-166. Accessed Date: 6 March 2023. From: https://journals.sagepub.com/doi/pdf/10.1177/1363461518794516