Erasmus University Rotterdam

# MASTER THESIS

Artificial Intelligence in Mental Health

How (potential) users make sense of AI driven mental health chatbot?

*Considering the welfare state and the public health care system.*

**Master Thesis**

Name: Maya Janecke

E-Mail: 669414mj@eur.nl

Date: 06-08-2023

Supervisor: Dr. Roy Kemmers

Second Evaluator: Dr. Bonnie French

# *Abstract*

*Mental health care is facing a global crisis, with a surge in demand for services coupled with limited resources and trained healthcare providers. Artificial Intelligence (AI) and Machine Learning (ML) have led to the development of AI-driven mental health chatbots, aimed at providing accessible and immediate support to individuals in need. This qualitative study aims to understand individuals' perspectives on AI-driven mental health chatbots and their perception of these technologies within the larger societal and healthcare context. Three distinct perspectives emerged: The Visionary, who sees chatbots as an innovative solution to address mental health challenges and emphasizes the potential benefits of privatization and market-driven approaches; The Pragmatic, who advocates for cooperation between the state and private companies, with regulatory standards to ensure ethical implementation; and the Skeptic, who expresses doubts about the effectiveness and ethical considerations of chatbots, preferring stronger state involvement in regulation and public healthcare expansion. The study highlights the need for comprehensive and inclusive approaches to mental health care, considering the welfare regime's orientation and its impact on healthcare systems. Addressing ethical concerns and promoting responsible implementation of AI-driven chatbots can enhance mental health care services while safeguarding individual well-being. Understanding users' perspectives is crucial for developing guidelines and policies that maximize the benefits of these technologies, especially for marginalized groups.*



Keywords: Mental Healthcare – AI driven Mental Health Chatbot – Welfare State – Public Healthcare System - Neoliberalism

**Table of concent**

**List of tables**



**List of abbreviations**

| | |
|---|---|
| AI | Artificial intelligence |
| e. g. | exempli gratia; for example |
| GDPR | General Data Protection Regulation |
| GT | Grounded Theory |
| ML | Machine Learning |
| NLP | Natural Language Processing |
| WHO | World Health Organization |

# 1. Introduction

Mental health care is facing a significant crisis worldwide, with an alarming increase in demand for services together with a shortage of resources and trained healthcare providers. Common mental health disorders, including depression, anxiety, and stress-related conditions, have become pervasive, affecting millions of individuals globally. The World Health Organization (WHO) reports concerning statistics that underscore the urgent need for improved mental health care services.

The numbers provided by the World Health Organization (WHO) emphasize the pressing urgency to address the mental health care crisis. Due to the Covid-19 Pandemic, there is an increase of 25% of the prevalence of anxiety and depression. Conditions like depression, anxiety, and stress impact a substantial portion of the global population, causing immense personal suffering and hindering overall well-being (WHO, 2017). The extent of these common mental health disorders demands a comprehensive response from healthcare systems.

However, the provision of effective mental health care is hindered by several critical barriers. One significant challenge is the lack of adequately trained healthcare providers to meet the growing demand for mental health services (Prakash, 2020). Other factors such as attached stigma, geographic boundaries, and long waiting lists often deter individuals from seeking timely help (Tewari et al., 2021; Prakash et al, 2020). Moreover, the ratio of mental health care providers to patients is significantly imbalanced, with a large gap in the availability of cost-effective and convenient services (see, for example, Tewari et al., 2021; Vallés-Peris et al., 2021; Kretzschmer et al., 2019).

Recent breakthroughs in Artificial Intelligence (AI) and Machine Learning (ML) have expanded automation beyond traditional tasks, creating opportunities to revolutionize knowledge-based care work, including mental health support (Vallés - Peris et al., 2021).
Technology companies have developed mental health chatbots, to provide innovative solutions for mental well-being. These AI-driven chatbots, aim to act virtual assistants, have the primary objective of offering immediate 24/7 support and ensuring easy accessibility to enhance the reach of therapy services (Tewari et al., 2021). However, these technologies come with significant risks, such as limitations in their functionality and ethical considerations related to privacy, transparency, and accountability. To address this, there are different approaches taken by different welfare regimes depending on their level of neoliberal ideologies. Research suggests that economic policies, particularly neoliberal ideologies that emphasize individualism and materialism, have contributed to the rise in mental illness and distress (Zeira, 2021). These

policies have exacerbated wealth inequality, limited access to mental health care, and caused emotional distress. Paradoxically, however, these same free market strategies are being relied upon to solve the problem of health care. Care used to be a task for the government/state, but because of the neoliberal trend of the past decades, it is increasingly becoming the responsibility of the individual to look after themselves (Zeira, 2021). There are many contrasting views on how to deal with the ethical risks and safe use of AI chatbots, and how to ensure a safe implementation in the healthcare system that does not hinder further innovation. The exploration and positioning in this regard depend on the orientation of the welfare states, which may either be more in favor of neoliberal developments or less.

The existing literature indicates that AI-driven mental health chatbots have been well-received by users, who appreciate their convenience and accessibility. Studies have highlighted the potential benefits of chatbots in mental health care, including personalized treatments and crisis care. (Hague & Rubya, 2023; Abd- Alrazag, 20; Kretzschmer et al., 2019). Research for AI in medical use has been made, focusing on the accuracy and efficiency of the technologies in clinical context (Lee et al., 2021; Graham et al., 2019). However, ethical considerations related to data security, transparency, and accountability have been raised as significant concerns (Carr, 2020; Sepaphor, 2020). Efforts to address these ethical concerns through regulations, such as the EU's AI Act, are underway. As these are highly debated, it is essential to locate viewpoints within the context of the system of welfare regime and its public healthcare.. The research objective of this study is to understand perspectives of individuals to gain a deeper understanding of the underlying attitudes, beliefs, and experiences with chatbots, considering the larger societal and healthcare context. By examining these perceptions, this research will provide valuable insights for policy development and innovation in mental health care and seeks to contribute to policy recommendations that enhance mental health care services, ensuring equitable access to support and addressing the challenges posed by the mental health crisis. First, I will introduce the chatbots in the mental health care system and then contextualize them within the public healthcare system of different welfare states. Next, I will address possible ethical considerations and attempt to take them into account to ensure a safer usage. Finally, I will outline the perception of individuals regarding these aspects and draw conclusions from it.

# 2. Contextual and Theoretical Framework

## 2.1. Artificial Intelligence driven Mental Health Chatbots

Within the rapidly evolving landscape of Artificial Intelligence, AI-driven chatbots emerge as a transformative embodiment of AI, capable of perceiving, interpreting, and learning from data to fulfill complex objectives in the digital realm. There are many different definitions for Artificial Intelligence. To keep it simple, I adopt the definition of the European Commission 2019.

> *"Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving and learning from their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal." (*European Commission, 2019)

As Artificial Intelligence is already widely used in modern western life such as customer service and other operating systems, it is not yet commonly used and implemented in the Public Healthcare System (Graham et al., 2019). AI adoption in mental health has been slower compared to other sectors. This may be because mental health practitioners rely on "softer" skills and subjective patient data. Still, AI holds potential to redefine mental illness diagnosis by having access to large amount of data, provide a 24/7 therapeutic approach, and develop better screening and risk prediction tools for personalized mental healthcare using big data (Graham, et al., 2019). In this context, AI driven chatbots has gained significant attention which is seen as a relevant tool for the next generation of psychological interventions (Tewari et. al, 2021).

A chatbot in general is not quite new: The first chatbot ELIZA was invented in 1964-66, which simulated the role of a psychotherapist to interact with users which became surprisingly successful. ELIZA pathed the way for further development in the field of human-computer interaction (Zemcík, 2019). Nowadays, there are at least more than 40 mental health chatbots and probably 1000 – 2000 apps that supply support for any health-related life situation. Most of them are implemented in the United States, but their usage is steadily increasing globally (Abd- Alrazag et al., 2019). AI driven chatbots nowadays have a certain level of artificial intelligence using Machine Learning (ML) and Natural Language Processing (NLP). This enables them to analyze user input, determine relations to understand and interpret NL statement to provide appropriate responses and support (Hussain et al., 2019). Some of the most

well-known AI driven mental health chatbots are e.g. Wysa, Woebot, and Emma. They are designed to provide cognitive behavior therapy support (Wysa, Woebot) or help managing diseases and medications (Emma) (Denecke et. at., 2021). They are developed for the target group of young adolescents dealing with mild depression, anxiety, and/or stress (Kretzschmer et al., 2019). By providing immediate and discreet support, these chatbots reduce stigmatization barriers and encourage self-reliance on mental healthcare support (Spephour, 2020; Nardanzynski et al., 2019).

Despite the potential of such systems, their practical application in clinical settings is infrequent due to several existing challenges (Yan et al., 2020; Prakash, 2020). AI driven mental health chatbots still face challenges in fully understanding conversational context and emotional cues compared to humans (Caldarini et al., 2022; Kretzschmer et al., 2019). The relationship between a human therapist and the patient is highly valuable and has proven successful in psychotherapy historically and in recent telehealth practices (Sepaphor, 2020). Mental disorders like depression are highly subjective, with intricate symptoms, individual variations, and significant socio-cultural influences, necessitating comprehensive consideration for accurate diagnosis (Yan et al., 2023).

Additionally, many chatbots are being developed without adequate consideration of ethical concerns (Omarov et al., 2023). They arise, particularly, regarding data privacy, transparency, patient autonomy, and algorithmic biases (Matheny et al., 2019). Furthermore, there are numerous issues with current sample data, such as artificiality, limited ecological validity, small sample sizes, and obligatory oversimplification of categories (Yan et al., 2023)

When integrating these chatbots into the public healthcare system of western welfare regimes, it is crucial to consider ethical implications to ensure safe and responsible use. These ethical considerations are influenced by the effects of neoliberal development, which in turn impact the accessibility, quality, and ethical aspects of mental health care services. As we delve deeper, we will explore the diverse welfare regimes and examine how neoliberal development shapes the landscape of public healthcare.

## 2.2 Welfare Regimes & Public Healthcare

According to Vallés-Peris et al. (2021), an approach to patients' perceptions involves considering patients not only in relation to their health or disease but also in connection with the entire healthcare system and the governing neoliberal regimes. AI technology is not simply an isolated artifact but a complex network consisting of devices, processes, and actors that are

interconnected. Technological innovations are conglomerates of various material, social, and semiotic relations, wherein technical, scientific, political, economic, and ethical considerations are complexly interrelated within a heterogeneous network. Interactions with AI technology occur within concrete healthcare systems, influenced by organizational forms, power structures, values, symbolic elements, and other contextual factors, highlighting the contextual nature of technological artifacts (Vallés- Peris et al.,2021).

McGregor (2001) describes health care policy as an integral part of social policy, aimed at protecting and enhancing human life and dignity. Traditionally funded by taxes, social policies encompass decisions affecting the cost, delivery, quality, accessibility, and evaluation of health care programs. The health status of a nation is often reflective of its health care policy, which involves considerations of hospitals, health care professionals, and public expenditure (McGregor, 2001). Neoliberal ideology advocates for the superiority of private markets, promoting the idea that the state is inefficient, and that privatization is more cost-effective, meaning that the supply and demand dynamics of the free market lead to an ideal distribution of resources throughout society (Card & Hepburn, 2022; McGregor, 2001). Card & Hepburn (2023) argue that the neoliberal ideology is "linked to poorer collective health and well-being" (p. 363). A significant aspect of neoliberal policies is the replacement of the concepts of the public good and community with individual responsibility. This shift is exemplified through the favoring of deregulation and privatization of public and state-owned enterprises, including education, health, and infrastructure (Card & Hepburn, 2022). The focus on individualism and the free market through privatization and deregulation comes at the expense of social solidarity, promoting competition, accountability, and consumer demand in the marketplace (McGregor, 2001). Consequently, there has been a reduction in social spending, including healthcare, in favor of an unregulated market and a minimal welfare state (Card & Hepburn, 2022). Decentralization is also a key element of neoliberal policies. Advocates argue that the market represents the ultimate form of decentralization, offering consumers a choice of suppliers (McGregor, 2001).

The neoliberal development cannot be understood without embedding it into the concept of the welfare state. It aimed to boost private capital profitability and played a crucial role in consolidating an accumulation regime centered on mass production and consumption. Embedded liberalism involves a welfare state where the government intervenes in the economy to ensure a minimum income for all, and essential services based on needs. The welfare state's core idea is to grant social rights to citizens, allowing them to be less dependent on the market for their well-being (Konuralp & Bicer, 2021).

In addition to this ideal concept, welfare states also vary in their approach to the relationship between the state, market, and the family. The institutional framework significantly influences institutions and their welfare policies as well as the people that live in such systems (Larsen, 2007). Using Esping- Andersen's (1990) well- known broad differentiation of liberal, conservative-corporatist, and social democratic welfare regimes, we can gain a deeper understanding of the underlying factors that contribute to the variety of viewpoints and stances concerning the distribution of welfare benefits where the provision of mental health care is located.

The liberal welfare regime, as exemplified by the United States, is characterized by its high level of selectivism, wherein benefits and services are provided selectively based on assessments of neediness and its extent (Larsen, 2007). Such a system has a low level of *decommodification*[1] and places a strong emphasis on individuality and the primacy of markets (Esping-Andersen, 1990).

Conservative-corporatist welfare regimes in countries like Austria, France, and Germany exhibit moderate decommodification (Esping-Andersen, 1990). Social rights are linked to status and class, with the state replacing the market as a welfare provider. However, the redistributive impact is constrained due to the emphasis on maintaining status, and traditional family structures influence these regimes (Esping-Andersen, 1990).

Social-democratic welfare states, predominant in Scandinavian countries, have a high level of decommodification and universalism (Larsen, 2007). Social reforms aim for high equality standards, extending social welfare benefits to the middle class too (Esping-Andersen 1990). These regimes prioritize the welfare state over the market, fostering universal solidarity.

The neoliberal development affected all types of welfare regimes as it resulted in reduced shares of national incomes and limited economic influence for them. This resulted in threats to security due to high costs under neoliberal policies. Konuralp and Bicer (2021) found difficulties in obtaining care and fears of declining quality were evident particularly for those countries with a liberal welfare system. This had a significant impact on the public health system. In healthcare, a neoliberal approach involving anti-health reforms, budget cuts, deregulation, and privatization was implemented cost-cutting in health spendings hit low- and middle-income families, leading to financial disasters. Neoliberal reforms limiting public healthcare have negatively affected population health in all welfare regimes (Konuralp & Bicer, 2021).

[1] „a process that reduced exposure to the market forces of price, supply and demand without necessarily removing it entirely" (Peredo & McLean, 2019 :822)

The focus shifted to viewing public services, including healthcare, as products for sale in the private sector, leading citizens to become customers purchasing these services. The shift towards a market-driven health system, treating services as commodities (McGregor, 2001), emphasizes a profit-oriented approach to mental health care provision, potentially sidelining the overall well-being aspect. Neoliberal policies justified this shift by presenting state-run systems as inefficient and promoting the for-profit market as more cost-effective and consumer-friendly (McGregor 2001). Efficiency through cost-cutting became a central aspect of this neoliberal perspective (Konuralp & Bicer, 2021).

Based on this assumption, the transformation of health care into a market-driven system, where services are seen as commodities, led to scarce resources for health care provided by the welfare regime. It can widen inequities and promote further socio-economic segregation. Those, who can afford private health care services benefit the most, while those with limited financial means may face barriers to accessing quality care (McGregor, 2001). This shows a reinforcing effect as the research by Allen et al. (2014), revealed that common mental disorders such as anxiety and depression are distributed based on economic differences. It results in a high demand for mental health care, while at the same time the provision faces limitations and is not capable to cover the need. In mental healthcare it results in long waiting list, difficult access, and/ or low-quality services (Dayioglu, 2022).

Addressing these challenges, governments invested high in amounts of money to increase the development of innovative solutions such as AI technology (Nedayvoda, 2021). Private companies invented AI driven mental health chat bots to substitute the lack of human care and to reduce the burdens of traditional mental health care support.

Despite the promise of chatbots, their implementation as new and unregulated technology comes with several challenges. Addressing ethical issues is crucial if chatbots are to effectively meet the growing demand for mental health services. To gain insight into the driving forces behind governments and their connection to the population's well-being, we examine the current AI regulations in different countries and their specific focuses.

## 2.3 Governmental Guidance & Ethical Considerations

Given the fast development of Artificial Intelligence, there have been intense discussions about the types of policies and governance that would enable the development of an ethical and trustworthy AI. To address ethical considerations there are different AI regulatory approaches. In 2021 the European Union recently came up with a proposal for the ‘Artificial Intelligence

Act'. This is proposed European Law to prevent risks and negative effects of the development of AI without hindering economic benefits and technological innovation. According to the most recent publication from June 2023, the calls for regulation have recently increased. Whereas the USA had taken "a lenient approach towards AI" (European Commission, 2023:2), the UK is currently working "on a set of pro-innovation regulatory principles" (European Commission, 2023:2) and The Cyberspace Administration of China is also working on a proposal which regulates AI. In the context of the recently established EU-US tech partnership (The Trade and Technology Council), the EU and the USA are seeking to develop a similar understanding on the principles underlining trustworthy and responsible AI, including advanced AI system such as ChatGPT (European Commission, 2023).

Relevant ethical considerations for this study on AI driven mental health chatbots include concerns of technical robustness and safety, transparency, data security and privacy, diversity and discrimination, accessibility, and accountability. These also apply to AI driven mental health chatbots. To prevent unintentional harm, the accuracy, resilience, security and reliability is crucial for AI systems (European Comission, 2019).

Transparency on the data, the system and the AI business model is essential for trusting, accepting, and adopting AI driven mental health chatbots. This enables humans to be aware of what they are using, including the capabilities and limitations these bots are providing (European Comission, 2021). AI technology is often seen as a 'black box' (Wirtz et al., 2022) which makes the algorithms mysterious and incomprehensible to citizens. This is one of the main risks, as stated by Yu and Alí (2019), that "AI systems can escape the control and understanding of their operators and programmers" (p. 2). This is referred to the problem of AI, in which decisions are made that cannot be traced back to humans entirely or only partially (Wirtz et al., 2022).

Transparency and awareness of these risks is crucial since users of bots often cannot distinguish them from human agents. Lack of transparency can be deceptive and compromise public trust, which is crucial for effective public health interventions. If chatbots are employed for healthcare purposes, it is essential to disclose that they are not human, promoting user autonomy. Transparency provides users with the necessary information to make informed decisions about using chatbot interventions (Sepahpour, 2020).

Human bias refers to the unconscious or conscious preferences, stereotypes, prejudices, and judgments that individuals hold toward certain groups or individuals based on their characteristics such as race, gender, ethnicity, age, sexual orientation, or other personal attributes. For instance, in 2016 Microsoft created a Twitter chatbot to learn from users that

quickly became racist, sexist, and harmful to other users. This incident occurred because the chatbot was trained on data from numerous racist users on Twitter (Victor, 2016). The potential negative implications include marginalization of vulnerable groups and the aggravation of prejudice and discrimination (European Commission, 2019). If the chatbot is trained on biased data that perpetuates stereotypes or stigmatizes certain mental health conditions, it can inadvertently reinforce harmful societal beliefs, further stigmatizing individuals seeking support (Siau & Wang, 2020). Biased training data can lead the chatbot to provide inaccurate or harmful advice to users. For example, it might recommend ineffective or potentially harmful treatments or suggest that certain mental health issues are not significant.

AI-driven chatbots for mental health have access to vast amounts of data, including sensitive and private information (Boucher et al., 2021). The increasing generation of data in society and business raises the risk of misuse (Siau & Wang, 2020). Health data can cause significant harm if not properly protected, leading to irreversible and long-term effects on individuals' lives and social environment (Van Kolfschooten, 2021). The growing interest of Big Data companies in health data due to its scarcity and expensive collection process poses threats to privacy. The use of AI-driven mental health chatbots may encourage users to provide more health data, potentially benefiting these companies financially (Van Kolfschooten, 2021). Additionally, passively collected data, gathered without users' control, raises concerns about privacy and data protection as users may be unaware of the specific types and volume of data being collected from them.

AI-driven mental health chatbots offer a promising approach to mental health assistance, as they are widely accessible and less stigmatizing than official mental health services (Kretzschmer et al., 2019). They can enhance doctor-patient communication, increase healthcare accessibility, and manage the rising demand for teleconsultations and medication monitoring. However, there are challenges related to the inclusion of elderly and handicapped citizens, which can impact overall well-being (Siau & Wang, 2020). According to Morley et al. (2002), it is crucial to prioritize equity and inclusion when implementing AI in healthcare to avoid disparities in health outcomes and ensure fair distribution of technology benefits.

Accountability for AI systems and their outcomes is essential to ensure responsibility. Auditability, allowing the assessment of algorithms, data, and design processes, is crucial, especially in critical applications, such as AI driven chatbots as they deal with vulnerable people who may be in serious or dangerous situations. Determining responsibility when the AI mental health chatbots fails or causes harm, often referred to as the "problem of many hands", can be challenging (Siau & Wang, 2020, p. 70). These consequences of AI mental health chatbots may

result from programming codes, entered data, improper operation, or other factors. It raises the question of who should be held responsible—the programmer, the data as it is still an ongoing discourse how ethical considerations are addressed in the context of AI in the mental health sector, it is useful to ask those who are expected to use these new tools (in the future). How do (potential) user deal with ethical challenges and how do they perceive the chatbot in the context of the welfare state and public healthcare system? Addressing this question sheds light on their ideological views on technological innovation (AI), their believe in the mental health care system and the need for governmental regulation to protect them.

# 3. Methodology

The aim of this is study is to examine how (potential) users make sense of AI driven mental health chatbots to gain a deeper understanding of the nature and meaning of AI driven mental health chatbots. Additionally, potential strategic utilization of these technologies by health care institutions (such as health professionals, insurance companies, government, hospitals etc.) is explored. This involves investigating the expectations of (potential) users, their intended use of these technologies, and their perceptions of associated risks and limitations, in consideration of the role of the welfare state regarding increasing neoliberal ideology.

This study employs a qualitative methodology to conduct empirical data which enables the acquisition of meaningful and in-depth insights and ideas regarding the perceptions and viewpoints of various individuals (Boeije, 2010). Qualitative research provides a profound understanding of individuals' attitudes, beliefs, motives, and behaviors, including their emotions, perceptions, and actions when dealing with a medical condition. It explores people's perspectives and experiences, seeking insights and identifying social structures or processes that clarify the significance of their actions (May & Pope, 2000; Wong, 2008).

## 3.1. Data Collection and Data Analysis

Due to the explorative approach for this study empirical data was collected to gain in-depth insights to answer the research question. Therefore, 13 qualitative semi- structured interviews were conducted, containing mainly open questions. This makes it possible for individuals to express their opinions in an environment where they feel comfortable doing so (Boeije, 2010) to address sensitive topics like mental disorders and personal therapy.

Interview participants were recruited through a sampling approach that involved both purposive and convenient methods. Purposive sampling describes the purposively selection of participants, that meet certain criteria that can answer the research question (Chun Tie et al., 2019). Convenient sampling describes the selection of participants due to convenient criteria, such as accessibility, availability, or the willingness to participate (Etikan et al., 2016).

Following purposive sampling, (potential) users of AI driven mental health chatbots were identified as those who have either (according to one's own statement) symptoms of a mild depression, mild anxiety, or mild stress or had previous personal therapy for the reasons mentioned above. Since this contains very sensitive information about individuals, a clear distinction was made to avoid causing harm or triggering emotions related to severe mental diagnoses that are beyond my capabilities to deal with professionally. As the respondents were also chosen through personal contacts, convenient sampling was also employed.

The conducted interviews specifically aimed to cover views and perspectives on AI driven mental health chatbots in the context of the public healthcare system regarding ethical implications as well as individual and social benefits and risks of implementing this technology. As one of the advantages of semi-structured interview is flexibility and the option to adjust questions (Harvey- Jordan & Long, 2001), depending on the interviewee's knowledge questions were left out or added to gain as much information for this study as possible. At any time, the interviewee was given the opportunity to skip questions or end the interview. Half of the interviewees chose to have the interview via a private telephone call, while the other interviewees preferred to meet in person either in a café or a private setting. The interviews were conducted either in English or in German with a length between 30 and 50 minutes.

To recruit the interviewees, I looked for diversity (which was admittedly challenging, given the time and scope for this study). The interviewees were mainly contacted through social media and other informal networks. These interviewees are not representative of all (potential) users of AI driven mental health chatbots since they are selected by who accepted an offer to participate and are thus more inclined to speak out and be aware of either AI driven mental health chatbots or mild mental disorders such as depression, anxiety, or stress in general. In fact, all participants were in the age range between 20 and 30 years, suggesting this study does not represent views and perspectives of other generations. In the context of this study, it is especially meaningful to investigate in young people's perceptions and views. Young people, being the largest consumers of the digital world, are a significant target group for digital interventions (Kretzschmer et al., 2019). Kretzschmer et al. (2019) found that they frequently seek health information online, particularly related to mental health issues like stress, anxiety,

and depression. Online content influences their health behaviors, indicating the potential impact of digital mental health interventions on this demographic (Kretzschmer et al., 2019). Therefore, young people's firsthand experiences with mental health chatbots are believed to be valuable in shaping the normative discourse surrounding AI in mental healthcare.

The interviewees all originated from western countries, except of one American person, all from countries in Europe namely Belgium, Germany, Bulgaria, Spain, and Switzerland. They currently live in 3 different countries: The Netherlands, Germany, and Spain. Table 1 provides an overview of the background information of the interviewees.

*Table 1 Demographic data of the interviewees*

| Interviewees | Age | Gender | Nationality | Residency |
|---|---|---|---|---|
| 1 | 25 | F | RU | NL |
| 2 | 25 | F | BEL | NL |
| 3 | 28 | F | GER | GER |
| 4 | 26 | F | GER | GER |
| 5 | 30 | F | GER | GER |
| 6 | 25 | M | GER | NL |
| 7 | 29 | F | US | NL |
| 8 | 29 | M | GER | GER |
| 9 | 27 | M | GER | GER |
| 10 | 28 | M | ESP/GER | ESP |
| 11 | 26 | M | GER | NL |
| 12 | 26 | M | GER | GER |
| 13 | 26 | M | CHE | NL |

To analyze the data, after conducting the interviews, the grounded theory approach was used. The process of interview transcription offers an initial opportunity to first engage with the data and identify key themes (Braun & Clarke, 2006). The transcribed data was then uploaded to the qualitative data analysis tool Atlas.ti. Using the traditional Grounded Theory approach to build a conceptual framework through inductive data analysis, various stages of coding were applied (Appendix B). The Traditional GT aims to produce a conceptual theory that explains a relevant and troubling pattern of behavior for people concerned (Chun Tie et al., 2019).

Starting with the initial coding, first codes were created to then sort them into categories. The next round of coding consisted of the intermediate coding, where core categories were selected. To build the storyline, advanced coding was used until theoretical saturation was reached (Chun Tie et al., 2019). These steps are carried out alongside constant comparative analysis for coding and category development, including writing memos to ensure the quality of the analysis and as comprehensive records capturing the researchers' reflections, emotions, and intuitive (Chun Tie et al., 2019).

### 3.2. Ethical Considerations

A particular focus was placed on protecting the collected data in order to adhere to ethical standards. Further ethical clearance was necessary due to the vulnerable group of participants. Participants were chosen if they have a mild form of depression, stress, or anxiety because this is the target audience for many AI-driven mental health chatbots, even though the interview questions do not ask for any sensitive information. The interviewees were informed about the purpose, scope, and data security before to the start of any interviews. The respondents additionally consented for the interview to be recorded and were given the option to withdraw from the study at any time before and during the interview. To prevent bias or misunderstandings during the interviews and make sure the participants' responses were appropriately interpreted, I summarized and shared with the participants' responses from time to time. The collected data – mainly mp3 files- were anonymized and kept in a secured place. For the duration of this investigation, only the researcher had access to the data. After completing the research any data was deleted from every device. For further description of the data management, please see Appendix C (Ethical Considerations).

## 4. Results

**Correlative and Contrasting Views**

*How (potential) users make sense of AI driven mental health chatbots? Taking into account the welfare state and the public healthcare sytem.*

To answer the research question this section focuses on the analysis of the interviews. This analysis is substantiated by employing direct statements from the interviews as evidence. English quotes are shown in their original form, while German quotes have been translated to English and grammatically adjusted to not interrupt the flow of reading. Interviewees are cited as I- 1-13. After analyzing the codes finding indicate that (potential) users have three different ways in making sense of these chatbots. Based on these differences three ideally typical discourses were discerned from the sample. They hold different views on the use of AI driven mental health chatbots, how they perceive them within the welfare state, and their intention to use it. The next section will first describe the shared views and will then delve deeper into the three differing perceptions.

## 4.1 Correlative Perspectives

AI-driven mental health chatbots remain largely unused by the interviewees, but with most of them being aware of their existence. Only one interviewee, I-5 had experience with an AI-driven mental health chatbot after completing a personal therapy. As a result, the majority of the interviewees could only speculate on the potential benefits of using these AI-driven chatbots in the broader context of public healthcare.

Despite their limited exposure, all interviewees expressed a shared perception of the potential advantages of AI-driven mental health chatbots. They highlighted two main benefits: resource-saving capabilities and time-independent accessibility. For instance, I-1 noted that these chatbots "can be cheap" and "save a lot of money and resources." This sentiment was echoed by I- 2, who emphasized the urgent need for mental health care, especially after the COVID-19 pandemic, as the mental healthcare system is under significant strain. Long waiting lists were mentioned by I-4, I-2, I-7, I-8, I-11, and I-13 and with some individuals having to wait for one to three years before receiving the support they require. AI-driven mental health chatbots, according to the interviewees, have the potential to reduce this burden on those seeking mental health support. Moreover, the technology-driven nature of these chatbots offers the potential for scaling up therapy and reaching a larger number of people in need of mental health care support, as stated by I-13 “You can scale therapy in the sense that you can base it on technology, so you don't need human resources, which are limited for the therapy itself“ (I-13).

Another perceived benefit is the immediate and continuous accessibility provided by AI-driven mental health chatbots. Interviewees highlighted that there is no waiting time to receive support, and the chatbots are "always available," making it possible for individuals to seek help at any time, even during late hours, as mentioned by I-2.

All interviewees agreed on the low barriers to access these AI-driven mental health chatbots, primarily due to their minimal or no costs. They offer a “partly free support” (I-5) so “barriers are relatively low” (I-5). Additionally, these chatbots provide a lower threshold for individuals who might feel uncomfortable or socially unable to communicate with a human therapist, as they offer a sense of anonymity (I-8).

Furthermore, the interviewees highlighted the persisting „taboo around“ (I-2) mental health disorders. AI-driven mental health chatbots have the potential to mitigate this stigma, making it easier for people to seek help without fear of judgment (I-2, I-8).

In summary, the interviewees recognized the potential benefits of AI-driven mental health chatbots, such as cost-effectiveness, continuous and obstacle- free accessibility, and reduced stigmatization. However, their limited exposure to these chatbots implies that their

perceptions are mostly based on assumptions, underscoring the need for further research and broader implementation in the public healthcare system.

All interviewees emphasize the importance of human elements, context, and emotional connection in therapy. Participants express that a chatbot „lack of the human element" (I-2) and "human interaction" (I-12), such as someone smiling empathically or observing facial expressions and gestures (I-1, I-5, I-8, I-13). Interviewees 4, 9 and 11 state that an AI driven mental health chatbot lacks the ability to understand the complexity of emotions. Interviewee 4 stated "It is important that you don't just go by a catalog (…) to develop a sense for the appropriateness of questions and a kind of intuition" (I-4). They highlight the „uniqueness of each individual" (I-5) and the difficulty of trusting a machine-based interaction (I-7). The absence of direct in-person interaction, especially in deep psychological discussions, is recognized by Interviewee 4, 8 and 11. Participants stress the significance of intuition, experience, and context-based understanding in therapy, beyond following a standardized approach. I-8 stated to not use the chatbot "when it goes in the direction of depth psychology" because it is difficult to have the computer do it" (I-8).

Interviewee 5 mentions the interplay between the digital age and the importance of human interaction "In a society where everything is becoming more digital, and somehow, we have many more options to avoid exposing ourselves to human interaction, so to speak, I believe that when it comes to mental health, it is crucial to reintroduce this human component and to foster a sense of belonging". Meaning, in a digitally advancing society with fewer human interactions, it is essential to prioritize the human component in mental health for a sense of belonging. Personal therapy with a human is considered „more sustainable to engage in therapy" (I-5), where someone listens, provides space for personal exploration, and offers a perspective that can lead to profound changes (I-8).

Overall, the participants value the richness of in-person therapy, which encompasses the emotional connection, intuitive understanding, and contextual awareness that a chatbot or digital interaction may lack. Although there are several aspects of therapy that respondents value, which AI bots currently cannot offer, they show openness towards the idea of AI mental health bots as they also value the benefits to overall mental healthcare system. Below, I outline three distinct ideal-typical types that illustrate how they make sense of AI chatbots for mental health therapy in the context of the public healthcare system. These types do not fit the respondent 1-on-1 and do not reflect an individual. This is due to the fact that human individuals are not simply defined in black and white terms; their perceptions can shift on different topics. Hence, these types are depicted as discourses that reflect an ideal-typical characterization of an

individual. Each discourse consists of how benefits and limitations of AI chatbots are perceived, the perspective on the welfare state's role in promoting accessible and reliable mental health services connected to perspectives on neoliberal development, influencing the implementation of AI-driven mental health technologies in society, and the intention to use it.

### 4.2 Contrasting Perspectives

*Table 2 Overview of the three discourses on AI mental health chatbots*

| *Discourses/ Views* | **Key Values** | **The use of AI chatbots** | **Public Healthcare vs. Privatization** | **Intention to use AI Chatbot** |
|---|---|---|---|---|
| **The Visionary** | | Chatbot available 24/7 to provide support and advice on | In favour of privatization to drive innovation | Engaging and using chatbot for behaviour therapy and daily life support |
| **The Pragmatic** | | Human-Machine collaboration: AI chatbot as a complement | Collaboration between state regulations and private companies | When guided/ regulated by state to prevent ethical risks, then intention to use |
| **The Skeptic** | | Lack of human element and unable to understand complexity | State should be more involved in public healthcare system | Skeptical about efficiency and concerns about ethical risks |

## The Visionary

The Visionary approach to AI driven mental health chatbots is characterized by seeing the potential of innovative developments for the public mental healthcare system. From this perspective the implementation in the society is perceived as positive and as a solution to solve problems society is facing. Therefore, the ideal-typical Visionary perceives the neoliberal

development as necessary to provide qualitative, innovative mental health care support and emphasize the benefits of free market forces such as competition and the focus on the individual. Governmental regulations and public insurances are considered a foundational framework, but they are not actively involved as players. Also, when it comes to ethics and accountability, they expect the companies being responsible in order to make profit and therefore ensure quality and security.

When it comes to the use of AI-driven mental health chatbots, interviewees perceive the efficacy of chatbots as sufficient when facing "mild mental disorders or problems with addiction" (I-6). I-11 shares the opinion that AI chatbots are suitable "for those that still function normally on a daily basis, having a 24/7 available guide“ (I-11) to provide support in daily-life situations.

I-6, I-8, and I-11 mention limitations to the use of AI-driven mental health chatbots, stating that they are not suitable for "mental extremities and individuals with acute suicidal tendencies or more severe conditions such as schizophrenia"(I-11). Since mental health is such a sensitive topic, extreme awareness „must be exercised in the language used to prevent potential risks of bias, as all data is based on human input“ (I-11). Interviewee 11 suggests, counteracting misdiagnoses can be achieved by refraining from hasty diagnoses and not assuming or attributing conditions like depression to patients without proper evaluation.

The Visionary’s perspective reflects a fundamental trust in the welfare state and governmental supervision concerning the integration of new technology into society, especially when it benefits the public sector. Individuals with a visionary perception of AI-driven mental health chatbots tend to consider data privacy and security less crucial during chatbot use, they stated, for instance, “Honestly, I don’t really care” (I-6) or “I don’t mind sharing it. I am a very open person” (I-1). I-11 exhibits a lack of concern about data security as long as confidentiality is maintained, akin to human therapy (I-11). In fact, I-11 even suggests that sharing more information can lead to a better outcome for the AI chatbot's quality, particularly regarding the machine learning mechanism. "It could even be quite good if the algorithm can learn from it"(I-11).

While the government should supply a basic general framework for using AI, such as “public health insurances” (I-6), Interviewee I-6 also highlights the potential benefits of private sector involvement, as it can foster competition and innovation in mental health support. Interviewee I-6 suggests that competition-driven innovation may lead to improved services because to make profit, companies supply high quality and take responsibilities for not causing harm to users to not lose them as a ‘customer’. I-6: „If people can make money from it, then

they will also receive a certain level of service or quality, so that exactly contributes to this general well-being“. In a company there are “Cleverer minds, sharper minds than those in the health ministry“ (I-6) referring to more capabilities to build high- quality AI driven mental health chatbots. The welfare state is perceived by Interviewee 9 as a proponent of investing in AI technologies and ensuring universal access to state-of-the-art mental health care services. Interviewee 9, for instance suggested,

> *“I'm just a huge fan, no one has ever done this before, but in my little limited mind, I have come up with the absolute brilliant solution. You say there's a prize money for the developer who creates a functional and excellent chatbot. They would receive a prize of six million, and all the teams will be funded until then. So, eight teams will start working on it and then you get the best possible outcome" (I-9).*

When it comes to responsibility and accountability companies are taken into focus: “Ultimately, the person who writes the program is responsible for any consequences that may arise. “(I-6) Interviewee I-9, I-11, I- 13 all share the same opinion that accountability and responsibility is with those “who make the money with it” (I-11).

However, Interviewees raise ethical considerations, such as ensuring equal accessibility and preventing harm caused by biases. To overcome barriers of accessibility, I-6 identifies the most effective adjustment in the design and interface of the chatbot. The interviewee refers to the development to the creation of "iPhones that are as easy to use for older people in the beginning" (I-6). With improvements in design and interface, older individuals and people with disabilities can also utilize the chatbot effectively. If this "relatively easy adjustment" (I-6) is implemented, AI-driven mental health chatbots can offer unlimited access to "every person with a smartphone" (I-5, I-6, I-9). To avoid bias and harmful mental health support, participants I-6, I-9, and I-11 emphasize the importance of continuous development and close supervision of the data used to train the chatbot. Private companies are seen to be responsible to ensure ongoing monitoring and refinement of the chatbot's training data which is crucial to prevent harmful outcomes. Transparency is crucial here: I-11 emphasizes that users should not have false expectations so that the best can be achieved from the chatbots. Transparency is "absolutely important (...), the crucial point to use the chatbots with a clear conscience (...). Before such a tool is used, it must be clearly communicated what the patient can expect" (I-11).

However, interviewee I-6 acknowledges the transfer of responsibility from the state to private entities, which could impact decision-making and the commercialization of mental health services. The recognition of the shift in power to private companies is evident, yet concerns appear minimal due to the assurance of governmental security. As stated, "health

insurance would cover, for instance, rising prices made by private companies when it comes to healthcare" (I-6).

The visionary perspective highlights a trust in the welfare state's role in integrating new technology for public benefit. Participants emphasize the responsibility of private companies to implement save use of AI chatbots. The involvement of private sectors is seen as beneficial for fostering innovation and competition in mental health support.

Interviewee I-8 recognizes the potential for human replacement and believes that preferences for human or chatbot support will depend on individual needs (I-8). Moreover, I-4 envisions chatbots becoming even more efficient and accurate than human therapists, suggesting that they could „make no mistakes when well-developed“ (I-4). I-8 and I-11 foresee chatbots becoming capable of understanding complex emotional contexts, reducing the importance of personal contact (I-9).

Interviewees 11 and 8 consider AI chatbots practical and believe they may eventually simulate human therapists effectively. However, the precise timing of this capability remains uncertain, and whether it is desirable is a question raised by I-11. Interviewee 13 envisions a "tandem between a human therapist and an artificial therapist" (I-13) for 24/7 healthcare support. The sharing of information between AI and human therapists is envisioned to enhance best possible outcomes for the quality of mental health care.

Contributing to the visionary perception, I-6 suggests that chatbots can encourage proactive behavior and empower individuals to take charge of their mental well-being through problem-solving strategies and guidance (I-6). This proactive support can lead to early intervention, improving mental health outcomes and overall well-being. AI-driven chatbots offer affordable and accessible support, potentially reducing the financial burdens associated with traditional mental health services (I-9) This could revolutionize mental health support and widen its availability to more people. AI driven mental health chatbots can contribute “successfully to the decrease of the gap between supply and demand” (I-11) of mental health care.

The visionary perspective on AI-driven mental health chatbots is characterized by a forward-thinking and visionary approach, inspiring long-term goals, and innovative developments in mental health support. Envisioning a future where bots can complement humans, interviewees perceive these chatbots as providing instant and personalized support. When facing mild mental disorders or addiction problems, they are willing to trust and engage with the chatbot, recognizing AI solutions to overcome societal challenges.

From a visionary perceptive AI-driven mental health chatbots can improve societal well-being, reduce disparities, and enhance quality of life. Acknowledging technological current limitations, the visionary perception highlights the potential in private sector involvement for innovation and envision powerful synergy between humans and AI, revolutionizing.

## The Pragmatic

Another approach is the pragmatic view on AI driven mental health chatbots. It is characterized by the pragmatic perception seeing them as a complement to current mental health care and as a first point of assistance for the intention of receiving help when facing mild mental disorders. They perceive governmental regulation as necessary if there is innovational outcome and imagine successful implementation into the mental health care system as a collaboration between the welfare state and private companies. For Pragmatics, it is essential to involve mental health professionals in the design and ongoing evaluation of chatbots to ensure their effectiveness. The pragmatic view was the most common perception among the interviewees. Interviewees perceive the chatbot as an efficient complement to personal therapy (I-2, I-10, I-13), especially “before and after a traditional therapy” (I-5). Interviewee 12 sees the chatbot for making “pre-diagnosis” (I-12) and I-13 states: “Those chatbots can help with onboarding, so that before you would talk to a therapist, you have a mental health bot that guides you with basic questions and tries to get crucial information to then redirect you to a specialist” (I-13), so “it saves time, finding the right personal therapists” (I-1) and can motivate users to take action to find actual human therapy (I-4). For Pragmatics, chatbots still feel “rudimental” (I-5) and are perceived as to give “generalized” information about mental health care (I-6).

Additionally, there are reservations about chatbots effectively addressing complex issues. For instance, Interviewee 9 doubts that the chatbot can help in a case like this “an art dealer feels bad because he has just started a company" (I-9). An AI chatbot might not grasp the complexity of this statement as it appears to be generally positive and not a necessarily something to worry about. Interviewee 11 highlights that „the reasons behind the success of some therapeutic approaches are not fully understood and „as long as one does not truly have solid knowledge in this area, one should not even begin to attempt to build or program an artificial agent that can completely replace human therapists" (I-11). Interviewee 4 also believes: "It won't replace psychologists, definitely not" (I-4).

The pragmatic perspective on the welfare state's role in AI-driven mental health chatbots recognizes the need for regulations and oversight to prevent the private sector from exploiting vulnerable individuals and prioritize patient wellbeing over profit (I-2, I-9). Interviewee 1 views the welfare state as responsible for „general well-being“(I-1). I-1 expresses confidence that the welfare state would not implement insecure solutions, stating „They would not implement something that is not secure enough” (I-1). Interviewees 1 and 2 highlight the significance of General Data Protection Regulation (GDPR) regulations in ensuring data protection and privacy compliance, contributing to the perception that using AI-driven chatbots is relatively safe (I-2). On the one hand, Interviewee 2 expresses a level of trust in certain regulations such as GDPR. On the other hand, there is also concern regarding the lack of understanding of artificial intelligence and data security, leading to uncertainty about how AI works and the safety of personal data (I-2). I-2 stated, "I do trust that we have certain regulations, for example, like GDPR. But on the other hand, we also don't really understand artificial intelligence. So that's why I'm also not completely sure how does it even work and how safe is my data".

Interviewee 2 highlights the potential danger of AI technologies and suggests that the government should regulate companies working on AI to ensure safety:

> *„AI could be super dangerous for humanity and there's not really someone who is working on the safety of AI. But I feel like if someone would be helpful, responsible, I think it's government, but then they should regulate the companies that are working on those technologies. The role of the government to regulate that. Going to be something that is used by everyone. All the private companies are using, but it is the government that set regulations for it like how everyone is using the internet. So I feel it's their responsibility” (I-2).*

Furthermore, I-5 expresses the importance of one’s own consciousness and believes that “first, and foremost, as an individual, you can think for a moment about what actually happens with the data” (I-5).

Regarding accessibility, I-5 emphasizes the responsibility of both individuals and the state to consider data usage and the necessity for certain regulations to ensure broader access to mental health support (I-5). The pragmatic perspective on human bias in AI-driven mental health chatbots recognizes the influence of training data and therapist interactions in shaping the chatbot's responses (I-9, I-11). Cultural differences and communication conventions also need consideration when deploying chatbots in diverse regions (I-11).

There are varied opinions on the level of transparency needed regarding chatbot utilization and outcome. Some participants feel that as long as the chatbot functions effectively and produces desired outcomes, the specifics of its operation may not be critical (I-8, I-13).

Interviewee 13 refers to an open minded approach “and would try to give it a chance and see like if the conversations are actually going into the direction, I mean if it works, it works“ (I-13). Others such as Interviewee 4 states that „some regulatory standards cannot be outsourced to private companies under any circumstances" (I-4) stresses the need for the government to „set standards just as for human therapy” and suggests to create a „mandatory first page for AI-driven mental health chatbots, clearly stating limitations and that chatbots cannot replace certified therapy “(I-4).

Interviewees expresses concern about the government's ability to develop effective apps promptly (I-4, I-5), „as seen during Covid-19 with the Covid-Warning App“ (I-4). I-4 also mentions that the government's regulatory „processes in the EU can be slow“ (I-4) and points out that companies have the potential to set their own ethical standards and codes of conduct within their communities (I-4). I-2, I-4, I-5 all stress the fact that private companies need to integrate specific rules into their apps to prevent harmful use and ensure equal access to mental health support.

From a pragmatic standpoint, "mental health bots are still very much in their early stages," (I-4) requiring substantial investment and research to become effective additional help. Interviewee 10 emphasizes the “a need for extensive discussions and debates involving various stakeholders, including mental health experts, IT developers, patient associations, and psychotherapist associations, to address the challenges and ensure the successful integration of AI-driven mental health chatbots“ (I-10). A larger exchange of diverse perspectives is essential to incorporate different viewpoints and address various security issues and concerns effectively.

Interviewee 13 emphasizes that the “human aspect is quite relevant and especially when it comes to mental health and emotional issues, we would be weird to assume that everything could be solved by a sort of rational and non-emotional entity“ (I-13). To overcome barriers to human-human interactions in mental health care. The Pragmatics are open to using chatbots as a complementary role alongside human therapy. “In the current developmental stage, it can already offer a very good assisting role, a complementary role to human therapy” interviewee 11 stated. Interviewee 13 could imagine positive outcome when it is used as “a mixed form (…) of having mental health bots in assistant to psychologists that maybe the workload of psychologists could be reduced through these assistants“ (I-13). Also, I-5 sees the AI chatbot as “an addition to (…) personal therapy and beyond, when there is no longer in face-to-face therapy” (I-5) and can be useful “as a tool for therapists as continuing support” (I-5), even when the actual therapy already ended.

The pragmatic approach perceives AI-driven mental health chatbots as a complementary tool to current mental health care, offering assistance for mild disorders, while emphasizing the need for governmental regulation, collaboration between the welfare state and private companies, involvement of mental health professionals in design and evaluation, and consideration of biases, cultural differences, and data privacy.

## The Skeptic

The Sceptics are characterized by doubting the use of AI driven mental health chatbots and its successful implementation in the public healthcare system. Several factors play a significant role. They are questioning their efficiency, potential biases, equal accessibility, and the quality of data they use. Additionally, when it comes to implementing the AI driven mental health chatbots into the public health care system, interviewees raised concerns about consequences of the neoliberal development and the role of the welfare state ensuring a safe implementation.

When looking at the efficiency and accuracy of them, I-7 feels skeptical because "most of these systems fail quite often" meaning that experiences were made with former chatbots that did not fulfill their tasks sufficiently. It is perceived as a "machine that cannot see gestures and mimics" (I-8) therefore lack empathy which is perceived as essential for successful therapy (I-8, I-2, I-3). Interviewee I-7 stresses the importance of having a therapist who can „empathize, validate, and ask meaningful questions because (…) therapy is so personal and private"(I-7). The perception is that AI chatbots might „lack the emotional understanding and depth of human therapists" (I-7).

I-2 states, that because of not being able to contextualize the mental disorder, the chatbots can only help with "basic things" and cannot give "real advice" (I-2). Interviewee 3 emphasized on mental disorders "that fall in between" one diagnosis what the chatbot is not able to differentiate, raising attention on experiences human therapists made (I-3). The sheer „volume of data may lead to generalizations rather than personalized treatment" (I-5), as pointed out by I-5. Building on that, if there is a „misdiagnosis" (I-6) and subsequently an inappropriate therapy is initiated, it could result in individuals being left to „manage the therapy on their own, which might not be beneficial, and lacking proper therapeutic supervision" (I-4).

The Skeptical raise concerns about the role of the welfare state and the development of neoliberal ideology in the context of AI-driven mental health chatbots. They voice concerns about the privatization of mental health care and the profit-driven motives of the private sector. There is worry that the pursuit of financial gains might overshadow the genuine well-being of

patients (I-1, I-5). Interviewee 5 has a clear opinion about the neoliberal shift of healthcare: “The healthcare system should not be privatized because, in my opinion, it leads to a shift in interests influenced by money. I believe that, in the context of healthcare, this is a wrong approach, and therefore, it should remain mostly in state hands, at least from my perspective" (I-5). Interviewee 4 adds on that saying “I’d rather trust the state” instead of private companies. They worry that personal health information might be misused “outside the health context” (I-4) or sold by private companies for „market research purposes“ (I-5).

To address accessibility issues, Interviewee 7 critiques the historical development of neoliberal policies, fearing that AI-driven chatbots may align with a privatized healthcare system, leading to „inequalities for minorities “(I-7) in access to mental health services, similar to the problematic examples in the USA (I-7). „Not everyone has internet access. Not everyone has like media and data literacy. Like what happens if like there's language diversity, right? Like do they operate just as well in Spanish as they do in English“ (I-7)? Additionally, the consequences of the neoliberal development can be crucial, referring to “the US like if you don't have insurance through your job, you're bugged. And that obviously depends on like your citizenship status, your race etc.” (I-7).

Concerns about the quality and equity of therapy arise, as chatbots may primarily appeal to lower socioeconomic groups and may not provide personalized treatment (I-4). This system of mental health care provision „creates a certain milieu where it is easier for specific groups (the wealthier) to obtain human therapy, while others may face barriers or have limited access, leading them to resort to alternative options (such as chatbots) and inequity for the quality of therapy” (I-4).

Furthermore, skepticism arises about the state's ability to create a suitable legal framework and effective regulations to govern AI-driven mental health chatbots. The fast-paced and ever-evolving nature of technology often outpaces the capacity of existing legal systems to hold these AI systems accountable (I-7). I-7 and I-2 state that the “legal framework is really behind technology” (I-7) and there are “only a few people working on it” (I-2).

The topic of data privacy is a central concern for skeptical individuals. Interviewee 1 perceives that “specifically in the Netherlands I know data security is usually not that good” (I-1). The lack of transparency in data usage and potential data breaches raise further doubts about the safety and ethical considerations surrounding AI-driven mental health chatbots (I-9). Interviewee 9 believes that „using data for algorithm training without clear explanation is critical and raises privacy concerns“ (I-9). To ensure the responsible use of AI-driven mental

health chatbots, skeptics emphasize the need for „documentation and transparency about the limitations" (I-7) which is essential for users to make informed decisions.

Skeptics also suggest that investing in traditional therapy approaches and increasing resources for psychotherapists would be a better use of funds compared to solely relying on AI-driven solutions (I-4, I-7). Interviewee 7 believes "that there's like a lot of other issues that energy and money could be put into" (I-7) and Interviewee 4 fears that this development "is seen as a solution so that no therapy places have to be paid for, and with that, the claim to a better therapy with a psychotherapist would then be eliminated" (I-4).

When it comes to the intention of using a chatbot for mental health, the Skepticals would rather prefer a human therapist. However, if that option is not available due to various reasons, Interviewee 5 perceives the chatbot as an "add-on" (I-5) and believes „that before someone doesn't engage with their mental health at all, a chatbot is still better than nothing" (I-5). Another interviewee compared a chatbot with Google, that gives information when asked. "If I feel down, I can google it (…) and if I haven't been able to help myself by looking those things up, then seeing them written out by a bot is not going to do anything for me either, because it's the same" (I-2). Skepticism exists about the feasibility and effectiveness of chatbots referring to emotional and highly individualized therapeutic support (I-13, I-5, I-7). Interviewee 13 states: "'cause usually it's a very emotional topic, so I'm not sure if I would use it" (13).

Expertise and the depth of knowledge are significant considerations. Interviewee 7 questions whether AI chatbots can truly be considered experts. Expertise, according to the Interviewee 7, is "so human and because it comes from such a depth of knowledge and experience where, to me the AI chatbot's not the expert" (I-7). This also raises concerns about the de-skilling of therapists and the impact on the quality of existing therapies (I-7). Interviewee 7 would „ assume that with AI driven chatbots, the profession of therapists would get de-skilled in ways, right? Where like maybe they're analyzing conversations or they're helping train models, but they're not actually speaking to people. And that's like a completely different skillset than the profession has been built" (I-7). Therefore, supporting such systems is seen as highly concerning by Interviewee 7.

In conclusion, skeptical perspectives on AI-driven mental health chatbots revolve around concerns related to their efficiency, potential biases, equal accessibility, and data privacy. A skeptical perspective expresses doubts about chatbots expertise to provide personalized and empathic therapy. The preference for human-centered therapy and caution in using AI chatbots are central themes among the skeptical. The Skeptical highlights the need for

transparency, informed usage and perceive potential benefits in basic support for mild mental health issues.

# 5. Discussion & Conclusion

This study aimed to shed light on how (potential) users make sense of AI driven mental health chatbots. To contextualize their perception, we investigated AI driven mental health chatbots how they work and what potential benefits and limitations could be. I have thus placed them on a macro level in the overall welfare system, shown that there are different approaches to ensuring general well-being and healthcare and what impact neoliberal developments have on the different forms of welfare states. AI chatbots can play an important role in solving the problems of the ever-increasing number of mentally ill people and the ubiquitous resource scarcity typical of neoliberal ideology. Implementing them in the current public healthcare system requires consideration of ethical issues that deal with data security, transparency, accessibility, and accountability. The interplay between ethical security and continuous innovation is also an interplay between government regulation and private enterprise. The individual viewpoints of (potential) users have provided insights into the allocation of responsibilities and the collaborative efforts required from both the government and private companies to ensure the secure implementation of new healthcare innovations. These opinions have also highlighted the various factors that influence the intention to use these innovations in the healthcare system.

The findings show three different ways of making sense of these AI mental health chatbots. The Visionary, the Pragmatic, and the Skeptic approach. All emphasized benefits the of resource-saving capabilities and increased (time-independent) accessibility of these chatbots, they all agree on the immediate and discreet support, and the reducement of stigmatization barriers, as also found by Sepahpour (2020) Nardanzynski et al. (2019).

However, all discourses perceive a major limitation of the chatbots as the lack of the human element and emphasize the complexity of emotions and these are perceived as essential for mental health therapy.

The visionary approach consists of hope and trust in mental health chatbots and focus on the potential benefit these can bring to the overall public healthcare system. Ethical considerations are considered but can be addressed and solved in terms of idealistic approaches. The responsibility of the state is primarily to encourage innovation by private companies, which

visionaries see as best suited to provide both quality therapeutic care and transparency of limitations and data security. The visionary approach does not focus on data security regarding the use of their data, but on solving problems of scarcity of therapeutic support and thus increasing the general well-being of the population, and hope that technological development can be continuously improved.

The pragmatic approach put a special focus on complementing human therapists. Chatbots are (not yet) capable of providing the same quality of care as human therapists due to a lack of understanding of complex situations. A collaboration between human health care professionals and chatbots are perceived as realistic and hold promise to improve the overall health care system. For the safe implementation and use of chatbots, they see the state as responsible for implementing regulatory standards regarding ethical issues and thus avoiding possible risks. At the same time, they also see the potential of private companies to ensure the effectiveness of chatbots. If there is continuous further development and analysis of risks with several stakeholders, they are positive about an introduction into mental healthcare.

The skeptical approach, on the other hand, emphasize mistrust on the effectiveness and quality of chatbots for adequate therapy and mental health support. They see the expertise, experience, and the empathy of human therapists as essential and doubt that chatbots can achieve similar success. Skeptics criticize neoliberal development and focus on potential risks such as human bias, unequal accessibility, and data misuse out of profit-oriented motivation. In their opinion, the state should be more involved in regulation, as they trust that the general well-being is more in the foreground. Therefore, a skeptic would not be likely to do so in nearer future and would rather advocate an expansion of the health care system with a focus on human therapy services.

Due to the neoliberal trend in all three types of welfare systems, there are difficulties in accessing care and concerns about declining quality in mental health care due to cost-cutting, deregulation, and privatization (Konuralp & Bicer, 2021). Additionally, the shift to neoliberalism has transformed the perception of health from a social obligation to an individual responsibility. As a result, health care has taken on a "health-economic" dimension, with health systems becoming more integrated into prevailing economic conditions and market forces (Dayioglu, 2022).

The visionary perspective highlights the potential of high-quality technological innovations driven by market forces, such as competitiveness. Chatbots are considered an effective solution to address issues in the mental health care system, enhancing accessibility and self-sufficiency in the mental health field. Studies by Prakash (2020) demonstrate the

positive perception of chatbots by users. Neoliberal development also offers prospects for quality assurance, as envisioned by the visionaries, given the increasing number of well-educated individuals working in the economic sector. Ethical considerations, like transparency and accessibility, can be ensured through clear communication of limitations and user-friendly design by companies. This reflects the principles of a liberal welfare regime, which emphasizes commodification and sees potential in privatizing public goods (Esping-Andersen, 1990). This explains why chatbots have been widely implemented in the USA (Adb-Alrazaq et al., 2019). Further research is warranted to ascertain if the visionary perspective is prevalent in other regimes. As a result of neoliberal developments and increasing privatization in other welfare systems, technological innovations in public healthcare are evolving. However, due to the shift towards profit-driven intentions, these innovations are costly and therefore accessible primarily to the affluent (McMaughan et al., 2022).

The pragmatic perspective, which favors cooperation between the state and companies, presents a different focus on neoliberal development in the welfare regime. Privatization involves not only transferring public institutions to private companies but also engaging the private sector in collaboration with the state to provide public services. Neoliberal structural adjustment policies, aiming to free market mechanisms from political interference and address perceived market failures and crises like climate change, paradoxically require the exercise of state power (Dayioglu, 2022). The pragmatic discourse acknowledges the potential of chatbots to complement mental health care but raises concerns about ethical considerations, human biases, and limitations when using it in mental healthcare, proposing regulation and quality standards. Collaboration between chatbots and human therapists, as well as private companies and charities, is suggested. This aligns with corporate conservative welfare regime principles, exhibiting moderate commodification (Esping-Andersen, 1990). The majority of interviewees expressed a pragmatic view of chatbot implementation, possibly influenced by growing up in these welfare systems. Further research is needed to explore any correlations.

The skeptical perspective on AI-driven chatbots for mental health is critical of their implementation in the welfare regime. Skeptics believe that supporting the healthcare system with ethical standards in mind is the responsibility of the state, and they do not see private companies as capable of resolving ethical considerations such as equal accessibility. From the skeptics' viewpoint, privatization leads to unequal access to mental health care, disadvantaging certain groups. Chatbots also contribute to making access more difficult for marginalized groups, citing human biases and different cultural contexts. From a critical perspective, chatbots risk being seen as a cost-saving solution and accessible only to poorer individuals, potentially

compromising the quality of mental healthcare and exacerbating mental health issues among this group (Allen et al., 2014). This aligns with principles of a social democratic welfare regime. Further research is needed to explore underlying factors influencing the skeptical opinion among individuals from different welfare regimes, especially those in the social democratic regime, as no interviews were conducted with them.

Possible limitations of this research could include sample bias, where the study's limited sample size, the focus on one social class or the focus on specific demographic groups may lead to biased results and limit generalizability. Self-selection bias may also occur, as participants with pre-existing attitudes or experiences with mental health chatbots might influence the findings. Social desirability bias could impact participants' responses, leading them to provide socially desirable answers rather than expressing their true opinions, especially concerning sensitive mental health topics. The research may have limited perspectives as it primarily focuses on young (potential) users and lacks input from mental health professionals and other stakeholders such as chatbot developers and companies. Cultural variations and older populations' perspectives are not explicitly addressed. Furthermore, the long-term effects of AI-driven mental health chatbots and their interaction with other forms of mental health support may not be fully explored as AI is such a new topic and rapidly developing. External factors, such as policy changes or technological advancements, could also influence users' perceptions and experiences over time. The lack of comparisons with other mental health interventions or technologies limits insights into their relative effectiveness and limitations. As most of the interviewees have not used an AI chatbot yet, the statements on efficacy and functionality of these chatbot are based on assumption. Cultural variations and their impact on users' perceptions and acceptance of AI-driven mental health chatbots in different regions or societies are not explicitly addressed. Research on trust in different welfare regimes and their health care systems among users could be a great addition to this research, to gain deeper knowledge about the influence of the context interviewee live in. Researcher biases could potentially influence the study's design, data collection, and interpretation of results. To address these limitations, future research could also adopt larger and more diverse samples and utilize mixed method approaches to gather both quantitative and qualitative data. Including perspectives from various stakeholders and conducting longitudinal studies could assess the long-term impact of AI-driven mental health chatbots and ensure ethical guidelines are followed to safeguard participants' rights and privacy. Comparative studies and cross-cultural analyses can provide a comprehensive understanding of how these technologies are perceived and utilized in different contexts.

In summary, understanding how AI-driven chatbots are perceived within the context of the welfare system is essential for ensuring their responsible integration into mental health care services. By considering the potential limitations and societal implications, researchers and policymakers can develop strategies that maximize the benefits of these technologies while safeguarding the well-being of all individuals, especially those from marginalized groups. Ultimately, this understanding fosters a more comprehensive and inclusive approach to mental health care, aligning with the principles of a social democratic welfare regime. Given the novelty of this topic and limited knowledge about the processes involved, developing a guideline or code of conduct for companies to implement AI-driven mental health chatbots responsibly would be beneficial. In the absence of governmental regulations, raising awareness among people about the potential benefits and risks could serve as an initial step.

# References

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.

Allen, J., Balfour, R., Bell, R., & Marmot, M. (2014). Social determinants of mental health. International review of psychiatry, 26(4), 392-407.

Boeije, H. (2010). Analysis in qualitative research. Sage Publications.

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., ... & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. Expert Review of Medical Devices, 18(sup1), 37-49.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. Information, 13(1), 41.

Card, K. G., & Hepburn, K. J. (2023). Is Neoliberalism Killing Us? A Cross Sectional Study of the Impact of Neoliberal Beliefs on Health and Social Wellbeing in the Midst of the COVID-19 Pandemic. *International Journal of Social Determinants of Health and Health Services*, *53*(3), 363-373.

Carr, S. (2020). 'AI gone mental': engagement and ethics in data-driven technology for mental health. *Journal of Mental Health*, *29*(2), 125-130.

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE open medicine*, *7*, 2050312118822927.

Dayioğlu, N. (2022). Assessing the Health-economic Crisis: The Case of Turkey. _Lectio Socialis_, _6_(2), 67-80.

Denecke, K., Abd-Alrazaq, A., & Househ, M. (2021). Artificial intelligence for chatbots in mental health: opportunities and challenges. *Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges*, 115-128.

Esping-Andersen, G. (1990). The Three Worlds of Welfare Capitalism, Cambridge: Polity Press.

European Commission. (2019). Ethical Guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai last accessed 06.08.2023

European Commission. (2019). European AI Alliance: ECS Definition of AI or How to Define Artificial Intelligence for Real and be Concerned. https://ec.europa.eu/futurium/en/european-ai-alliance/ecs-definition-ai-or-how-define-artificial-intelligence-real-and-concerned.html last accessed 06.08.2023

European Union. (2023). Artificial Intelligence Act. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf last accessed 06.08.2023

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics, 5(1), 1-4.
Grewal, D. S. (2014). A critical conceptual analysis of definitions of artificial intelligence as applicable to computer engineering. IOSR Journal of Computer Engineering, 16(2), 9-13.

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*, 1-18.

Harvey-Jordan, S., & Long, S. (2001). The process and the pitfalls of semi-structured interviews. Community Practitioner, 74(6), 219.

Haque, M. R., & Rubya, S. (2023). An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews. _JMIR mHealth and uHealth_, _11_(1), e44838.

Hussain, S., Ameri Sianaki, O., & Ababneh, N. (2019). A survey on conversational agents/chatbots classification and design techniques. In Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019) 33 (pp. 946-956). Springer International Publishing.

Konuralp, E., & Bicer, S. (2021). Putting the neoliberal transformation of Turkish healthcare system and its problems into a historical perspective. Review of Radical Political Economics, 53(4), 654-674.

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People's Advisory Group. (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, 1178222619829083.

Larsen, C. A. (2007). The Institutional Logic of Welfare Attitudes: How Welfare Regimes Influence Public Support. Comparative Political Studies, 41(2), 145-168.

Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H. C., ... & Jeste, D. V. (2021). Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. _Biological Psychiatry: Cognitive Neuroscience and Neuroimaging_, _6_(9), 856-864.

Matheny, M., Israni, S. T., Ahmed, M., & Whicher, D. (2019). Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington, DC: National Academy of Medicine.

Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. BMJ, 320(7226), 50-52.

McGregor, S. (2001). Neoliberalism and health care. International Journal of Consumer Studies, 25(2), 82-89.

McMaughan, D. J., Oloruntoba, O., & Smith, M. L. (2020). Socioeconomic status and access to healthcare: interrelated drivers for healthy aging. Frontiers in public health, 8, 231.

Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: a mapping review. Social Science & Medicine, 260, 113172.

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digital health, 5, 2055207619871808.

Nedayvoda, A., Delavelle, F., So, H. Y., Graf, L., & Taupin, L. (2021). Financing deep tech.https://www.jaai.de/post/mensch-oder-maschine-der-turing-test-einfach-erklaert last accessed 06.08.2023

Nadikattu, R. R. (2016). The emerging role of artificial intelligence in modern society. International Journal of Creative Research Thoughts.

Omarov, B., Narynov, S., & Zhumanov, Z. (2023). Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review. *Computers, Materials & Continua*, *74*(3).

Peredo, A. M., & McLean, M. (2020). Decommodification in action: Common property as countermovement. _Organization_, _27_(6), 817-839.

Prakash, A. V., & Das, S. (2020). Intelligent conversational agents in mental healthcare services: A thematic analysis of user perceptions. Pacific Asia Journal of the Association for Information Systems, 12(2), 1.

Sepahpour, T. (2020). Ethical considerations of chatbot use for mental health support (Doctoral dissertation, Johns Hopkins University).

Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. Journal of Database Management, 31(2), 74-87. IGI Global. The definitive version is available at https://doi.org/10.4018/JDM.2020040105

Tewari, A., Chhabria, A., Khalsa, A. S., Chaudhary, S., & Kanal, H. (2021, April). A survey of mental health chatbots using NLP. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

The World Health Organization. (2022). WHO highlights urgent need to transform mental health care https://www.who.int/news/item/17-06-2022-who-highlights-urgent- last accessed 06.08.2023

Vallès-Peris, N., Barat-Auleda, O., & Domènech, M. (2021). Robots in healthcare? What patients say. _International Journal of Environmental Research and Public Health_, _18_(18), 9933.

Van Kolfschooten, H. (2022). EU regulation of artificial intelligence: challenges for patients' rights. _Common Market Law Review_, _59_(1).

Victor, D. (2016). Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk. https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html, last accessed 06.08.2023

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. _Government Information Quarterly_, _39_(4), 101685.

Woebot Health.
https://woebothealth.com/woebot-for-adults/

Wong, L. P. (2008). Data analysis in qualitative research: A brief guide to using NVivo. Malaysian family physician: the official journal of the Academy of Family Physicians of Malaysia, 3(1), 14.

Yan, W. J., Ruan, Q. N., & Jiang, K. (2022). Challenges for artificial intelligence in recognizing mental disorders. _Diagnostics_, _13_(1), 2.

Yu, R., & Alì, G. S. (2019). What's inside the black box? AI challenges for lawyers and researchers. Legal Information Management, 19(1), 2-13.

Zeira, A. (2022). Mental health challenges related to neoliberal capitalism in the United States. _Community mental health journal_, _58_(2), 205-212.

Zemčík, M. T. (2019). A brief history of chatbots. *DEStech Transactions on Computer Science and Engineering*, *10*.

# Appendix A: Questionnaire

**Introduction**

Brief explanation of the purpose and goals of the interview and the use of data regarding the ethics for this research.

With this interview I try to find out how people make sense of AI driven mental health chatbots in the context of the development of the welfare state system and how AI driven mental chatbots would fit within this perspective.

This interview is completely anonym, will use pseudonym when transcribing.

Storing the data secured on my laptop and will delete immediately as soon as its not needed anymore for my research.

Background information about the interviewee (age, gender, nationality, former personal therapy, occupation?).

**Understanding of AI-driven mental health chatbots**

Awareness and familiarity with AI-driven mental health chatbots.

Former use of chatbots in general? Experiences with former chatbots?

Former use of AI tools? What do you know about AI?

If you heard of mental health chatbots, how did you hear about it?

**Perceived Benefits**

What potential benefits do AI-powered chatbots have for mental health care?

**Perceived Limitations**

What potential limitations do AI-powered chatbots for mental health have?

**Dehumanization**

What role do you think AI-driven chatbots should play in relation to traditional therapeutic relationships and human professionals?

Do you worry that using AI-driven chatbots might diminish the importance of human connection in mental health support?

Can a chatbot truly understand and respond to the complex emotional needs of individuals seeking help?

**Collaborative approach**

Do you think a collaboration between mental health professionals and developers and service provider of mental health chatbots help in responsibly deploying AI-driven chatbots for mental health support?

**Impact on Human Interaction**

Did or would your attitude towards seeking professional help change after engaging with AI-driven chatbots for mental health support and why?

Can chatbots complement or augment traditional mental health care, and if so, how?

**Trust and credibility**

How do users perceive the expertise and accuracy of chatbots in providing mental health support?

How important is transparency in disclosing the limitations and capabilities of mental health chatbots to users?

How to ensure that users are well-informed about the nature and boundaries of chatbot interactions?

How can chatbot developers and service providers address issues of bias, fairness, and inclusivity in their algorithms and decision-making processes?

**Privacy and data security**

How concerned are you about the privacy and security of your personal information when using AI-driven mental health chatbots?

What steps do you think should be taken to ensure the confidentiality of user data collected by these chatbots?

**Accountability and responsibility**

Who do you believe should be held accountable for any negative consequences or harm caused by AI-driven mental health chatbots?

Are you aware of any ethical guidelines or regulations that should be in place to protect user privacy in the context of AI-driven mental health support?

How do public and private providers ensure user trust in chatbot technologies and maintain ethical practices?

Are there any concerns regarding privacy, data security, or informed consent specific to mental healthcare chatbots in each system?

What specific actions should each side take to ensure user safety and mental health support?

**Introduction welfare state and mental health care**

Can you provide a brief overview of your understanding of the development of welfare states and their significance in addressing societal well-being and healthcare needs?

**Welfare state and mental health care**

How do you perceive the role of mental health care within the broader framework of welfare states?

**Benefits and challenges of mental health AI bots**

From your perspective, what potential benefits can mental health AI bots bring to the overall mental health ecosystem within a welfare state context?

**Equity and accessibility considerations:**

Are there any concerns regarding the differential access to mental health AI bots based on socioeconomic status, education, or other factors?

How can the welfare state ensure that the integration of mental health AI bots does not exacerbate existing disparities in access to mental health care?

**Future implications and considerations:**

How do you envision the future development and integration of mental health AI bots within the evolving landscape of welfare states?

**Future expectations and improvements**

Hopes and expectations for the future development of AI-driven mental health chatbots.

What are your views on the role of AI-driven chatbots within the broader mental health ecosystem? Did this change over time?

**Conclusion**

Thank you and appreciation for the interviewee's participation and any closing remarks.

# Appendix B: Codebook

For privacy reasons, individual quotes within the coding frameworks are not displayed and transcripts are not provided.

| Grand theme | Theme | Subtheme | Codes |
|---|---|---|---|
| Benefits | Improvement of Mental Healthcare system | Ressource saving | Reducing psychologists workload<br>Cost- effectiveness<br>Waiting lists |
| | | Increase of mental health demand | High demand for mental healthcare than supply<br>Increase due to Covid-19 Pandemic<br>Scaling up therapy |
| | | Capabilities | Developed for mild disorders<br>Daily life struggles<br>Mental extremities, suicidal emergency<br>Depth psychology therapy<br>Behavioural therapy<br>Misdiagnosis |
| | | Easy accessibility | No waiting time<br>Always available<br>Decentralication |
| | Less Stigmatization | Societal Taboo | Shared goals/concerns with other activists |
| | | Private | Self-Disclosure |
| Dehumanization / Lack of human component | Empathy | Lack of human element | Human interaction<br>Gestures<br>Expressions<br>Intuition<br>Sense of Belonging |
| | Understanding complexity | | Uniqueness of individuals |
| | Experiences | Trust | Lack of trust a machine |
| The Visionary | Chatbots for therapeutic use | Efficiency | Views activism as an accumulation of acts |

| | | | |
|---|---|---|---|
| | | Limitations | Lack of human empathy |
| | Favouring neoliberal development | Accessibilty | 24/7 Availability<br>Smartphone accessibility<br>Design/ Interface Languages |
| | | Transparency | Capabilities of chatbots<br>Communication |
| | | Saving Resources | Cost- effectiveness<br>Human resources |
| | | Data Security | Less concerns<br>Confidentiality<br>Monitoring |
| | | Responsibility | Private Companies accountable<br>Developer Supervision<br>Continuous development |
| | | Regulation | Competition<br>Innovation<br>Profit-driven motivation<br>Prize-money State insurance<br>Basic general framework |
| | Acceptance of use of AI in mental health | Mental Health Support | Hope to bridge gaps<br>Trust and engagement |
| | | Self-reliance | Pro-active behaviour<br>Problem- solving management |
| | | Future Development | no mistakes<br>complex emotional context<br>Therapist simulation<br>Human Replacement |
| The Pragmatic | Chatbots accompanying to human therapy | Reorganising therapy | Collaboration<br>Before and after therapy |
| | | Effectiveness | Evaluation<br>Pre-diagnosis<br>Basic Questions<br>No Social Pressure<br>Generalization |

| | | | |
|---|---|---|---|
| | Governmental Guidance for Innovation | Risk of Data Security | GDPR<br>AI dangerous for humanity |
| | | Various perspective Transparency | No need for transparency for limitation<br>Fake expectations<br>Clear communication |
| | | Regulation | Uncertainty of how the chatbot works<br>Risk prevention<br>Government set standards |
| | | Innovation | Ineffective state<br>Slow state |
| | | Bias | Data is biased<br>Mindfulness<br>Cultural variations |
| | | Responsibility | Regulation of private companies to prevent misuse |
| | Chatbots as a complement to human therapy | Collaboration | No human replacement<br>Support and Assistance<br>Onboarding<br>Directing to human therapists<br>Chatbot as a tool<br>Facilitate exchange with human therapists |
| | | Evaluation | Monitoring and evaluation with all stakeholders involved<br>Diverse perspectives |
| The Skeptical | Chatbots not capable of therapeutic support | Expertise | Lack of human expertise<br>Human Intelligence<br>System Failing<br>Highly Emotional<br>Individualized Support |
| | | Misdiagnosis | Generalization problems<br>Inappropriate Therapy<br>Prejudgment |

| | | | |
|---|---|---|---|
| | | Human Abilities | Human relevance for therapy<br>Gestures and Mimics<br>Meaningful questions |
| | Neoliberal development negative impact on healthcare system | Regulation | Trust in state<br>Companies: Profit-driven motivations<br>Slow Regulation |
| | | Accessibility | Age<br>Lack of digital literacy<br>Lack of Internet Access<br>Language Barriers<br>Minorities |
| | | Human Bias | Fear of Harm<br>Cultural Differences<br>Language Barriers |
| | | Transparency | Lack of transparency in data usage<br>Lack of clarity<br>Training algorithms |
| | | Data Security | Data Concern<br>Misuse of Data outside Healthcare system |
| | No adaption of chatbots | Efficiency | Google comparison |
| | | Lack of Humanity | Emotions |

# Appendix C: Ethical Considerations

CHECKLIST ETHICAL AND PRIVACY ASPECTS OF RESEARCH

INSTRUCTION

This checklist should be completed for every research study that is conducted at the Department of Public Administration and Sociology (DPAS). This checklist should be completed *before* commencing with data collection or approaching participants. Students can complete this checklist with help of their supervisor.

This checklist is a mandatory part of the empirical master's thesis and has to be uploaded along with the research proposal.

The guideline for ethical aspects of research of the Dutch Sociological Association (NSV) can be found on their website (http://www.nsv-sociologie.nl/?page_id=17). If you have doubts about ethical or privacy aspects of your research study, discuss and resolve the matter with your EUR supervisor. If needed and if advised to do so by your supervisor, you can also consult Dr. Bonnie French, coordinator of the Sociology Master's Thesis program.

PART I: GENERAL INFORMATION

Project title:

Name, email of student:
Name: Maya Janecke
Email: 669414mj@eur.nl

Name, email of supervisor:
Dr. Roy Kemmers
Email: kemmers@essb.eur.nl

Start date and duration:
Start date 16.04.2023
Duration: 3 months

Is the research study conducted within DPAS YES

If 'NO': at or for what institute or organization will the study be conducted?
(e.g. internship organization)

PART II: HUMAN SUBJECTS

1. Does your research involve human participants. YES

*If 'NO': skip to part V.*

If 'YES': does the study involve medical or physical research? NO

*Research that falls under the Medical Research Involving Human Subjects Act (WMO) must first be submitted to an accredited medical research ethics committee or the Central Committee on Research Involving Human Subjects (CCMO).*

2. Does your research involve field observations without manipulations that will not involve identification of participants. NO

*If 'YES': skip to part IV.*

3. Research involving completely anonymous data files (secondary data that has been anonymized by someone else). YES

*If 'YES': skip to part IV.*

PART III: PARTICIPANTS

1. Will information about the nature of the study and about what participants can expect during the study be withheld from them? NO

2. Will any of the participants not be asked for verbal or written 'informed consent,' whereby they agree to participate in the study? NO

3. Will information about the possibility to discontinue the participation at any time be withheld from participants? NO

4. Will the study involve actively deceiving the participants? NO

*Note: almost all research studies involve some kind of deception of participants. Try to think about what types of deception are ethical or non-ethical (e.g. purpose of the study is not told, coercion is exerted on participants, giving participants the feeling that they harm other people by making certain decisions, etc.).*

Does the study involve the risk of causing psychological stress or negative emotions beyond those normally encountered by participants? NO

Will information be collected about special categories of data, as defined by the GDPR (e.g. racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a person, data concerning mental or physical health, data concerning a person's sex life or sexual orientation)? YES

Will the study involve the participation of minors (<18 years old) or other groups that cannot give consent? NO

Is the health and/or safety of participants at risk during the study? NO

Can participants be identified by the study results or can the confidentiality of the participants' identity not be ensured? NO

Are there any other possible ethical issues with regard to this study? NO

If you have answered 'YES' to any of the previous questions, please indicate below why this issue is unavoidable in this study.

As this study focuses on participants characterized by mental health issues, I will have to know that they have them. The study is not about their diagnosis but as I use semi- structured interviews, I cannot control what information regarding mental health issue they share.

What safeguards are taken to relieve possible adverse consequences of these issues (e.g., informing participants about the study afterwards, extra safety regulations, etc.).
Complete anonymity, deleting of data as soon as the research finished, protecting data on my devices so no one else has access to data. Inform them, that they can stop the interview any time.

_____

Informing participants about the study afterwards, complete anonymity, deleting data as soon as it is not needed (after max. 3 months). My device is secured with a password. I am the only one that has access to my laptop.

_____

_____

_____

Are there any unintended circumstances in the study that can cause harm or have negative (emotional) consequences to the participants? Indicate what possible circumstances this could be.

Participants who get more awareness of the use of mental health bots might oppose the apps which could impact their progress of mental therapy. The consequence and harm of this is not predictable, solution: Motivating or referring to human therapist after the interview?

*Please attach your informed consent form in Appendix I, if applicable.*

*Continue to part IV.*

PART IV: SAMPLE

Where will you collect or obtain your data?

I will collect the data from using in person interviews, online interviews or telephone calls. Trying to reach people from online platforms such as Reddit, Instagram, Twitter and also from personal and University context etc.

_____

_____

*Note: indicate for separate data sources.*

What is the (anticipated) size of your sample?

12- 15 participants

_____

_____

*Note: indicate for separate data sources.*

*Continue to part V.*

Part V: Data storage and backup

Where and when will you store your data in the short term, after acquisition?

I will record the interviews and store them on my computer (hard drive). I will delete the data as soon as the research is finished. I am the only one who has access to my laptop.

*Note: indicate for separate data sources, for instance for paper-and pencil test data, and for digital data files.*

Who is responsible for the immediate day-to-day management, storage and backup of the data arising from your research?

Myself

How (frequently) will you back-up your research data for short-term data security?

In case of collecting personal data how will you anonymize the data?

I will not record and ask for name, gender, race, appearance.

*Note: It is advisable to keep directly identifying personal details separated from the rest of the data. Personal details are then replaced by a key/ code. Only the code is part of the database with data and the list of respondents/research subjects is kept separate.*

PART VI: SIGNATURE

Please note that it is your responsibility to follow the ethical guidelines in the conduct of your study. This includes providing information to participants about the study and ensuring confidentiality in storage and use of personal data. Treat participants respectfully, be on time at appointments, call participants when they have signed up for your study and fulfil promises made to participants.

Furthermore, it is your responsibility that data are authentic, of high quality and properly stored. The principle is always that the supervisor (or strictly speaking the Erasmus University Rotterdam) remains owner of the data, and that the student should therefore hand over all data to the supervisor.

Hereby I declare that the study will be conducted in accordance with the ethical guidelines of the Department of Public Administration and Sociology at Erasmus University Rotterdam. I have answered the questions truthfully.

Name student: Maya Janecke  Name (EUR) supervisor: Dr. Roy Kemmers

Date: 16.04.23  Date: