Utrecht University

Master's Thesis

Value-Oriented Approach to Socially Assistive Robots Design for Young Adults' Mental Health

Benedetta Ghedi[1]

Supervised by Dr. M.M.A. de Graaf

December 2023

[1]MA Artificial Intelligence. Student No.: 7005571. Email: b.ghedi@students.uu.nl.

Contents

Abstract

This thesis aims to explore the potential of Socially Assistive Robots (SARs) for mental health support of young adults with emotional complaints by initiating a value-oriented design process. In line with the AI4SG-VSD framework, this research provides an analysis of the context. This includes, first, framing the societal challenge related to young adults' mental health, related tools and practices, and considerations on SARs; second, by discussing perspectives and needs of selected groups of stakeholders, psychology and Human-Robot Interaction professionals, obtained through qualitative interviews; it further presents a conceptual investigation of selected values relevant to the field, such as autonomy and legitimacy. The insights gathered through the context analysis and value investigation are used to identify potential applications and design recommendations, with the goal of providing direction for further research. The potential applications identified include therapy support, positioning SARs as an entry point to mental health services, and prevention. Recommendations for design and design practices include interdisciplinary and intercultural collaboration, and a focus on the integration of technology in current practices. Future research recommendations include the investigation of the design for the applications identified and the exploration of SARs personalisation strategies.

1 Introduction

Mental healthcare resources are currently insufficient in meeting the requirements of people in need of such assistance [1, 2, 3, 4, 5]. Staff shortages, high costs, and uneven territorial distribution are some of the factors contributing to the gap between demand and adequate supply [1].

Technology-mediated interventions offer a possible solution to mitigate this issue by reaching under-served populations, reducing costs, and improving accessibility to support [1]. Mobile interventions such as chatbots and smartphone applications have shown beneficial effects on users [6], however, they are subjected to poor adoption rates in private and clinical environments [7, 8]. This could be affected by inadequate integration with users' routines [8] or the scarce engagement of these interventions [9, 10].

Embodied Artificial Intelligence has been gaining attention in research and clinical settings [7] due to its ability to interact socially with patients, building affective relationships [1]. Robots that are made to support people socially and emotionally are known as "socially assistive robotics" or SARs. SARs are designed to interact with people in a way that is considerate of their needs and emotions, in contrast to conventional industrial robots. They are employed in contexts related to healthcare, education, and therapy. SARs exist at the intersection between assistive robotics and socially interactive robots: the former comprehends mechanical artifacts designed to assist in a wide range of tasks, such as lifting, transporting, or crafting; the latter defines robots that are able to interact with people through speech or body language [1, 2, 11].

SARs research in therapeutic and well-being contexts has mostly focused on two target populations: the elderly and children with Autism Spectrum Disorder. The evidence for social robot-assisted health, well-being, and psychosocial interventions is still in its early stages, mostly limited to the two main target groups [12, 13].

The potential use of SARs in assisting the mental well-being of the young adult population is still open for investigation [12, 13, 14], with only a few studies on the topic [15, 16]. Exploring innovative care delivery methods, such as SARs, to better support this demographic is a pressing matter: 62% of individuals suffering from mental disorders present symptoms before age 25 [17], with only one-third of those affected accessing mental health services [18, 19, 20, 21]. The onset of issues at a young age can have detrimental long-term effects, influencing education, employment, health, and social outcomes [22, 23, 24].

Developing SARs interventions for emotional disorders has the potential to help a large number of people while also addressing a critical public health need. SARs can be utilized as tools to maintain treatment protocol adherence and keep active between appointments with a real therapist. Robots may be useful in encouraging participation in self-help programs and be a source of interaction and engagement [11]. The focus of this research is directed at young adults with emotional complaints, including both individuals who have not been formally diagnosed with any mental health disorders and those who have.

Introducing novel technology in a high-stakes field such as healthcare requires careful ethical examination to ensure its application respects, supports, and promotes human values. Researchers and organizations have stressed the importance of integrating ethical frameworks at the design phase due to a lack of guidance and standards for the development of these interventions [7, 25, 26, 27]. This research aims to explore the potential of SARs for young adults' mental health support through initiating a value-oriented design approach. Potential applications, recommended features, and research directions are provided through an empirical and conceptual examination of the context and values to respect and promote in mental healthcare technology, applying the AI4SG-VSD framework developed by Umbrello and Van de Poel [28]. To the author's best knowledge, this study is the first to conduct value-oriented research in the field of mental healthcare SARs, contributing to the field of Artificial Intelligence (AI) through a Responsible Research and Innovation [26] approach to the design of an AI-embedded application, focused on positive, socially acceptable and desirable outcomes.

Mental healthcare practices, supporting technology and potential SARs applications are investigated through related work research and the analysis of qualitative data collected through semi-structured interviews with mental health workers and roboticists, following the AI4SG-VSD framework [28]. These insights provide a contextual understanding of the mental healthcare and robotics research landscape and identify stakeholders' perspectives on the application of SARs in mental health support, their anticipated benefits, and potential challenges. Building upon these insights, the values identified as relevant in the context of this research are conceptually examined. Subsequently, leveraging the understanding derived from both the contextual landscape and value analysis, this research identifies potential applications and provides recommendations for future design and research of SARs in support of young adults with emotional complaints.

Three of the four research questions are formulated based on the AI4SG-VSD methodology's phases, which are described in section 2.2, namely context analysis, value identification, and design requirements. The identification of potential applications is included in the research questions, despite not being part of the AI4SG-VSD framework [28]. This is justified by the exploratory, context-driven approach of this research: while design studies usually begin with an application, the starting point for this research is the context analysis. This allows for the identification of potential applications which are in line with stakeholders and societal needs. The absence of a pre-defined application also justifies the choice of presenting design recommendations instead of requirements, as the insights are more general and exploratory, aiming to guide future potential research and applications rather than dictate exact design parameters. It is important to note that the AI4SG-VSD framework is iterative, and does not have a defined starting point: each phase provides insights for the following ones, and, through repetition, the practice allows for continuous improvement [28]. This research presents the first iteration of the aforementioned phases, which inherently aligns with its exploratory nature.
The research questions are formulated as follows:

1. Context Analysis:

   (a) What is the current state of young adults' mental health? Which tools and practices are currently used to support it?

   (b) What are the challenges, understandings, and considerations of psychology and Human-Robot Interaction (HRI) professionals with regard to their profession, the mental health crisis, and the potential of SARs?

2. Value Identification and Conceptualisation: What are the values to promote and respect in the design of SARs for young adults' mental health support, and how are these conceptualised?

3. Potential Applications: In which scenarios can SARs support young adults with emotional complaints?

4. Design Recommendations: What are design recommendations and insights for future research?

This thesis is structured in the following sections: the Related Work section (2) is comprised of two subsections. The first, "Context Analysis - Part One" (2.1) provides background information about young adults' mental health, digital tools, and SARs applications for mental health support, answering research question 1a. The second (2.2) outlines the AI4SG-VSD framework and defines relevant concepts. The methods section (3) describes the procedure of the empirical investigation, including the interview process, the interview structure, and data analysis. Section 4 presents the results of the empirical research. Section 5 is divided into four parts: "Context Analysis - Part Two" (5.1) provides a discussion of the results, answering research question 1b. Section 5.2 addresses the identification of values and their conceptual examination, answering research question 2. The potential applications are outlined in section 5.3, answering question 3. Lastly, section 5.4 presents design recommendations, answering research question 4.

The motivation behind the division of the context analysis, which addresses research question 1a and 1b, in two parts lies in its conceptualisation in the research methodology. The context analysis, as described in the AI4SG-VSD framework [28], involves the identification of societal challenges, the exploration of current technology, and the analysis of perspectives, needs and values of stakeholders. The former two elements coincide with performing a literature review and therefore are located in section 2.1. The latter requires the analysis of the data acquired through the empirical investigation performed in this research and is therefore found in the discussion section (5.1).

2 Related Work

2.1 Context Analysis - Part One

2.1.1 Young Adults' Mental Health

Mental health is a complex and multifaceted concept, understood and approached differently across various paradigms [29]. It can be defined, according to the World Health Organization, as "a state of mental well-being that enables people to cope with the stresses of life, realize their abilities, learn well and work well, and contribute to their community" [30]. According to the biopsychosocial model, proposed by Engel in 1997 and widely accepted in modern psychiatry, mental health is influenced by complex interactions of biological, psychological, and social factors which can have a cumulative effect on an individual [31]. Amongst the factors influencing mental health, cognitive and social skills are considered important, together with emotional regulation skills, flexibility, and ability to cope with adverse events [32, 33, 34, 35, 36]. In literature, the relationship between mental health and well-being is not clear [37]. In this paper, the term mental health is used to refer to psychological well-being and vice-versa.

The most common mental health disorders are emotional disorders, which comprise depressive and anxiety disorders. They are chronic and frequently recurring psychiatric disorders that cause significant impairment in quality of life, productivity, and interpersonal functioning which affect about 25% of the population in Europe [38, 39]. Treatment for emotional disorders typically includes pharmacotherapy and psychotherapy [40]. Psychotherapy involves a structured, supportive, and collaborative approach that enables individuals to explore their thoughts, emotions, and behaviors in a safe and non-judgmental environment [41]. Evidence-based treatments (EBTs) are the gold standard in current mental health practices, built on a foundation of scientific research and clinical evidence [42]. The success of these treatments is not merely a matter of application but is deeply influenced by a variety of factors, such as the quality of the therapeutic relationship, and patient and therapist's personal characteristics [43], with the therapeutic relationship being considered the most influential factor [44].

Adolescence and young adulthood are the stages of life in which most mental disorders emerge, with up to 74% of documented cases of depression occurring before the age of 24 [19, 16, 45]. Research has suggested that the early onset of anxiety and depression leaves

more psychosocial scars and poses greater risks for comorbid mental disorders [46, 47]. Prevention and treatment at this stage of life are therefore crucial in determining long-term outcomes [46, 48, 49]. Mental disorders account for 45% of the global burden of disease amongst people younger than 25, calculated as the sum of the reduction of life expectancy and the diminished quality of life [50, 51]. Suicide is the fourth leading cause of death among 15-29 year-olds [52]. The notion of adolescence and young adulthood is debated amongst cultures and expert groups [20, 53, 54]. Recent social and economic forces, especially in Western countries, have extended the path towards independence in adulthood, posing significant challenges for young adults in modern society [55]. In epidemiological studies, young adults are defined between 18 and 35 years old [46].

Despite the significant frequency of mental health problems among young people, the majority of mental health services have shown to be ineffective in offering healthcare at this important age, with only one-third of young adults with mental disorders seeking professional help [18, 19, 20, 21]. The reluctance to seek help is attributed to negative views on help-seeking behavior, a deficit in mental health literacy, and the stigma and embarrassment linked to mental health concerns [56]. Those who do seek help face accessibility barriers such as high costs and long waiting lists [57]. This failure of current practices to provide support to large segments of the young adults population motivates the development and investigation of novel ways of delivering care.

2.1.2 Digital Tools

Technological advancements opened new avenues for research and innovation for mental health support. Digital tools dedicated to mental health and well-being employ different media and techniques to provide users with support, education, and resources.

Teletherapy refers to employing video conferencing tools to provide remote mental health services. This has grown in popularity during the COVID-19 pandemic as it allows therapists to reach patients regardless of location. Studies have demonstrated the effectiveness of online delivered therapy, making it a valid substitute for traditional face-to-face therapy [58].

Computer-delivered therapy, on the other hand, allows patients to access educational content, exercises, and other components of therapy through digital platforms. While the guidance of clinicians has yielded better results, self-help programs have also been found

to be successful in treating anxiety and depression [59, 60, 61].

Therapy can also be delivered by chatbots or virtual agents. These computer programs utilize natural language processing and machine learning algorithms to simulate conversation with a human therapist, with the aim of providing mental health support. They are typically designed to offer evidence-based therapeutic interventions such as cognitive-behavioral therapy [6] and can provide support for a range of mental health concerns, including anxiety, depression, and stress [62].

A multitude of smartphone applications have been deployed to assess, treat, and alleviate mental health conditions [3, 63]. These apps can be used to manage symptoms of mental illness, improve mood and emotional regulation, reduce stress and anxiety, promote mindfulness and relaxation, and connect with mental health professionals or support groups [64].

Wearable devices can provide insights into patterns or changes in mood, anxiety, or other mental health-related symptoms by monitoring metrics such as sleep, heart rate, and physical activity through embedded sensors.[65].

In mental health treatment, Virtual Reality has been employed in a range of applications and it is showing promising results. It involves computer-generated 3D environments accessed by users through goggles, creating an immersive sensory experience suitable for exposure therapy, cognitive behavioural therapy, and other evidence-based treatments. It creates a safe and controlled environment for patients to interact with simulations of different kinds, such as anxiety-inducing situations. [66, 67, 68, 69]

From supporting users' well-being through meditation, monitoring mood, progress, and therapy adherence [3] to cognitive behavioural therapy chatbots [6], technological interventions are improving accessibility to care. One of the main issues of many of these tools is low long-term engagement, possibly influenced by limited social presence [1, 3, 63].

2.1.3 SARs

Socially assistive robots are designed to interact with people socially through verbal, non-verbal, or affective modalities [3, 70]. Social robots vary greatly in terms of appearance, social capabilities, levels of autonomy and intelligence, roles, proximity, and temporal profile [71]. Socially assistive robots are fully programmable, making them suitable to adopt different roles in various contexts [3]. Their social presence, which is the extent

to which they are perceived as a social entity, is influenced by their embodiment and social capabilities. Researchers have found that SARs' social presence can have a positive impact on the engagement and motivation of users, potentially making interventions more effective. [72, 73]

SARs have been researched and deployed to support mental health as companions, play partners, and coaches [11]. They have been used to provide comfort and reduce stress in the elderly population [74, 75, 76, 77, 78, 79], young adults [80] and children [81, 82]. Some of these studies showed mixed or negative results [78, 82]. Robots have also been used for medication reminding [83], exercising coaches [84] and motivators [85].

In the children with autism population, the applications are wide, ranging from building and improving social and emotional skills [86, 87] such as taking turns [88, 89], communication [90] and emotion recognition [91]. The hope is that, through exercising different skills with the aid of robots, these can be applied with human peers [7].

Lastly, robots have been employed to deliver or aid therapy: Jeong et al. deployed a robotic coach in college dorms to improve students' well-being [15]. The robot provided positive psychology exercises and auxiliary activities to build rapport with its users. A plant robot was designed by Bhat et al. for behavioural activation reminders for young adults with depression [16].

Reviews [2, 12] discovered methodological weaknesses and limited or mixed evidence of the positive impact of SAR in mental health interventions. The results are, however, promising, and the difficulties in conducting strong studies in the field are attributed to its infancy and a range of barriers; these include characteristics of the social robot, for instance, the robot's weight, and technical limitations and issues, such as connection instability and deficiency of speech recognition [13, 2].

The integration of technology in mental healthcare, including the use of SARs, represents an innovative approach to augmenting traditional treatment and support methods. SARs hold promise in enhancing accessibility, mitigating stigma, and supplementing existing therapies [7]. Having mentioned the potential advantages, it is imperative to pivot our focus to the ethical considerations that emerge: ethical concerns identified in previous research include potential harms arising from malfunctions, reduced human contact, emotional deception, privacy, and data security issues [2, 7]. Taking a step further in ensuring technology's adequacy entails considering the system in which they would be

integrated: the introduction of novel technology can be disruptive, shifting responsibilities, roles, understandings, and expression of values integral to the system [92, 27, 93]. Previous research identified a lack of guidance in development, integration and training, the potential misuse of the technology to reduce the provision of mental health services, and the potential impact on care-related values [7]. Previous research deems it crucial to acknowledge that the development and integration of SARs in mental health care are influenced by the mutual shaping of society and technology [94]. This means that while these robots are shaped by societal values, norms, and needs, they, in turn, have the power to reshape aspects of society, including human values and social interactions. It is therefore of utmost importance to pay thorough consideration to social, cultural, and ethical aspects in the design and implementation of robots.

2.2 AI4SG-VSD

Building on the Value Sensitive Design (VSD) methodology, which aims at integrating values into the design process, the AI4SG-VSD framework expands VSD to address the unique challenges posed by AI [28]. VSD is a design approach that accounts for human values with the goal of integrating them into the design process. It is an iterative process composed of three types of investigations:

- conceptual, philosophical analysis of values and concepts under investigation;
- empirical, research on understandings, contexts, and experiences of stakeholders;
- technical, analysis or identification of mechanisms and designs to support particular values. [95]

VSD has been applied to the design of information systems, such as a web browser with a novel mechanism for cookies and informed consent [96], a screen displaying real-time outdoor scenes, and a simulation package for the prediction of urban development patterns [97]. It has more recently been applied to the design of an app targeting people with dementia [98]. In previous research, the identification of values has been carried out in a top-down approach, from lists of values redacted by researchers, bottom-up, eliciting them directly from stakeholders, or using a mixture of both [28]. VSD has received

criticism due to the lack of a clear normative methodology [27, 99, 100]. In light of this limitation and as an answer to the ethical issues regarding robotics in care, Wynsberghe proposed the Care Centered Value Sensitive Design [25], which integrates VSD with ethics of care. The framework includes analysis of context, practice, actors involved, type of robot, and manifestation of moral elements: responsibility, competence, reciprocity, and attentiveness, identified as fundamental values of care.

The AI4SG-VSD framework was applied to elderly care robots by Umbrello and Van de Poel [27] as an extension of the CCVSD, expanding the value sources to be considered in the design process to take into account challenges posed by AI:

- AI specific values from the EU HLEG on AI, namely human autonomy, prevention of harm, fairness, and explicability. These high-order values are translated into design norms using the AI4SG principles as described by Floridi et al. [101], as shown in Figure 1. A definition of the values and norms can be found in the following section.
- UN SDG are included to serve as orientation for socially desirable outcomes (Figure 2)
- Context-specific values

The AI4SG-VSD approach is composed of four phases, context analysis, value identification, design, and prototyping (Figure 3), and is meant to continue throughout the complete design process.

The AI4SG-VSD framework emphasizes the significance of understanding environments in which technology is designed and used. The context analysis consists of framing the societal challenges, addressing existing technology and systems and eliciting perspectives, values and needs of stakeholders. The second phase is value identification. This phase concerns the identification and conceptual exploration of a set of values relevant to the design and the context. Next, design requirements are formulated based on the context analysis and value identification. The authors suggest translating values into design requirements through a value hierarchy [28]: values are translated to norms, as illustrated in Figure 1. These norms can then be operationalised through design requirements. Lastly, the prototyping phase consists of building prototypes for testing based on the design requirements.

The following section outlines definitions of the norms and values that are integral to this framework.

Definition of values and norms

Values from the EU HLEG on AI:

- **Human Autonomy**: the right of a person to make rational decisions and moral choices, and be allowed to exercise their capacity for self-determination [102].
- **Prevention of harm**: the obligation not to harm individuals or groups [103].
- **Fairness**: Fairness is generally interpreted as fair, equitable, and appropriate treatment of persons [102]. In Floridi's account of the AI4People [103], fairness is understood as using AI to correct past wrongs such as eliminating unfair discrimination; Ensuring that the use of AI creates benefits that are shared (or at least shareable); Preventing the creation of new harms, such as the undermining of existing social structures.
- **Explainability**: it comprises intelligibility, meaning being able to understand and explain how a system works, and accountability [103].

AI4SG principles:

- **AI4SG#1 Falsifiability and Incremental Deployment**: the need for AI systems to be testable and falsifiable. It advocates for the gradual deployment of AI technologies, allowing for careful assessment and adjustment based on empirical evidence at each stage [101].
- **AI4SG#2 Safeguards Against Manipulation of Predictors**: This principle focuses on preventing AI systems from being controlled, particularly when it comes to the data or predictors they employ. It entails putting in place safeguards to verify the data's integrity and authenticity, ensuring that AI choices are based on correct and reliable information [101].
- **AI4SG#3 Receiver-Contextualized Intervention**: AI interventions should be adapted to the recipients' individual situations and needs. This principle advocates for the development of AI systems that take into account the specific circumstances and challenges of the target users, resulting in more effective and relevant solutions [101].

- **AI4SG#4 Receiver-Contextualized Explanation and Transparent Purposes**: AI systems should provide clear explanations of their processes and decisions to their users. Furthermore, the purpose of AI should be transparent, ensuring that users are fully informed and can trust the system [101].
- **AI4SG#5 Privacy Protection and Data Subject Consent**: This principle emphasizes the significance of safeguarding individual privacy in AI systems. It entails ensuring that personal data is handled responsibly, with the consent of the individuals to whom the data pertains, and in accordance with privacy legislation [101].
- **AI4SG#6 Situational Fairness**: AI systems should be developed and operated to be fair in a variety of contexts. This entails avoiding biases and ensuring that all users, regardless of their background or circumstances, are treated fairly [101].
- **AI4SG#7 Human-Friendly Semanticisation**: AI must allow people to foster their "semantic capital", which can be defined as content that can empower someone to give meaning and understand something [101].

Care Values:

- **Responsibility**: a willingness to respond and take care of need [27].
- **Attentiveness**: a proclivity to become aware of need. [27]
- **Reciprocity**: the care-receiver's capacity to guide the caregiver and the instauration of a reciprocal interaction [27].
- **Competence**: the skill of providing good and successful care [27].

Research Justification

The previous sections highlighted the urgency of developing better care for young adults' mental health, the scope and limitations of current technological tools and the potential of SARs. The current state of AI and robotics technology presents significant challenges [13, 2], particularly when engaging with vulnerable populations. Young adults' acceptance of technology-mediated mental health support [104] and rapid technological advancements, however, indicate research on the possible application of SARs for this population to be relevant and timely. SARs, thanks to the possibility of varied, natural interaction modalities, are versatile tools that open new avenues for technology-supported

mental health research. Given the importance of an empathetic alliance in therapeutic contexts [40], SARs' social capabilities may prove to be an effective medium for providing support to users. Due to the increased severity of consequences of malfunctions or misjudgments in mental health support, the adequacy of the technology must be ensured to promote safety and effective integration of robots in mental health care. The AI4SG-VSD framework offers a suitable methodology for exploring SARs potential in the field, focusing on integrating stakeholders' perspectives and values from the design phase. This research aims to contribute to Responsible research and innovation [26] in the field of mental health SARs by providing insights into the context, values and practices of stakeholders, indicating potential applications and features, with the goal of supporting design practices that respect and promote human values.

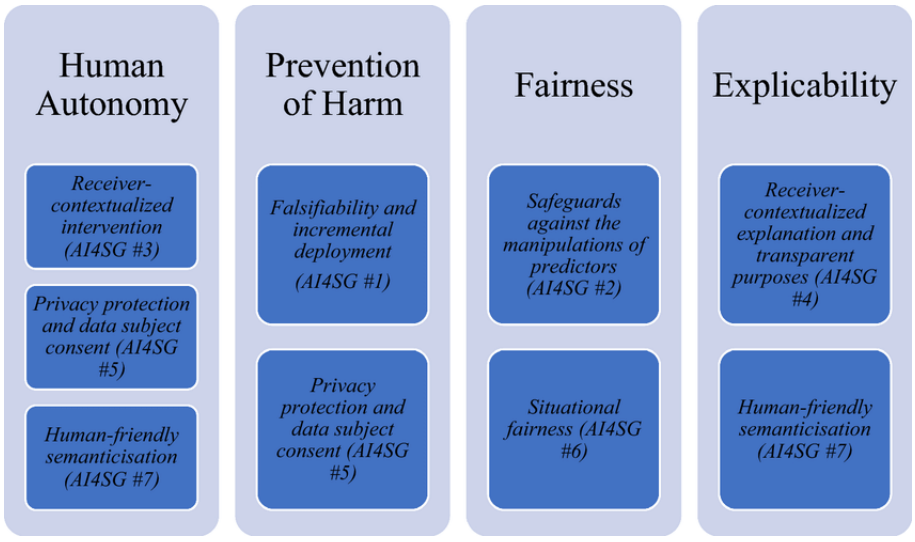

Figure 1: Relationship between higher-order values of the EU HLEG on AI and AI4SG norms. Source: Umbrello and van de Poel [28]

Figure 2: United Nations Sustainable Development Goals. Source United Nations [105]

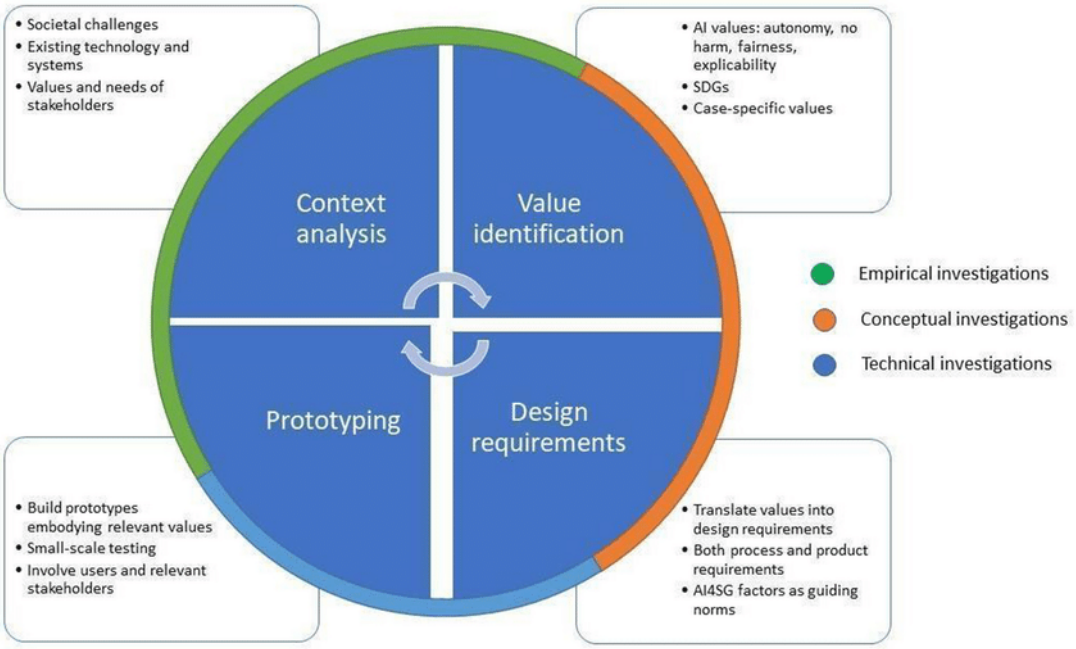

Figure 3: AI4SG-VSD design process. 2021 Source: Umbrello and van de Poel [28]

3 Methods

This research consists of an empirical and conceptual investigation in the field of mental healthcare robotics for young adults with emotional complaints, following the AI4SG-VSD design process as described by Umbrello et. al [27]. The framework is applied performing a context analysis, value identification and investigation, and providing design recommendations. As mentioned in the introduction of this research (1), the identification of potential applications is not part of the AI4SG-VSD framework. It is, however, one of the subjects addressed in this thesis. This is supported by the context-driven methodology of this research, which allows for the discovery of prospective applications based on context and value analysis, instead of being a pre-requisite for design. Furthermore, design recommendations are a modification of the design requirements phase of the AI4SG-VSD framework. This choice is in line with the exploratory nature of this research, which does not allow for defining strict requirements but is instead aimed at suggesting potential features and avenues for future research.

Following is an outline of the research process:

1. The context analysis was carried out first through related research, presented in 2.1, then through the analysis of interviews conducted to understand the context, elicit values and perspectives of two groups of stakeholders: psychology professionals and robotics researchers.
2. The analysed data is used in three ways: to expand the context analysis of section 2.1; to inform the identification of values in addition to the ones suggested in the framework [28], as outlined in section 2.2; to inform their conceptualisation. The values suggested by the AI4SG-VSD framework and included in this research include autonomy, prevention of harm, fairness, explainability, attentiveness, competence, responsibility and reciprocity. Definitions for those values can be found in section 2.2.
3. Potential applications and design recommendations are derived from insights from the context analysis and the conceptualisation of the selected values. Promising potential applications are selected through combining previous research, insights from the qualitative data and their potential impact on the values relevant to this research. Design recommendations are obtained through a translation of values into norms, as shown in Figure 1, and norms into design recommendations.

The following sections describe the methodology adopted for the acquisition and analysis of the qualitative data.

3.1 Data Acquisition and Analysis

3.1.1 The Interview Process

Participants were identified from the researcher's personal network and through online outreach. 13 participants were recruited: 5 psychologists, 1 researcher in organizational psychology, and 7 researchers in Human-Robot interaction. Every HRI researcher holds a PhD degree or higher and their background ranges from engineering, psychology, and public health. With reference to the interview transcripts (Appendix **??**) and coded segments (Appendix **??**), P1, P2, P3, P8, P9, P10 and P12 are HRI researchers. P4, P5, P6, P7 and P13 are psychologists, and P11 is a researcher in organizational psychology.
The interviews were conducted in person or via teleconference depending on logistics and availability. Participants were informed on the nature of the research, the type and purpose of the data collected, and consent was obtained prior to data collection following Utrecht University guidelines. The consent form can be found in Appendix A.

3.1.2 The Interview Structure

The interviews aimed to identify practice-specific challenges and to unveil context-specific understandings of care practices, the mental health crisis, the role of technology, and the potential of SARs, with the goal of answering research question 1b and gathering insights for the remaining research questions. The interviews were value-oriented and semi-structured [106]. They were informed by the VSD methodology and included stakeholder-generated value scenarios [107] elicited by interview questions. Value scenarios are tools used in VSD research that are used to explore imaginaries and anticipate the impact of technology on human values, enabling designers to create more ethically informed and socially responsible technological solutions [107].
Following the interview structure and motivation. The interview questions can be found in Appendix A.

- Personal background

  This topic addresses participants' motivations and difficulties related to their job.

Difficulties and limitations participants encounter in their profession are important aspects to enrich the understanding of the context and unveil either situations where technology might help or situations that might hinder the design and development of technology, contributing to answering.

- Mental health support practices

  This topic starts with an introduction on the current state of young adults' mental health and challenges related to the low accessibility of services. The introduction sets common ground with regard to the social issues relevant to this research. The value scenario involves identifying an ideal way to tackle the mental health gap previously introduced. Its aim is to understand the limitations of current systems and practices and the shape of a possible solution without boundaries. Identifying an ideal solution allows for reasoning outside the constraints of the current structure and points toward potential directions for new avenues. Probing important features of a carer aims at discovering important factors of care practices and those involved in them.

- Technology in mental health

  Personal attitudes and experiences with technology for mental health are relevant to this research in two ways: firstly, they unveil personal values with regards to the object in question, namely technology; secondly, they might yield insights on the integration process of these technologies, both from an individual perspective and a professional, systemic one.

- SARs in mental health

  This topic tackles directly some of the research questions' topics, investigating potential applications, design requirements and recommendations; questions regarding risks, benefits and considerations aim to unveil the values and opinions of participants.

  This topic includes a value scenario that uses the element of pervasiveness to lead the imagination of participants possibly beyond what is already discussed, inviting reflections on long-term implications of the potential application of SARs, on consequences, and on values in a fictive but concrete way.

3.1.3 Data Analysis

A coding manual was generated from reviewing part of the data, consisting of 5 interview transcripts, and the domain conceptualisation which informed the interview structure, as described by Friedman et al. [106]. 10 main themes were defined, each containing subcategories emerged during the coding process or pre-defined during the earlier stages of this research, namely the EU HLEG values: autonomy, fairness, explainability, and prevention of harm; and the CCSVD values: reciprocity, responsibility, competence, and attentiveness. The coding manual can be found in Appendix B. Table 1 presents, for each code and subcode, the number of participants who addressed that topic, "P", with the maximum amount corresponding to the number of participants, 13, and the percentage of occurrence of the topic relative to the total amount of coded segments "R", with the total amount being 454. The percentages were calculated with precision to one decimal place, employing a rounding methodology where figures were rounded down for values up to .05 and rounded up for values of .05 and above.

Name	**P**	**R**
1. Challenges in Research and Practice for HRI professionals	7	4.4%
1.1 Collaboration	4	1.5%
1.2 Systemic Issues	4	1.5%
1.3 Inconsistencies	1	1.3%
2. Challenges in Research and Practice for psychology professionals	6	3.1%
2.1 Difficulties of Therapy	5	1.8%
2.2 Institutionalisation and Funding	3	1.3%
3. Challenges of Applying SARs in Mental Health Support	13	14.9%
3.1 Feasible Integration	9	3.7%
3.2 Social Consequences	7	2.2%
3.3 Dependence	8	2%
3.4 Limited Human Connection	4	1.8%
3.5 Regulation and Safety	7	1.8%
3.6 Misinterpretations	7	1.5%
3.7 Technical Limitations	4	1.1%

Name	P	R
3.8 Assessment	4	0.9%
4. Scenarios and Roles for SARs in Mental Health Support	13	14.7%
4.1 Therapy Support	9	3.5%
4.2 In-between	7	2.4%
4.3 Therapist, friend, coach	6	2.6%
4.4 Skill-building	6	2%
4.5 Complementary	5	1.8%
4.6 Prevention	3	1.1%
4.7 Data Collection	3	0.7%
4.8 Crisis Intervention	2	0.7%
5. Benefits of SARs in Mental Health Support	13	7.9%
5.1 Accessibility	9	3.1%
5.2 Impact on People	7	3.1%
5.3 Availability	5	1.1%
5.4 Reduction of Provider's Burden	3	0.7%
6. Features and Design Requirements for SARs in Mental Health	11	12.1%
6.1 Communication	6	2.6%
6.2 Collaboration	5	2%
6.3 Personalisation	5	1.8%
6.4 User Education	6	1.5%
6.5 Appearance	3	1.3%
6.6 Privacy	3	1.1%
6.7 Accessibility	3	0.9%
6.8 Support and Monitoring	4	0.9%
7. Considerations about Technology in Mental Health	10	6.6%
7.1 Role	8	3.3%
7.2 Accessibility	6	2%
7.3 Different Tools work for Different People	5	1.3%
8. Considerations about Mental Health Support Practices	10	8.4%
8.1 Prevention	5	1.5%
8.2 Features of a Helper	6	1.5%
8.3 Therapeutic Relationship	6	1.3%

Name	P	R
8.4 Goal	6	1.3%
8.5 A Societal Issue	3	1.1%
8.6 Personalisation	3	0.9%
8.7 Accessibility	3	0.7%
9. EU HLEG Values	13	15.4%
9.1 Autonomy	13	5.5%
9.2 Prevention of Harm	11	4.8%
9.3 Explainability	8	2.6%
9.4 Fairness	6	2.4%
10. Care Values	13	12.3%
10.1 Competence	9	3.5%
10.2 Reciprocity	10	3.3%
10.3 Responsibility	8	2.9%
10.4 Attentiveness	8	2.6%

Table 1: Codes, subcodes and counts. P stands for the number of participants who addressed the topic (N=13) and R for the percentage relative to the total number of codes (N=454)

A subsection of 30% of the data was coded by a second researcher for determining intercoder-reliability. The second coder analysed the data autonomously, assigning codes to arbitrary sections of text based on the coding manual.

As it is often the case in the data of multiple coded sections or overlapping sections with different codes, no predefined segments were provided. Given the variability of the identified block of text, agreement was determined where the first and second coder sections overlapped by any measure The reliability score was computed with two metrics: first with Holsti's method [108], which can be applied to situations in which two coders code different units of the sample [109]. The score, calculated on the subcategory level, was computed as 68%, using the following formula:

$$X = \frac{2A}{(N1 + N2)}$$

where X is the score, A the number of agreements, N1 the number of decisions made by the first coder and N2 by the second. At the parent level, the score is 77%. Cohen's kappa

was also computed on each of the parent codes to provide a more nuanced evaluation of the inter-coder reliability which accounts for chance agreement, using the following formula:

$$k = \frac{P(a) - P(e)}{1 - P(e)}$$

Where P(a) represents the observed agreement and P(e) the chance agreement, meaning the probability of the two coders agreeing by chance [110]. The weighted average of those values was computed to provide a score that reflects the general inter-coder reliability, with the weights being the count of each code in the data. The weighted average of Cohen's kappa across all parent codes is 0.75. This value provides empirical support for sufficient agreement between the two coders [111].

4 Results

4.1 Challenges in Research and Practice for HRI professionals

4.1.1 Systemic Issues

Some researchers highlighted the difficulties related to carrying out interdisciplinary research in an academic setting. Because of the interdisciplinary nature of the field, carrying out good research requires knowledge from fields outside of one's expertise. Acquiring such knowledge, keeping up to date with current research, and applying it is time-consuming, and creates tension with the higher amount of publications necessary to succeed as a researcher in the field. In applications in the healthcare domain, this poses a bigger challenge, because of higher standards in designing, testing and deploying technology for vulnerable populations, which translate into longer timeframes. In contrast, in psychology or medicine departments, the amount of publications has a lower significance. There are different expectations from researchers of different fields, and this is also reflected in the way projects are funded: shorter projects in HRI/Computer Science (CS), and longer-term projects in psychology or medicine.

"I'm not sure if that's really supported by the research community in the way that we need to publish how fast it is going and you know quickly publishing is one of the prerequisites to get a Ph.D. it seems whereas if you want to be really good research it takes a lot of time then you might have a few publications which is important for your career thereafter and I think in other fields that might be less of a problem for instance in psychology it's more accepted if you just have one or two papers I think and I'd say medicine if you have one paper that's already a very good job I think" - P2, HRI researcher

"Their [Medicine researchers] projects are really typically done in a very long term, like four to five years, and that's very typical for them. Whereas in CS world, people are usually funded, they're shorter or more short timeline kind of projects. So it's just different expectations, right?" - P1, HRI researcher

One participant explained that there is fear regarding SARs in mental health which translates in difficulties in carrying out research.

"I think I will just say that one of the things we're up against, and I've seen this in reviewers for grants I've written and stuff is there is a there is a unfounded fear of robots in mental health

[...] I find it really ironic that I can submit like a proposal about a virtual reality environment that does this that and the other and the teen uses it by themselves and it helps them you know work on emotional regulation or whatever and I won't get a single comment that's like could that be dangerous, is it unhealthy [...] every single robot grant I've submitted there's been a like but what if the team becomes too attached to the robot, couldn't robots be unhealthy for teens, wouldn't this increase their addiction to technology. It's so strange to me, something about robots like triggers an alert system.." - P10, HRI researcher

4.1.2 Collaboration

Some participants indicated interdisciplinary collaboration as one of the difficulties to face in the profession. Human-Robot interaction is a highly interdisciplinary field. Projects involve the collaborations between engineers, philosophers, social scientists and other actors involved, for example healthcare providers. The difficulty arises from different ways of thinking, articulating and understanding of professionals from different fields, which generates friction in establishing common knowledge and understanding of priorities, tasks, and methodologies. Ownership was mentioned as a factor hindering collaboration between researchers of different groups.

"Everybody speaks a different language. - P8"

"It's challenging to find participants willing to help, to find venues to collect data, for example, in schools or hospitals, because it's not their primary goal, right? They have a set of goals and stuff they do on a daily basis. And then it seems like you're intruding, and they have to make room for you. So that's very difficult." - P9, HRI researcher

"...how can I say that this work is really mine how can I defend it in my thesis and how can I claim I am the expert when I actually collaborate with someone else." - P3, HRI researcher

4.1.3 Inconsistencies

One participant identified inconsistencies in the HRI field. First, they explained how the social aspect of HRI and the technical, experimental approach are bundled together when their objectives and approaches are very different. Secondly, they challenged the idea that humanoid robots are the right tool and the trend that robots need to be human-like:

they believe that pushing towards human-like robots will feed the fear of SARs, hindering current research which is not following that line.

"I think there's a huge unfounded bias that robots are to replace or to be human like." - P10, HRI researcher

4.2 Challenges in Research and Practice for Mental Health Professionals

4.2.1 Difficulties of Therapy

A few participants described difficulties inherent to therapy itself. Some participants pointed out the difficulty in balancing the family's wishes and the patient's wishes in therapy settings. Self-diagnosis was also indicated as an obstacle, narrowing the views of the patient, of the family and posing a barrier to the therapist's intervention. Another issue raised by a few participants is the difficulty of assessing whether an intervention works.

"Minors, so people under 18, children, adolescents. Difficulties, limitations is that you're not just providing support to the client but there's a big factor which is their family, their parents. They live with their family, they live with their parents. So if there's difficulty from them, then my support is limited for the client. " - P7, psychologist

"I think now it's like with all the evidence based treatments, it's like, okay, in this setting, it works, but it's different in the room. And can you really measure if something works? And I think that's why the money problem is there as well, but because with the healthcare, you have cancer, you can say, okay, this treatment, you have to get eight chemo, for example, and then work because we know that, but that's not how it really works with mental health." - P6, psychologist

4.2.2 Institutionalisation and Funding

A few participants explained that becoming a therapist is very challenging in many countries. There are limited spots for both education and internships. The challenging path proves to be a barrier for many wishing to embark on the profession. This is perceived as

poor policy making, given the long waiting lists of people waiting for treatment experienced in many regions. Furthermore, a few participants discussed the institutionalisation of mental healthcare as problematic: the way the system is built makes it so that it is difficult to receive appropriate funding.

"For example, in Germany, all I hear about from students is that you need to have 100% grades to study psychology. And they also have a problem with waiting lists, so why is that standing so high? It should be high, but there's also the problem that there's a lot of need, and there are a lot of passionate students. So for me, there needs to be more training, I guess." - P7, psychologist

4.3 Challenges of Applying Socially Assistive Robots in Mental Health Support

4.3.1 Feasible Integration

Some participants discussed the feasibility of applying SARs to support young adults' mental health, from the issues of integrating new technology into a system to whether the technology would bring value, and whether it is appropriate for the purpose and target population.

"In terms of healthcare workers, the technology and systems that they would need would need to be inter-operable. They would need to work with the systems that they already use because we see this often is that when we introduce new technologies, they disrupt existing workflows and they create a big burden on the healthcare workers in the system before they become efficient. And so if we introduce new technologies, they need to be, I mean, not only easy to use, but also they need to be compatible with the systems and the workflows that people are already using to try and reduce that workload." - P8, HRI researcher

"the question would be what kind of value would I be providing." - P2, HRI researcher

"sometimes I wonder like maybe like a phone-based interventions is more suitable for young adult population because they're they always have phones near them right it's more accessible." - P1, HRI researcher

4.3.2 Social Consequences

Introducing an interactive agent in someone's life raises concerns regarding the possible consequences on one's social sphere. From furthering isolation by decreasing one's perceived need for social interactions to decreasing one's social skills, applying the use of SARs to vulnerable populations needs careful consideration.

"Teaching you social skills that doesn't necessarily mean that your well-being is being improved so for instance it could be that whatever you learn doesn't generalize to the interaction with people it could be that the interaction is so filling that you start losing out on the interaction with people which again might have negative impacts on your well-being in general in life." - P2, HRI researcher

4.3.3 Dependence

Most participants expressed concern regarding the possibility of users developing dependence on the technology, feeling like they cannot function without it. In particular about the consequences of SARs not being available to their users anymore, due to for instance technical issues, and the possibility that young adults might become over-reliant on this technology and avoid furthering their mental health path with humans.

"I would worry that if a teenager now would learn to do some mental health exercises with the robot that they would be very afraid to do it with a human. And I know this is true because they are generally more reluctant to interactions with people, than maybe when I was a teenager, where you know that was the only the only option I had if I needed assistance, so I would worry that it would rely too much on this safety and kind of non-judgmental robots that also don't understand them completely." - P3, HRI researcher

"it has a risk that reducing autonomy in the sense that they could get addicted to this interaction with robots and then that would have negative consequence for the well-being." - P2

4.3.4 Limited Human Connection

SARs limited human connection was discussed by a few participants, wondering if it is possible to build a genuine connection with a robot, given the complexity of people and that the therapeutic relationship is the most important factor of successful therapy.

"The most important factor of successful therapy is the relationship between you and your therapist. And I'm just asking, you know, wondering, when you are fully aware that you have a robot in front of you, like, how well can you connect? It's that's just a question that I have. Maybe it's possible. I don't know, of course, but I think that would be a very important factor. Like, will the robot be able to build, like, genuine relationships?" - P4

4.3.5 Regulation and Safety

Some participants mentioned regulation, liability and accountability as challenges. Accountability was used as a justification for explainability: it is really important to know how the technology works for someone to be held accountable. The challenge of regulation spans many critical situations: data storage, privacy and storage, malfunctions liability, and intervention in critical situations, meaning who is responsible for intervening in the event the robot has detected warning signals. Additionally, whether the data should be shared with the health provider or the family of the patient under 18. Some participants also discussed the possibility of the technology being hacked.

"I think that's one of the same problems with sort of AI and chat bots in particular is that people think that they're real and that maybe on the other side there's a human, but a lot of people don't understand that it's not a human and so there's sort of that ethical concern as well about liability, so what happens if something goes wrong and the robot can't detect when something is going wrong and sort of flag that to maybe a mental health professional?" - P8

4.3.6 Misinterpretations

Misinterpreting the robot's intentions can cause significant issues: by not fully grasping the nature and inner workings of SARs, its behaviour can be misinterpreted causing distress. Vice-versa, the robot misinterpreting one's behaviour can have dire consequences in vulnerable populations.

"I could imagine that the situation comes worse, the mental situation with a person, if they do not feel understood" - P11

4.3.7 Technical Limitations

Technical limitations were given as a reason for the inadequacy of these tools at this moment in time. With regards to language models, value alignment was mentioned as one of

the challenges, and so was topic sensitivity and how the complex nature of communication and of people communicating makes language models' standards too low for interacting with vulnerable populations.

"I think language generation models generate quite a bit of hype recently and we start looking into developing those models specifically for certain kind of health care applications which is already something that's being done but before they're actually usable I think it's still a quite long way because right now they can really generate a text quite well but I think the value alignment problem is still quite large.." - P2

4.3.8 Assessment

A few participants raised the issue of assessment: given the multifaceted and contextual nature of mental health and well-being, they argued that assessing whether SARs will be more beneficial than harmful, and in which situations, will prove to be complicated. One participant generalised this as an issue in mental health: evidence-based treatments' success can be influenced by many factors and the assessment is not as clear as it is in other areas of healthcare.

"And I think it's very hard, like to really measure it works. That's very hard because even without robots, I think now it's like with all the evidence based treatments, it's like, okay, in this setting, it works, but it's different in the room. And can you really measure if something works?" - P6

4.4 Scenarios and Roles for Socially Assistive Robots in Mental Health Support

4.4.1 Therapy Support

Some participants proposed the idea of using SARs to aid already existing therapies, making them more variate and enjoyable, extending their scope to a different environment, for example at home, and fullfilling different or multiple roles. SARs could be useful tools for young adults to stay activated in between therapy sessions, for example assisting in completing homework given by the therapist or encouraging them to carry out certain activities. They could also be used to aid in delivering certain therapies, support treatment adherence, or enhance the telepresence of therapists for patients in rural areas or ones

that feel more at ease with a robot. SARs could be a useful tool for patients who need 24/7 availability from their therapist: they could ease the burden and and additionally allow patients to stay at home in situations where they are to be admitted inpatient.

"Yeah, maybe, I would like to try out and then also give it to the person to take it home. That would be like, I think I wouldn't like, uh, enjoy it as much if I only had it in my session. It also, it needs to broaden its therapeutic interventions. So it's a, it's a tool that you use in your therapy as an extension to your own settings, but they can have at home as well." - P13, psychologist

"What's just come to my mind spontaneously is people who have borderline personality disorder. It's well known that like the people who exclusively treat borderline personality disorder or this might just be a generalization but need to be available for their clients 24 seven, which is always very extreme and probably appropriate for the treatment. That would be effective, I think, for someone that really does for someone who, for example, is on suicide risk, someone that needs that 24 seven attention and they could have that in their home as opposed to being admitted to as an inpatient. " - P7, psychologist

"So maybe they, I don't know, maybe they go and see a therapist first. And then the therapist says, this is your homework for the week as a robot to help you, you know, do the homework that you need to be doing. Maybe it's cognitive restructuring. And then we're going to have a follow-up appointment in two weeks' time. I'd like you to share your experience. Okay. And then maybe the robot helps them with the sort of in between, the in between work, but doesn't sort of replace the human. " - P8, HRI researcher

4.4.2 In-between

Some participants suggested that SARs could be perceived as non-judgemental and less stigmatising to approach, therefore improving accessibility to mental health services. They could be used as a 'pathway to humans', to practice talking about one's problems and emotions to then, once comfortable, switch to opening up with humans. SARs could be adopted as an early step of a stepped-care approach: they could be used for the first intervention and, when no improvement is seen, switch to a therapist. SARs could be used in the screening process, in diagnostics, carrying out clinical interviews.

"..the robot could bridge the gap between starting a mental health practice therapy and having moments of mental health. So the robot could be at home of the person for a couple of weeks and could deliver interventions that are very short, like you know breathing exercises or any other type of intervention that would reduce stress, anxiety, depression." - P3, HRI researcher

"There are these huge barriers so could the robot become have a role in that pathway to help accelerate that person getting to another person or being more comfortable with another person and being able to access those resources. " - P10, HRI researcher

4.4.3 Therapist, Friend, Coach

A few participants proposed the use of SARs to deliver talking therapies. Some participants envisioned them as a stand-alone solutions in certain situations. Others saw SARs fulfilling the role of a friend or a coach, rather than a therapist.

"I think it could be a stand-alone solution, I think of it as sort of stepped-care, so you start with this robot, and if, for example, in one or one and a half month, it doesn't improve, you can also get a real therapist, and where you can first check, are your settings okay for your robot, like what was the problem, and otherwise, maybe something else is needed." - P13, psychologist

"With a therapist, you have to give advice that should be really, really valid. And again, let's say there's a technical strain, because again, that's a really very challenging situation to actually give proper advice. Rather than give a friendly advice, then you know, you can like take a bit of grain of salt, right?" - P12, HRI researcher

4.4.4 Skill Building

Skill building was mentioned by some participants as a potential application of SARs. The value of delivering education and practicing skills with a robot lies in the fact that they can be perceived as low stakes, non-judgemental, which could potentially lead to lowering the social stigma so that people, after practicing talking about one's emotions with a robot, might feel it is easier to talk about them with people. There is uncertainty about whether these skills can transfer to humans.

"..can we embed mental health practices long before the person's in crisis. They have some tools on board, they're used to talking about stuff you know so that's where I think something like our

educational robot could come in, of like practicing things, practicing some social things with somebody that's not gonna get mad at you or you could repeat the same silly exercise 25 times if you wanted to because it isn't a person and the stakes are very low. " - P10, HRI researcher

"So I think there would be a way of using the robot to essentially provide basic education and skills building around the way a person views their thoughts. And that can apply multiple situations, whether it's about the thoughts that they have following a breakup, or the thoughts that they have about themselves, or being able to manage disappointment and setbacks career-wise or study-wise. So there's various ways of, I think, that that could be useful. So I think probably along that psycho-education viewpoint of, yeah, developing awareness, developing mental health literacy, and probably starting with the way a person thinks about themselves in the world." - P5, psychologist

4.4.5 Humans and Robots: complementary roles

Some participants explained the complementary role robots should have to humans. Humans and robots have different skills, and robots are tools to support humans. They should not be thought of as replacing the human, nor designed for it. Robots, as tools with a certain human aspect, can prove to be useful in the field, with the potential to expand therapeutic options and efforts.

"...a human could just be a calming presence to be there or you know to hug you or to do something else which I don't think the robot can really do or shouldn't really do." - P3, HRI researcher

" [Humans are better at] ... all kinds of things I think, like assessing true risks right, utilizing resources right, like they can actually connect with other humans. I can talk to someone's family, I can reach out to someone's teachers and I think what humans actually want, this is why I'm like it's never robots or humans, what humans actually want is that human engaged, what a teen really wants is a counselor to hear them talk about how they want to harm themselves and to help them with that right. The robot can't really help you with that necessarily, not in the way that's really good for you so I think, again I think of the robot is like part of a pathway to humans." - P10, HRI researcher

4.4.6 Prevention

A few participants indicated prevention as an area where SARs could contribute. From providing knowledge and education, delivering evidence-based therapies, and checking in with young adults to detect worsening symptoms to observing and reflecting on group interactions, SARs could have become part of prevention practices.

"..it could be because it's more available maybe it could be a way to kind of check in like casually check in hey like how are you feeling today or um just to keep tabs on them [...] they can sort of like detect um earlier on like early signs [...] they have like these diagnosis right and you don't hit certain score you're not diagnosed with depression [...] they might not be eligible for certain services [...] so I think those like for those mild situations like it that could be a really good opportunities for social robots to kind of like catch people and then support them so they don't like dip down" - P1, HRI researcher

"Yeah, for the prevention, I think it's also very good to just provide knowledge, to see it as sort of a teacher. Yeah, and a robot could do that, too. " - P13, psychologist

4.4.7 Data Collection

Some participants explained how collecting data through the robots could be useful: for example to provide information to therapists and doctors, to deliver personalised interventions through data-driven insights.

"To be honest, there's a big chance I see that robots and technology in general, they can be very data driven, and they can record lots and lots of data that humans could not. So that data can always be recorded, it can be analyzed. And with that, the care of the person can also like the human care can be improved. Yeah. So for example, knowing like, this person always wants to take so much time to talk about this, or, I don't know, this person has these needs, or there's a pattern in their behavior here, you know, it can be really data driven, which is great." - P11, psychology researcher

4.4.8 Crisis Intervention

A few participants indicated the potential use of robots in detecting distress and promptly act upon it.

"...I think the robot would be it would be important for the robot to detect if the person it is living with it's coexisting in the home with um is feeling distressed if there is high levels of heartbeat or something that would tell the person is not well right now..." - P3, HRI researcher

4.5 Benefits of Socially Assistive Robots in Mental Health Support

4.5.1 Accessibility

Participants identified accessibility as one of the potential benefits of SARs in mental health. SARs could be perceived as non-judgemental and approachable by young adults, lowering the barrier to access support and expanding the range of interventions available.

"...I can imagine that for young adults the barrier to talking to a robot might be less and that's higher than talking to or actually taking a step to go to apply for mental health care and that this could be an in-between step to make that process easier" - P2, HRI researcher

4.5.2 Impact on People

Some participants discussed the nature of SARs, a tool with some human-like features, as a benefit in terms of engagement and range of potential applications in assisted therapy.

"no one is indifferent to robots and I think that's a very good leverage that we have when we work in HRI because we can really make a difference in terms of health, in terms of education if we use a robot because it's such a physical entity, it's an agent that interacts with you." - P3, HRI researcher

4.5.3 Availability

SARs potential availability could be a benefit for skill-practicing, in-between therapy session support and ongoing support for patients who need it.

"somebody that's not gonna get mad at you or you could repeat the same silly exercise 25 times if you wanted to because it isn't a person and the stakes are very low."

4.5.4 Reduction of Provider Burden

A few participants saw the potential of SARs to help with staff shortages and professionals' high workloads.

"[About the use of the robot for clinical interviews] And that would save the clinician a lot of time in terms of being able to have those more information gathering parts done and leaving the more technical things like skills intervention to the clinician." - P5, psychologist

4.6 Features and Design Requirements for Socially Assistive Robots in Mental Health Support

4.6.1 Communication

Participants discussed the desirable communication abilities of SARs from different points of view: robots should be able to communicate naturally and adapt the interaction to the setting and needs of their users, displaying situational awareness, the ability to recover from mistakes, and empathy. A few participants mentioned the importance of relatability to build connection, which could be achieved by feeding SARs pop culture.

"If I could build a robot for young adults, I think, I think also to be, because connection is very important in, in psychology. So to somehow be relatable, you want to also ask about your day and talk about Instagram and, you know, that, that builds connection and connection is very important, um, in therapy." - P7, psychologist

One participant explained how robots do not need to express emotion to be engaging:

"I like the idea that the robot doesn't ever express emotion. It can be empathetic without expressing emotion." - P10, HRI researcher

4.6.2 Collaboration

Collaboration was deemed a very important practice for successfully designing SARs, to ensure that the context, the problem, and the relevant expertise are integrated and the solution can provide value.

"I think therapists need to approve them. I think they should be developed with therapists and um basically the therapist and the computer scientists, whoever is implementing it in the robot, need to be very tight, very tight communication, um so that the mental health strategy for recovery is there, is present when the robot interacts with a human. So I think this is very important." (P3)

4.6.3 Personalisation

Participants suggested personalising the robot to one's preferences and needs as an important feature for engagement and efficacy, through data collection, therapist or user input.

"I think this is how you need to program your robot, and I think it should be programmed according to the person, so I think an intake should be done first, and then you adjust the settings." P12, psychologist.

P12 introduced the concept of continuous learning, which endows SARs with long-term memory, and explained its importance for personalisation:

"And again, personalization is very important. It also will make the person feel like they're heard and listened to instead of just an app that just tells the same thing every day to every person and has to be more personal, has to be referring to them and their experiences." - P12, HRI researcher

4.6.4 User Education

Some participants explained the importance of reminding users about the non-infallibility of the robot, how it works, its role, and its limitations. They also indicated that manuals are often non read by users, therefore the robot should provide information directly to them.

"I think it's also important the robot to explicitly tell the user what so to not make false expectations and to be very blunt about what it cannot and can do." - P3, HRI researcher

4.6.5 Privacy

A few participants indicated the importance of designing for privacy. They suggested the insertion of privacy protection mechanisms to shield users from potential access of their data from third parties and proposed the adoption of an easy way for users to delete personal data.

"And again, we talked before about sort of privacy and things like that, maybe having an option that the robot would be able to almost, I don't want to say shut down or hide when someone else comes into contact with it, but sort of be able to store that data that isn't easily accessible. So if

someone was, you know, sharing or doing sort of a CBT session with this robot, that someone that it wasn't being recorded and someone was able to easily sort of listen in and be exposed to that data." - P8, HRI researcher

4.6.6 Appearance

Some participants discussed the importance of the robot's appearance, mentioning a soft body and approachable features. A few participants suggested carrying out research to understand what young adults prefer and how their appearance affects the interaction.

"I'm afraid that developers of robots are thinking about functionalities and not, uh, as much as, um, uh, the look on the outside or the, the sympathy you create by attractiveness." - P13, psychologist

4.6.7 Support and Monitoring

A few participants stressed the importance of providing ongoing support throughout the testing and deployment of SARs through, for example, assigning someone the role of checking in and monitoring how the intervention is going. Another suggestion was the creation of a profession to support users with robots, much like other technical support available.

"also when implementing it, just providing that level of technical support and training and ongoing training as well that they might need." - P8, HRI researcher

"There should be some job created around supporting users that have robots and not just like problems with robots, but understanding robots." - P3, HRI researcher

4.6.8 Accessibility

Some participants mentioned the importance of accessibility and ease of use. This could include mechanisms to support different kinds of interaction based on the person's requirements or to recognise the user and operate without the need for technical knowledge.

"In terms of accessibility, there should also be options for that. So whether it has, you know, maybe a touch screen for people who can, I don't know, want to read it instead, maybe people who can't hear or, you know, has different accessibility options in terms of the speed of the

speech, the languages, I guess how loud it speaks as well, that would be hugely important." - P8, HRI researcher

4.7 Considerations about Technology in Mental Health

4.7.1 Role

Participants discussed the role of technology in mental health. Technology was envisioned as a tool to achieve autonomy and self-management, as a means to transition towards independence, as a temporary solution while waiting for in-person care, or as a medium for psycho education. It can be used to collect data, to support and strengthen existing therapies.

"I'm not sure if it can be a like long-term solution. How I envision it is that maybe during that waiting time, you know, it could be a temporary solution so that the bridge, I mean, that's already enough, but with the more severe problems, they will eventually need to see someone in person. " - P4, psychologist

"I think there needs to be, there needs to be sort of a hybrid approach to care where there's people providing the care, you know, these therapists and qualified people, but then also supported by technology. And I think using, I always say using existing technologies is sort of easier because people already know them. So things like smartphone apps or computers or things we have lying around and that young people would know anyway, and just having those to sort of follow up and collect data for monitoring. " - P8, HRI researcher

4.7.2 Accessibility & Transparency

Accessibility, meaning ease of use, affordability and availability, and transparency were regarded as important features. Some participants mentioned the importance of indicating evidence-based therapies from other kinds of interventions and the importance of expertise in building technology for mental health.

"It should be very easy, uh, adjusted to education level, because I often think we are making therapies, um, for people who are educated a bit more, a bit better. So I think, uh, that would be good. Easy to use. User friendly. User friendly and very, uh, you don't have to think about how it works. Intuitive. Um, and so easier than we do it now, because I think, uh, therapy is more

open to high educated people. I think, uh, definitely all the online materials should be for free. And I think they need to have, uh, a label to, to make the distinction between.. How do you say, uh, research based therapies? " - P13, psychologist

4.7.3 Different Tools Work for Different People

Some participants discussed how certain tools work for some people but not for others. Cultural differences were mentioned, and so were individual preferences and attitudes, hence the importance of personalising tools and interventions to their individual and their context.

And I think it's difficult to say whether or not the technologies are working now because maybe they're working for some people, but they're not working for other people. And across different healthcare systems as well globally, it's really difficult to sort of understand whether technology would work because of different cultures essentially, but also different treatment pathways. " - P8, HRI researcher

"I mean, it's not a one-size-fits-all, I think. One of the things that with digital mental health, in particular, people often, I think early on, thought, well, this is going to be what happens. People are just going to do all their mental health by digital means. And that's not the case." - P5, psychologist

4.8 Considerations about Mental Health Support Practices

4.8.1 Prevention

Some participants viewed prevention as an area to invest in and allocate resources. Spreading knowledge and awareness from an early age and in general to society was described as an effective way to ameliorate the mental health situation.

"I think maybe on the preventative side, that's where more resources need to be put, because it's usually a lot more accessible, shorter waiting times, shorter therapy also. But I feel like all of the money goes into when it already has become a problem rather than just preventing it." - P4, psychologist

4.8.2 Features of a helper

Discussing features someone who is able to help young adults should possess, participants mentioned genuinity, attunement, and being relatable. Furthermore empathetic, honest, confronting, fun and non-judgemental. They should come across as accepting and normalize the young adult's issues.

"So make it active while talking and then using also fun. And I think it will help to use it like this, because you will face many, many problems in your life. And that it doesn't become stereotype as a problem is always very heavy. So to have also to try have fun around it." - P13, psychologist

"So, and a lot of people think they're not normal when they have anxiety or etc. And I think the first step is like normalizing that because it's okay to ask for help." - P6, psychologist

4.8.3 Therapeutic Relationship

Some participants pointed to the therapeutic relationship as the foundation for successful therapy.

"I think the most important part of therapy, because you can just do like a protocol, but yeah. And of course that is evidence based. Yeah. But it's not the real world. So do you know what I mean? You know, I think like that's, that's the most important thing. Like first make a connection and is there like a click? " - P6, psychologist

4.8.4 Goal of Therapy

A few participants explained that the goal of therapy is for patients to detach from the therapist and reach a higher level of autonomy and empowerment.

"I don't want them to rely on me. I want to give them tools and help them practice tools that then they can use themselves. So they become independent in dealing with their own issues." - P7, psychologist

"To help people understand themselves. And to feel that they have their own power to improve, to change, to choose. Instead of the feeling that they are sort of entrapped in a system, a family. So empowering them to just become." - P13, psychologist

4.8.5 A Societal Issue

A few participants talked about mental health as a societal issue: it should be part of daily lives and it should be everyone's responsibility.

"it's a community societal issue. We can change how we practice mental health we can increase access, we can do all these things over here but I promise you it's not gonna have a real impact unless we also do all the things over here, which means like literally changing how schools function, changing parent education, creating I think of it as like it's everyone's responsibility to improve teen mental health. " - P10, HRI researcher

4.8.6 Personalisation

Some participants explained how mental health treatments should be personalised to one's needs and preferences, and responsive, there is no one-size-fits-all.

"...we know even with therapies that have very good results, like CBT, it's not going to be right for everyone, so it needs to be tailorable to the individual and responsive to whether it's working for them or not. " - P5, psychologist

4.8.7 Accessibility

The difficulty in promptly finding the right kind of care was described as an issue impacting accessibility; so was the type of care offered.

"...therapy is more open to high educated people" - P13, psychologist

4.9 AI4SG Values

4.9.1 Autonomy, dependence

Some participants discussed the potential impact of SARs on autonomy in terms of over-reliance and dependence: a few raised the question of the consequences of the technology not being available anymore; others that young adults might fail to make the switch to humans, becoming over-reliant on technology and potentially furthering isolation; some drew the parallel of avoiding becoming dependent on your therapists to robot.

"it has a risk that reducing autonomy in the sense that they could get addicted to this interaction with robots and then that would have negative consequence for the well-being.

another aspect so in that sense I have less autonomy in choosing who they want to interact with because this interaction is so fulfilling that they neglect all kinds of interactions so that could be a big risk. you also have the risk for autonomy in the sense that a machine is telling you to do something. " - P2, HRI researcher

"Well, I think that's the same with any type of therapy, right? It's a process, and in fact, there's a part of the process where people feel like they cannot do stuff on their own without asking the therapist. So I think that needs to be also integrated. " - P9, HRI researcher

"You should not rely on me, right? And I think the same should then happen to the robot because otherwise you might be really reliant on the machine and not be autonomous anymore in your life and just do things by yourself." - P11, psychology researcher

The impact on autonomy was described by some participants as personal and contextual: what for some might be autonomy enhancing and promote their well-being, it could have the opposite effect on someone else.

"...we don't want a person reliant on something, we want them to be self-reliant, but at the same time, it could be a good stepping stone in terms of, you know, if that person wasn't leaving the house anyway, then if they leave the house with their item, that's a good step in the right direction. So I think in terms of autonomy, it would depend on the individual in terms of whether that's actually, you know, reducing it, or whether this is a step up from the previous functioning that they were at. " - P5, psychologist

"...there could be certainly people who for whom that exercise with the robot is so satisfying and they can never make the jump to the human. I'm still not sure that's a negative because maybe they were never going to jump to that conversation with a human anyway" - P10, HRI researcher

Autonomy was also discussed in terms of giving people options to choose from regarding mental health interventions.

"...people should always still have the choice in what they're doing." - P5, psychologist

A few participants indicated autonomy as one of the goals of therapy and adulthood; SARs could help young adults with regards to autonomy, when programmed accordingly.

"I hope the robot also realizes that autonomy is like one of the goals of adultship, so it needs to like also encourage this, by this and this are your options, but whatever you choose, you choose, and all the choices you make are good, something like this." - P13, psychologist

4.9.2 Prevention of harm

Responsibility and liability issues were discussed by a few participants. The importance of expertise and collaboration in building these technologies was discussed as a way to prevent harm, and so was the importance of keeping someone in the loop to monitor and check in with the users.

"Even if you make money from the app, I'm not against that. But even if you are doing it for profit, it still needs to be very carefully considered." ... "At the end, the people developing it should be the ones really concerned about it. Who's going to regulate it? Yeah, difficult question." - P9, HRI researcher

Some participants talked about the risk of the technology being hacked or data used in a non-transparent way, and the fear surrounding the possibility of it happening.

[About confidentiality] "I think that's a big part because we had like a code, you know, how do you say that? A conduct code. Um, and of course you can do something like that with a robot as well, but it's like it is technology and it can be hacked. And yeah, where is the safety of that." - P6, psychologist

Some participants worried of the possible consequences of assigning human-like machines to interact with vulnerable populations: while it could prove to be an effective way of tackling the mental health treatment gap, it could also be considered a form of deception, it could further isolation, create dependency and cause harm when malfunctioning or unavailable.

" But all the efficiency in the world won't bring you anything if you're not effective. So you need to do it sustainably. That's a challenge to overcome for sure. We could risk that people feel even more isolated if the robot is not human-like enough. So it's really like a weird device that you don't really feel attached to. It could be a risk to feel even more alone." - P11, psychology researcher

"we are kind of playing with like an ancient instinct of the human to be in a collective group, to be social. But not really because you are interacting with the machine and not with other humans. So can you actually get all the benefits from interacting with another human?" - P11, psychology researcher

"when we're trying to do what humans are doing in like counseling sessions and digging deeper into emotions and whatever I think we're way out of scope for for a healthy interaction, only because think about like triggering stuff and whatever like there's then there's nobody there if you're suddenly triggered and having a panic attack and there's nobody really with you" - P13, psychologist

The current limitations of large language models were also touched upon as potentially harmful:

"they can actually suggest very negative things to do to the person. And it's always you need to be really careful and have these filters to overcome any such situation. But of course, these filters might be very limiting as well that if you want to talk about your mental health." - P12, HRI researcher

4.9.3 Explainability

Some participants explained that one does not need to know the inner workings of a technology to accept it; users should however be informed about what robots can or can not do and there should be transparency with regards to personal data.

"I think it's also important the robot to explicitly tell the user what so to not make false expectations and to be very blunt about what it cannot and can do." .. "it should be important for the user to know where the data goes, what is happening with what they're saying to the robot and what they're doing together because this is a very vulnerable situation to be under a mental health therapy."

"I think every technology that is created needs to clearly convey its limitations and I think that even just things like smartphone apps, wearables, you know, everything that we create including robots needs to be very clear that this is not a healthcare worker, this is not a person, this is a technology that can help support the treatment or delivery of care but it can never replace a person and so if in doubt, you need to talk to a healthcare provider." - P8, HRI researcher

One participant, on the other hand, explained how it is not necessary to stress the fact that robots are not human:

"I think they [people] choose to interpret such things as human behavior, like we do with pets, for example. And I don't think that's necessarily bad. Unless it would be to an extreme where people really do not understand that the robot is not a person, but I think that's very unlikely." - P9, HRI researcher

A few participants thought is important to know how the technology works, to create awareness and for accountability reasons.

" I think it's important to know how it works. I think also always with big, you know, improvements technology wise, there's like a big group of skeptical people. So I can imagine parents being skeptical. So I think just to put people at ease with good information, it's important to know how it exactly works, but also your it's human lives that we're talking about, you know, so it's also really regarding that's really important to know how it works because there's like, yeah, accountability, I think. " - P4, psychologist

For those users who wish to know more about the technology, there should be an easy way for users to access information, either through the robot, the health professional or a support team dedicated to answering users' inquiries:

"There should be some job created around supporting users that have robots and not just like problems with robots, but understanding robots." - P3, HRI researcher

4.9.4 Fairness

Participants discussed fairness from different points of view: from algorithmic bias to degrees of accessibility, affordability, and suitability of SARs, to a teleological view on fairness: if it works it is better than nothing.

"If we train all that bias in the data, if we train machines to kind of interact with all kinds of human beings, we need to make a diverse picture and not just real white, cis male kind of situation, I think. Yeah. So it's a matter of fairness. We have to work on fairness." - P11, psychology researcher

"if it is built in a informed way with clinicians then I think it would be a fair thing to have. It would be better than nothing." - P3, HRI researcher

" .. "the ultimate goal is people's well-being. So whatever is faster and whatever works" - P7, psychologist

4.10 CCVSD Values

4.10.1 Competence

Some participants stressed the importance of relying on expertise in building these systems, to ensure their use and interventions are appropriate and provide value. A few participants discussed using SARs to supplement therapy, expanding the ability of healthcare providers to deliver care. A few participants mentioned that data-driven insights could help deliver better care. On the other hand, some participants explained that current technical limitations make the technology inadequate to provide care.

"..it needs to, I guess, really be informed by evidence and theory and developed with, you know, expertise.." - P5, psychologist

4.10.2 Reciprocity

Some participants explained how the therapeutic relationship between therapist and patient is the foundation of successful therapy, with genuinity, empathy, relatability and honesty being some of the features contributing to a successful connection. In these terms, SARs limited human connection was seen by some participants as a reason to doubt their potential: they expressed uncertainty regarding the possibility of building a genuine connection with a robot. On the other hand, a few participants saw the potential of SARs interactive capabilities in aiding assisted therapy by building a connection with their users.

"And I think the, uh, how the robot is adding its value, it's that you, that it's, it's giving you something. Also, if you don't ask for it. Um, and I think that is a very, uh, beneficial option for people and for young adults who are not easy to activate. So I think it, if it works a bit like a dog. I hope. So people with a dog, they have to go out every day to like walk their dog. Blah, blah, blah. And I think you feel sort of also a bit of responsibility maybe for your robot, which gives you value of, you can take care of something else. Um, and maybe that's also the bond you create, like you care for the robot and the robot cares for you." - P13, psychologist

4.10.3 Responsibility

Responsibility was touched upon by some participants discussing personal data regulation, malfunctions liability, and accountability issues. Making sure that users are aware of the scope and limitations of SARs also contributes to the responsibility of care practices.

"Risks again, ethical things about who is able to access that data. Again, if the data or an algorithm or whatever is showing that there's an increased severity, who's liable to respond to that? Is there enough health or are there enough healthcare professionals who can react when things or maybe when the data is suggesting that someone is maybe likely to relapse, so if there's no health services available to catch that person, then that becomes a huge ethical issue." - P8, HRI researcher

4.10.4 Attentiveness

In mental healthcare practices, some participants explained the importance of personalising treatments to one's needs and making sure they are responsive. Attentiveness was also discussed by some participants in terms of endowing SARs with situational awareness and the ability to detect distress. Moreover, care practices could see an improvement in attentiveness thanks to personalisation to users' needs and to SARs availability; if used to supplement therapy, SARs could attend to users' needs in-between therapy sessions. On the other hand, SARs misinterpreting users or failing to detect certain situations could be detrimental to the attentiveness of the care practice.

..it would be important for the robot to detect if the person it is living with it's coexisting in the home with um is feeling distressed if there is high levels of heartbeat or something that would tell the person is not well right now" - P3, HRI researcher

4.11 Results Summary

Interdisciplinary collaboration, systemic issues and inconsistencies within the field were indicated by HRI researchers as challenges. Psychology professionals explained certain difficulties inherent to therapy and the problematic institutionalisation and funding in mental healthcare. Participants identified several challenges with regards to applying SARs in mental healthcare: feasible integration to current practices, possible consequences on users' social sphere, risk of developing dependence, issues arising from the limited human

capabilities of robots, regulation and safety, negative consequences of misinterpretations, technical limitations, and assessment. Participants identified therapy support as a potential area of application, followed by positioning SARs as an access point for mental health services. Some participants envisioned SARs as therapists, friends, or coaches. Skill-building was identified as a further potential application, as were prevention, data collection, and crisis interventions. Some participants highlighted the importance of envisioning SARs as tools, to complement human efforts. The benefits identified by participants range from accessibility, impact on people, and availability to reduction of provider burden. When discussing SARs features and capabilities, communication-related features were the most prevalent, followed by an emphasis on the importance of interdisciplinary collaboration in design, personalisation, user education, appearance and privacy mechanisms. Furthermore, features addressing ease of use and accessibility, and the importance of maintaining human supervision and facilitating feedback and support were regarded as important. Personalisation and accessibility were mentioned as important factors in technology and mental health practices in general. Prevention was discussed as an area to invest in. The therapeutic relationship was mentioned by some participants as the most influential factor in therapy success. Mental health was discussed as a societal issue, requiring efforts from multiple sources. The potential impact of SARs on users' autonomy was discussed as personal and contextual, pointing to the risk of developing dependence. Supervision and collaboration were indicated as important with regard to harm prevention. The risk of data being hacked, of malfunctions and deception which may further isolation were mentioned as potential sources of harm. Explainability was addressed in terms of transparency, accountability, and the importance and degree of user education. The potential impact of SARs on care values was discussed as both supporting or enhancing, and detrimental and harmful.

5 Discussion

Despite the high prevalence of mental health issues among young adults, the effectiveness of existing mental health services for this age group remains limited. Technological innovations, such as SARs, emerge as a promising avenue to address these shortcomings, potentially enhancing accessibility and reducing stigma. However, these innovations

also bring forth ethical concerns that must be carefully considered. This research aims at initiating a value-sensitive design process with the goal of exploring the potential of SARs for mental health support of young adults with emotional complaints, following the AI4SG-VSD framework [28] and supporting a Responsible Research and Innovation approach [26]. This paper presents the results of an empirical investigation, consisting of the collection and analysis of qualitative data from interviews with representatives of selected groups of stakeholders, namely Human-Robot interaction researchers and psychology professionals. The data collected was used to elicit difficulties, understandings, perspectives, and values of stakeholders, to enrich the understanding of the context, inform a conceptual investigation of values, the identification of potential areas of applications and design recommendations. The discussion section of this thesis is structured as follows:

- Context Analysis - Part Two [5.1]: this section presents and offers an interpretation of the main results of this research, expanding the context analysis of section 2.1 with insights from the empirical investigation, answering research question 1b.
- Identification and Conceptual Investigation of Values [5.2]: this section answers research question 2: "What are the values to promote and respect when designing SARs for young adults' mental health support, and how are these conceptualised?"
- Potential Applications [5.3]: this section answers research question 3: "In which scenarios can SARs support young adults with emotional complaints?"
- Design Recommendations [5.4]: this section answers research question 4: "What are design recommendations and insights for future research?"

5.1 Context Analysis - Part Two

The qualitative data collected in this research reveals insights into the challenges experienced by HRI researchers and mental health professionals in their job, on the mental health crisis, current practices, and on the potential of SARs in the field. The following sub-section offers an outline and interpretations of the main findings.

5.1.1 Challenges in Research and Practice for HRI and Mental Health professionals

Researchers indicated systemic issues as posing difficulties in carrying out their jobs, ranging from the pressure of publication to a lack of resources for conducting quality inter-disciplinary research. This obstacle was deemed greater in applications in the healthcare domain: because of the end-users being vulnerable populations, the demanded standards for technology and interventions are higher, as the possible consequences of mishappenings are more severe. These findings are in line with previous research, which voiced the need for adequate funding to foster quality inter-disciplinary projects, proposing a change in the incentives driving the field and the publication process [112]. Another issue raised was collaboration between stakeholders and professionals with different expertise. Collaboration was also mentioned by participants as one of the main design requirements for SARs, necessary to build trustworthy systems. These findings indicate the necessity of efforts towards facilitating and encouraging collaboration and the factors that contribute to it. Issues and suggestions related to inter-disciplinary collaborations in HRI can be found in previous research, although limited and mostly consisting of conference papers [113, 114, 115, 116]. These insights underscore the need for further research on the topic and policy changes that foster collaborative environments in academic and research institutions. One participant described inconsistencies within the HRI field, questioning the feasibility of bundling together the technical approach and the social, human-centered approach, with widely different objectives and approaches. This division in approaches was observed by Gooding et. al [117] with regard to algorithmic technology in mental healthcare, which supports the claim and invites further consideration. Another inconsistency identified in this research is the unjustified utilisation of humanoid robots, taking for granted their suitability and legitimacy. A scoping review from Guemghar et al. [2] states that humanoid SARs show the highest levels of acceptability and usability among participants, referring to studies with older adults. This indicates that while humanoid SARs may be a suitable tool for older adults, their use is not always justified. Before assuming the applicability of humanoid robots, it is vital to critically analyze the context and specific needs of the intended user group. Future research should include an assessment of the effectiveness and user acceptance of selected SARs for the target population and setting to ensure their appropriate and justified use. These reflections suggest that

questions of suitability and legitimacy should be incorporated into research from the design phase.

Psychology professionals voiced difficulties arising from institutionalisation and funding of mental healthcare, often inadequate in addressing needs and providing prompt assistance, blaming bureaucracy for its lack of expertise and complex decision-making structure. These challenges are particularly pronounced in youth mental healthcare, where the decentralization of services has resulted in inconsistent access across different geographical areas. As noted by Ronis et al. [118], this decentralization, coupled with complex referral systems, creates additional barriers for young people seeking mental health support. This highlights the importance of improving mental health service accessibility, particularly for vulnerable populations such as young adults, in order to provide fair and timely access to mental health treatments. Other difficulties expressed by therapists were inherent to providing therapy: aligning the family and patient's will and assessing interventions are a few examples. The limitations of the current system and the difficulties of therapy expressed by participants highlight the importance of focusing on the feasibility of the integration of technology in current practices and systems: technology does not only need to be effective, it also needs to ease the burden of mental health institutions and professionals, or at least avoid adding to it. Previous research found a lack of focus on integration in HRI studies [7, 93], which corroborates the claim and calls for more efforts toward the issue.

5.1.2 Considerations about Mental Health and Mental Healthcare

Prevention was identified as an area to invest in and allocate resources to increasing literacy from a young age and endowing individuals with the skills to open up and express themselves, together with lowering the stigma through normalising behaviours and conditions related to mental health and mental healthcare. Prevention is also deemed an important area to allocate resources to in literature [119]. The emphasis placed on prevention by the participants and by research underscores its critical role in this field, which justifies investigating the potential role of SARs in support of preventative efforts. These strategies encompass early identification, mental health education, and the development of resilience and coping mechanisms, particularly among young adults, with the goal of curbing the onset of mental health issues. [50, 120]

The mental healthcare crisis was described by some participants as a societal issue, requir-

ing societal restructuring and revision of current systems, such as the education system, and the development or adaptation of current practices towards more accessible and personalised to young people. The current crisis was seen as reflecting the inadequacy of current practices and societal structures to promote and support mental health in the younger population, as also highlighted in research [119]. These findings suggest a need to investigate and understand mental well-being from different points of view, integrating the current cultural and societal context with insights from a multi-disciplinary stance, with the goal of empowering individuals, communities, and institutions with mechanisms and skill-sets to promote well-being. Evidence suggests that the responsibility for promoting and preventing mental health issues cannot be realistically confined to mental health professionals alone [50].

As previous research highlights [19], participants saw accessibility of services as problematic, in a system where finding the right type of care can be a complex process, and the type of care offered is oriented more towards educated individuals. Evidence of inequity in mental health services use by education level was found by Steele et al. [121]. This points to the pressing need for reforms in the mental health care system to enhance accessibility and inclusivity.

A few participants described the pivotal role of the therapeutic relationship for therapy success, widely discussed in research [44, 122, 123], and described what characteristics are important when dealing with young adults, for example relatability, attunement and empathy. Participants were divided on whether these human qualities should or could be effectively incorporated into robots for therapeutic purposes. The discussion points toward a need for ongoing dialogue and research to carefully assess the feasibility and impact of such design choices, with the goal of safeguarding and supporting therapeutic relationships in youth mental health practices.

Furthermore, fitting the treatment and outcome measures to patients' needs and preferences was deemed crucial by some participants of this research, as it can enhance treatment initiation and outcomes [124].

Personalisation was described by some participants as an integral component of delivering youth mental healthcare, as also supported in research [125], and it extends to technology for mental healthcare: there is no one-size-fits-all. Personalisation in digital mental health is a well-discussed approach, aimed at enhancing patient adherence to treatment

protocols and improving clinical outcomes. However, there are still questions about what constitutes effective personalisation, how it is currently implemented in practice, and the specific benefits it offers [126]. These findings highlight the need for a deeper understanding and development of personalization strategies for SARs for youth mental health.

5.1.3 Considerations about SARs

The data analysis revealed a multitude of challenges, benefits and considerations about the potential use of SARs in mental healthcare. Benefits include improved accessibility, potentially reducing the barrier to accessing mental health services, the impact SARs have on people, in particular with regards to engagement, increased availability and reduction of the burden of healthcare providers. These findings are in line with benefits identified in previous research [7, 2].

Participants identified several use cases: supporting existing therapy practices through activities that benefit from SARs availability, such as in-between sessions support where the robot could potentially fulfill different roles; SARs could also be positioned as an entry point for mental health services: their presence might be perceived by young adults as less stigmatising to approach; their availability could also be exploited for learning and practicing skills of various nature, or to carry out preventative interventions. The potential of SARs for skill-learning and to support therapy was also identified in previous research [11, 2]. A preliminary study on delivering micro-interventions through robots to adolescents was carried out by Alves-Oliveira et al. [127]. Previous research has also suggested finding applications for populations with difficulty accessing care [2, 1]. Fiske et al. proposed using AI applications as an entry point to mental health services for those populations, in line with the findings of this research. Some participants highlighted the importance of conceptualising and using SARs as tools, to complement the human counterparts, and not as a stand-alone solution. A similar position can be found in literature [7, 2]. On the other hand, a few participants described a potential use-case of SARs as a stand-alone solution, for example in the first levels of a stepped-care approach: if the robot is not able to respond appropriately to one's needs, further help should be provided. This idea is similar to the one proposed by Fiske et al. [7], where the authors suggest applying technology to support mild cases of depression. Lastly, some participants viewed SARs features as useful to collect and analyse data, with the goal of providing better care, as

also discussed in [11]. The alignment between this research's findings and previous literature warrants further exploration of these applications of SARs for youth mental health. Some participants elaborated on the kind of features and design requirements which should be integrated in the technology: communication-related features, collaboration amongst experts and stakeholders, user education, personalisation, continuous support and monitoring of users, privacy-promoting mechanisms, accessibility-oriented features and attention to the appearance of the robot. All these features point towards the safeguarding of patients' safety, well-being, preferences, and awareness. Previous research highlights the importance of collaboration between roboticists and psychologists in research on SARs in mental health [11, 1]. Additionally, some of the features mentioned were identified as facilitators to the implementation of SARs in mental health facilities, namely appearance, personalisation, user education, and supervision [2], which supports the validity of the features identified in this research.

Participants identified several categories of challenges, ranging from discussing SARs feasible integration in current, existing systems, to questions of the legitimacy of its potential applications in young adults' mental health support, and the inadequacy of current technology to carry out tasks appropriately. Challenges also included potential harms to their users, exacerbated by the vulnerability of the target population of this study: risk of developing dependence, negative impact on social skills and social engagement, SARs limited human connection, which calls into question the possibility of building rapport with a robot, and distress caused by misinterpretations. These findings are in line with previous research [7, 92]. Looming over the potential harms to users, next to the vulnerability of the population under scrutiny, is the issue of assessment: given the complexity of the sphere we refer to as mental health, given the numerous factors impacting the success or failure of interventions and, furthermore, the complications triggered by the introduction of a social machine interacting with vulnerable populations, assessing whether SARs could be beneficial or not is problematic. With regards to mental health, different paradigms, ranging from biomedical models to psychosocial approaches, offer diverse perspectives, leading to varying definitions, methodologies, and treatments [128, 129, 29]. This plurality adds complexity to the assessment of treatment success, as metrics for success can differ significantly across paradigms. For instance, while some approaches may prioritize symptom reduction, others might focus on personal recovery outcomes [130]. Assessment

is problematic also in HRI: it has been brought forward in several papers with regards to the methodological weaknesses of HRI studies [131, 112]. These considerations strengthen the need for more adequate funding, allowing for longer-term, interdisciplinary collaborative projects.

The ambiguity of the potential effects of SARs is also reflected in the implications for the values relevant to this research; besides the values selected through the AI4SG-VSD methods, these findings suggest collaboration and legitimacy as additional values. Collaboration frequently recurs in the results of this research both as a difficulty encountered by HRI professionals in their jobs and as an important requirement for the design of SARs. Given the complexity of mental health described by some participants and the breadth of the challenges identified, combining different expertise seems crucial to promoting well-being; furthermore, the framing of mental health as a societal issue strengthens this claim, calling for a collaborative and inter-disciplinary effort to tackle the mental health crisis. The issue of legitimacy was raised in different instances: one participant questioned the legitimacy of the widespread use of humanoid robots in research and practice; a few participants expressed doubts on the legitimacy of applying SARs in the field of young adults' mental health, questioning the appropriateness of the tool and warning about the drivers behind technological innovation. Taking these considerations into account, exploring legitimacy includes unveiling the forces behind innovation in order to understand current directions and foster a more informed discourse.

According to the AI4SG-VSD methods, the following section presents the identification and conceptual investigation of the values relevant to this research.

5.2 Identification and Conceptual Investigation of Values

The values identified as relevant for the design and application of SARs for young adults' mental health support comprise those defined in the methodology sections and those identified through the context analysis: the promotion of well-being will be discussed alongside collaboration; following, fairness and legitimacy, presented together as closely

related, prevention of harm, autonomy, explainability and care values, namely competence, attentiveness, responsibility and reciprocity.

5.2.1 Promoting Well-being

The potential introduction of SARs in support of young adults' mental health supports #SDG 3, 'Ensuring healthy lives and promoting well-being at all ages' [132]. The integration of social robots holds the promise of enhancing accessibility and broadening the scope and type of mental healthcare interventions. SARs can potentially deliver personalised and continuous support, providing companionship, monitoring, and therapeutic interventions. Their non-judgmental presence and consistent availability could establish a safe and comfortable environment for young adults contending with mental health challenges, acting as a low-stakes entry-point to mental health services or enhancing existing therapies. Furthermore, social robots have the potential to augment the capabilities of healthcare professionals, affording them the capacity to allocate resources towards more complex cases.

Amongst the large number of paradigms in mental health, there are two main perspectives on well-being that are often discussed in psychological literature: eudaimonic and hedonic well-being. Hedonic well-being is centered on happiness, pleasure attainment, and pain avoidance. This approach to well-being focuses on the balance of positive over negative affect and general life satisfaction. Hedonic well-being is often measured by the extent to which people experience subjective happiness and contentment. [133, 134] Eudaimonic well-being, as reflected in Ryff's Psychological Well-Being Scale, emphasizes self-realization and the degree to which a person is fully functioning. It is based on the concept of living in accordance with one's true self and virtues. Its components are self-acceptance, positive relations with others, autonomy, environmental mastery, personal growth, and purpose in life. [135, 133] To enhance well-being, it's important to evaluate how SARs-supported interventions effectively improve relevant outcomes. It is critical to investigate if these connections promote young adults' well-being and enhance human relationships, or whether they exacerbate social isolation and reliance on technology. Because of the contextual, personalised and multi-faceted nature of well-being, ensuring that SARs truly serve the best interests of young adults calls for a collaborative effort between technologists and mental health professionals and a commitment to continuous evalua-

tion and adaptation in practice, with a strong focus on personalisation, patient's safety and awareness. To ensure a responsible approach to the development and deployment of these technologies, collaboration across multiple disciplines should involve not only technologists and mental health professionals but also ethicists, sociologists, policymakers, and representatives from the communities these technologies are intended to serve. Promoting collaboration in design and research effectively supports the promotion of other values. Given the breadth of challenges and potential consequences on the individual, on the healthcare system, and on society, such an inter-disciplinary approach can help ensure that the drivers behind technological innovation are ethical, culturally sensitive, supportive of the population's needs and of society's values, and so are the innovation themselves.

5.2.2 Fairness and Legitimacy

Floridi's account of Justice in the context of AI [103] determined various understandings of the principle:

- "Using AI to correct past wrongs such as eliminating unfair discrimination" [103];
  SARs could serve as a force for social equality by providing mental health support to those who might otherwise lack access due to geographic isolation, socioeconomic barriers, or stigma. However, concerns arise regarding algorithmic bias, where SARs might inadvertently perpetuate existing prejudices if not carefully designed and monitored. Additionally, the accessibility and affordability of these technologies are crucial: without equitable distribution, SARs might exacerbate existing inequalities rather than ameliorate them.

- "Ensuring that the use of AI creates benefits that are shared (or at least shareable)" [103];

  Promoting distributive justice requires a deeper understanding of current inequalities and careful planning for resource allocation. In the context of SARs for young adults' mental health, it is important to ensure that the potential advantages these technologies offer are not exclusively accessible to a selected few, but are, instead, widely shareable across various demographics. Considerations about affordability, accessibility and design choices such as ease of use and cultural sensitivity are part of the effort.

- "Preventing the creation of new harms, such as the undermining of existing social structures" [103].

  An over-reliance on technology, negative consequences on users' social sphere, privacy concerns and consequences of malfunctionings are amongst the potential harms identified in this research. The design and integration of SARs should not undermine or pose risks to the role of human caregivers or the importance of human connections. Additionally, there is a need to consider the sustainability of SARs, both in terms of their environmental impact and the long-term feasibility of their deployment. Careful consideration must be given to how SARs fit into the broader politics of mental healthcare to ensure they complement rather than compromise existing structures and relationships, safeguarding their integrity.

Influenced by fairness, some participants highlighted the importance of investigating the legitimacy of SARs introduction. This means, on the one hand, ensuring that the tools reflect users' preferences and needs, which calls for user-centered design approaches and personalisation strategies; on the other hand, ensuring that the tool is appropriate for the task: it needs to fulfill its purpose and sustain practice related values [92]. Closely related to legitimacy and fairness is the discussion regarding the drivers behind technological innovation, ranging from economic to political and ideological. Research shows concern regarding the development of SARs being driven by technological solutionism, which entails an over-reliance on technology to solve complex social problems [92, 136]. This is problematic from different points of view: a technological solutionist approach might assume that SARs are inherently neutral and beneficial, without thoroughly considering potential biases, ethical dilemmas, or unintended consequences. It is important to recognize that SARs, like any technology, have limitations and raise challenges that must be carefully evaluated. With regards to social equality and distributive justice, while holding the promise of improving accessibility to care through lowering the barrier to seeking help and reaching under-served areas, their implementation might only benefit segments of the population, due, for example, to the cost of the technology. Furthermore, the design might reflect current inequalities. An example of how this might present itself is algorithmic bias: SARs trained on existing data will not be trained on details of those who currently do not access or have access to support, which might result in the technol-

ogy not being suitable for them. Lastly, the complexity of the challenges identified in this research unveiled a critical range of potential harms arising from the deployment of this technology, which are presented in the next section.

Technological solutionism may lead to an uncritical belief that SARs can fully address the multifaceted nature of mental health issues. This perspective could potentially oversimplify the complexities of mental healthcare, overlooking the importance of human factors, therapeutic relationships, and the broader socio-cultural context in the treatment process. This was criticised by participants disagreeing with the idea of SARs replacing human carers and expressing doubts on the suitability of SARs in a complex system of patients, families, and care institutions, to tackle an issue that was framed by some participants as societal, towards which we should all take responsibility. This could potentially undermine existing social structures and potential new ones.

Moreover, technological solutionism may lead to the prioritisation of commercial interests over the needs of people seeking mental health care, as well as those of mental health practitioners and institutions. Market drivers and the potential involvement of private technology firms in public services were criticised by Sharon et. al [137] in virtue of the blind repurposing of data storage and processing capacities as an entrepreneurial endeavor, without taking into account the implicit values, norms, and skills of existing services and actors.

SARs could positively impact social equality and distributive justice, potentially improving accessibility and ubiquity of services, offering personalised and youth-oriented support. Without a proper understanding of the socio-cultural context and nature of the technology, they could also negatively impact fairness in mental healthcare, through, for example, reproducing current inequalities and providing inadequate support. Determining and understanding factors contributing to fairness and legitimacy of SARs introduction in mental healthcare practices requires the consideration of a wide variety of perspectives, ranging from socio-political to ethical and psychological; user-centered design approaches, interdisciplinary and inter-cultural collaboration are recommended to ensure the suitability of these tools in specific contexts.

5.2.3 Prevention of Harm

The bioethical principle of non-maleficence refers to a duty and responsibility of preventing harm, intended or non-intended [103]. SARs hold the potential to prevent harm by providing support to underserved populations, ensuring users' safety and well-being through activities ranging from companionship, monitoring, and on-going support.
With consideration to the potential deployment of SARs in support of young adults' mental health, many potential sources of harm can be identified. Due to SARs monitoring, collection, and re-elaboration of user data, of concern is the potential infringement of personal privacy, which refers to the right of the individual to access and control how personal data is used. [92] In the context of mental healthcare, the importance of privacy is paramount, exacerbated by the vulnerability of patient information, and it is in current practices protected through confidentiality. A few participants expressed worry with regards to the risk of the technology being hacked, or data being used in non-transparent ways and the fear surrounding the possibility of it. An inappropriate use of data could prove to be discriminatory towards certain groups. With personalisation being an important factor in mental healthcare, it is crucial to ensure fair and effective handling of user information and preferences. It is therefore important for designers to focus on transparency, practice informed consent, and adopt secure privacy mechanisms and practices.
Deception and malfunctionings are two further potential sources of harm: whether SARs are deceptive or non-deceptive is a highly debated topic in HRI [92]. Deception can be considered a risk of harm in principle, or because of the negative consequences it can have on users, for example by encouraging them to build an emotional rapport with a machine, incapable of reciprocity [92]. This can create dependency and harm users' autonomy, emotion regulation and social skills. SARs could further isolation in certain circumstances: if a vulnerable user has access to a non-judgemental artificial entity to engage with, they could choose to retract from their social life, or not engage in human-delivered care. Robots' malfunctionings could cause distress or reinforce negative beliefs in vulnerable users. Furthermore, negative experiences are not only harmful in the context in which they happen and their immediate consequences but can affect future help-seeking behaviour and negatively impact trust toward institutions. To prevent the risks arising from deception and malfunctionings, it is important to first ensure the safety and legitimacy of these systems and their design choices. It is then also crucial to ensure accountability, provide compre-

hensive training for all stakeholders, and implement continual evaluation and feedback mechanisms, effectively maintaining human oversight.

Research in the field of autonomous systems is often characterised by a deep divide between technical research and human-centered, ethically involved ones [117], as was indicated by one participant. This divide can contribute to techno-solutionist tendencies, elaborated in the previous section. Intensive and continuous collaborations between experts and researchers of different fields, end users, care providers, and policymakers are key to ensuring an informed evaluation of the legitimacy of the technology, evaluating the potential benefits for users and care providers against potential sources of harm. Maintaining human oversight throughout the testing and deployment of SARs and facilitating feedback mechanisms from users contribute to harm minimization and foster responsibility.

5.2.4 Autonomy

In the context of healthcare, autonomy refers to a patient's right to make decisions about their own medical treatment, including the right to refuse treatment, based on informed consent [138].

In psychology, autonomy is frequently described as a dimension of psychological well-being, characterized by the ability to self-regulate and make decisions independently of external influences [139]. It is also seen as a crucial aspect of healthy psychological development, allowing individuals to form a sense of identity, transitioning from dependence on caregivers to the ability to make independent decisions and take responsibility for one's own actions[140].

In philosophy, there are different theories of personal autonomy [141, 57], and further theories aim at reconceptualising autonomy, for example, from a relational perspective [142, 143]. These various conceptualisations of autonomy reveal a difficulty in providing an operational definition of autonomy [144, 145].

According to Floridi's account of the AI4SG, encouraging the concept of autonomy within the realm of AI involves finding a middle ground between the authority humans hold in making decisions for themselves and what they entrust to artificial agents [103]. Yet, the distribution of autonomy between technology and humans is not a zero-sum game[141]. Following Formosa's account of the impact of SARs on human autonomy, there are 3

ways robots can enhance human autonomy: by achieving more valuable ends, improved autonomy competencies, and encouraging more authentic choices[141].

In the context of mental health, more valuable ends can be achieved by expanding the range of support available for young adults, reducing barriers and improving accessibility of resources. For instance, they could help young adults reach the right type of care promptly; they could free professionals from certain tasks, allowing them to provide better care; Additionally, among young adults, a perceived barrier to seeking professional help is the desire for autonomy and control [146]. Robots, being non-human, could evoke a lessened threat to perceived autonomy, reducing stigma and making reaching out for help more accessible. In some circumstances, robots could provide the support needed to achieve certain means regarded as valuable by the user, for example, by supporting them in achieving certain goals. The relationship between dependence and autonomy is one that needs to be determined in context: becoming dependent to increase one's autonomy can be beneficial in some cases, for instance when the user is able to fullfill certain tasks they would not otherwise do, while detrimental in others. Creating dependence to increase autonomy makes autonomy vulnerable: a few participants raised the question of what would happen when the technology is taken away, or malfunctioning, once the user has grown accustomed to it. Becoming over-reliant on the technology can also negatively impact autonomy overall, reducing in fact valuable means.

Improved autonomy competencies can be achieved through skill learning and positive social interactions. Humans have the capacity to nurture and enhance their autonomy skills through positive social interactions that reinforce qualities like self-respect, self-love, and self-trust [141, 147, 148]. There is limited evidence that this could translate to the interaction with social robots [15]. Learning and practicing skills to support one's mental health can lead to increased confidence and self-efficacy. Some participants suggested that SARs could be programmed to push for autonomy and social interaction. Deploying SARs in mental health support for young adults could potentially result in negatively impacting autonomy competencies: accessible, non-judgemental, personalised support could translate to an over-reliance on interactions with the technology and further social isolation, hindering one's social, decision making and emotion regulation skills.

People can also be aided in making more authentic choices by being guided to reflect on one's values, available information, choice options and emotional state[141], in a similar

way shared decision making processes adopted by health providers aid patients in making choices about their treatment: through understanding of risks, benefits and their alignment to personal values, people are empowered to take decisions that promote their goals, respecting their autonomy. [149]

In the context of young adults' mental health support, the ability to aid one's choices can be considered very valuable: mental complaints can cause problems in one's decision making process. Supporting decision making can therefore be a crucial aspect of support: by practicing the principle of informed consent, SARs can help users navigate choice options. On the other hand, privacy concerns and surveillance threats can diminish people's ability to make authentic choices. SARs can also perpetuate oppressive norms through the bias showcased by language models. Addiction and over-reliance also would negatively impact users ability to make authentic choices.

Whether SARs would positively or negatively impact autonomy depends on the conceptualisation of autonomy, on the specific context, and on the types of features and mechanisms endowed in the technology. Promoting autonomy in the context of SARs for mental health support requires a contextual understanding of autonomy and continuous evaluation of users' experience, attainable through human oversight and feedback mechanisms.

5.2.5 Explainability

Explainability is described by Floridi et al. as comprising of intelligibility, understanding how a system works, and accountability, who should be responsible for it. It complements the other principles: through understanding how the technology works, it is possible to determine where and how the technology creates good or causes harm, evaluate the impact on users' autonomy and care practices, and foster trust in clinicians and patients alike. By holding individuals or organisations accountable, it promotes fairness. [103]

Promoting explainability can mitigate the shortcomings of the technology by aiding a clear definition of its scope, use, and limitations. Even though, as some participants explained, some users might not be interested in the inner workings of technology, having easy access to such information has a positive impact on trust and accountability.

5.2.6 Care values

The care values of responsibility, attentiveness, competence, and reciprocity are values that arise within the relationship between the care-receiver and the care-giver. The evaluation of how these values might change due to the introduction of certain technology aims at ensuring that these practices are enhanced and not hindered. [25] The comparison of care practices with and without the technology is not straightforward: systems with a certain degree of autonomy and interactive capabilities open the avenue to new care practices, making the comparison unattainable [27, 150]. Furthermore, SARs might be considered reciprocal partners, as engaging in degrees of attentiveness and competence[27, 150]. Whether artificial systems should be considered or built as reciprocal partners or not, the nature of care practices themselves undergoes a profound transformation: from care practices consisting of a response of the caregiver to the needs of the care receiver, an inter-subjective relationship between two humans, the introduction of a new entity with certain degrees of autonomy presupposes a redistribution of roles and responsibilities, and the emergence of new factors to evaluate [27, 93].

Ensuring the systems' competence, intended as the ability to execute a task, is paramount to prevent harm and foster trust and acceptability. Competence is supported by transparency, as it is necessary to understand how the system operates to determine its suitability in carrying out certain care practices: this includes evaluation of efficacy and the respect of practice-related values, such as taking into account users' preferences. SARs could enhance care practices' competence by offering support to therapists, enhancing the scope of their interventions. Competence could be compromised when the technology is not fit for the task.

Attentiveness is also crucial to harm prevention: the ability of the system to respond to a patient's need promote their well-being, the inability to do so is potentially harmful. SARs could increase attentiveness in care practices due to their monitoring capabilities and availability. Attentiveness could be negatively impacted when the system fails to recognise events or intentions or respond accordingly.

Responsibility would be greatly impacted by the introduction of SARs: from care practices in which responsibility is in the hands of the caregiver, the introduction of technology shifts it in ways that are not yet defined: issues of accountability are still open to debate. Responsibility is closely aligned with the value of trust [27**?**], which indicates that it

would be positively impacted by explainability [27].

Reciprocity is of great importance in mental healthcare, given that therapeutic relationships are the most influential factor in the success of interventions [44]. While the capacity to build rapport is one of the benefits of SARs, potentially allowing them to aid in assisted care, there is widespread doubt as to whether this rapport could contribute to promoting well-being and recovery due to SARs limited human connection capabilities.

The potential impact of SARs on care values is to be understood in the specific context of application and the uncertainty as to whether SARs could be considered reciprocal partners calls for further research on the topic.

5.3 Potential Applications

Therapy support is a promising potential area to exploit SARs availability: robots could be an optional, personalised tool to support in-between session engagement of young adults, functioning, for example, as a medium to administer therapy-related homework, mindfulness exercises, or exercising skills. This application has the potential of promoting well-being through an expansion of the scope of therapeutic efforts, potentially impacting the attentiveness of the practice in a positive way and enhancing patients' autonomy through improved competencies.

Another area of potential application is prevention. Prevention was indicated by participants as a promising area to invest in: through intervening with, for example, psychoeducation interventions, SARs could contribute to preventative efforts, complementing human efforts through their extended availability and low-stakes accessibility. Prevention-oriented robots could aid young adults with emotional complaints by lowering the stigma around mental health, encouraging help-seeking attitudes and awareness through, for example, practice talking about one's emotions. Preventative efforts contribute to promoting well-being and enhance autonomy competencies.

In this context, as an addition to prevention, SARs could also act as a low-stakes entry point to mental health services, with the goal of aiding young adults in navigating the mental healthcare landscape, understanding their options and needs. This could contribute to users' autonomy by aiding them in making authentic choices. The potential of this application is justified by the difficulty in accessing the right kind of care expressed by participants and the low rates of young adults with emotional complaints seeking pro-

fessional help.

5.4 Design Recommendations

The findings of this study suggest special attention to be paid to safeguarding users, respecting their preferences, promoting fairness, awareness, and autonomy, mitigating harm, and contributing to better care. Following is a discussion of norms and related design recommendations related to the values identified in this research.

- Collaboration between technologists and mental health professionals is paramount in the design and development of robots and interventions. Collaboration was indicated by an HRI researcher as one of the main difficulties in their job. Efforts in understanding barriers to collaboration, ranging from systemic to cultural, should be understood and addressed in order to facilitate research and design of SARs to promote well-being. Collaboration with experts from different fields and cultures and user-centered design practices should be encouraged, to foster a deeper understanding of ethical issues and the socio-cultural context.

- In order to promote the legitimacy of these tools it is recommended to ensure that the innovation respects users' needs and preferences and is appropriate to be integrated in the context, meaning respect and promotion of context-related values and fulfillment of the task.

  - In order to understand and respect stakeholders' needs and preferences, it is recommended to adopt user-centered design approaches that actively involve mental health professionals alongside representatives of young adults experiencing emotional complaints. This approach is especially important in the context of this research, as personalisation has a meaningful role in mental health. Given the multi-faceted and subjective nature of mental health, it is important to further investigate personalisation strategies for SARs. For instance, these might involve empowering users and therapists to select appropriate features, appearance, behaviours, and outcomes. Personalisation could also be pursued through SARs monitoring capabilities, which have the potential of providing insights toward a more personalised care delivery. With regard to mechanisms

involving data collection, special attention towards data privacy and fair use is warranted: algorithmic bias may exacerbate current social inequalities and discrimination. Personalisation can be supported through features such as user recognition and continuous learning. Personalisation contributes to respecting and promoting users' autonomy, as per AI4SG#3, receiver contextualised intervention.

  - Determining the tool's appropriateness and feasibility of integration involves more than ensuring the tool's efficacy: when exploring potential applications in therapy support it is crucial to design for supporting context-specific values, such as the therapeutic relationship, paying special attention to factors that could contribute to the strengthening or deteriorating the rapport. Other values to take into consideration are confidentiality, which invokes privacy protection, as per AI4SG#5, accessibility, and affordability, to promote equity. Furthermore, the systems in which these technologies are introduced must be ready to accommodate them. This means taking into considerations the shifting of roles and responsibilities. Design decisions should be aligned with the goal of integrating new technologies in ways that enhance existing services, and reinforce rather than replace human connections and therapeutic relationships. The introduction of SARs and other technological interventions should be presented as an option, should not come at the expense of human-centered services, and should be sensitive to the broader socio-political context. Understanding mental health as a societal issue means that solutions, perspectives, and responsibilities should be spread out and collaborative, avoiding an over-reliance on technology.

- Fairness can be respected and promoted through practices and designs aimed at evaluating and ensuring equitable treatment, as dictated by AI4SG#6 (situational fairness). In the context of SARs for mental health support, accessibility features such as appropriate communication modalities or interfaces, inter-cultural collaboration in design, and attention to safeguard against potential bias in algorithms are recommended. With regards to AI4SG#2, safeguards against the manipulation of predictors, the findings of this research did not yield any particular insight.

- Maintaining human oversight throughout the testing and deployment of SARs and facilitating feedback mechanisms from users contribute to harm minimization and foster responsibility. These recommended design and deployment practices support the goals of AI4SG#1, falsifiability and incremental deployment, namely ensuring effectiveness, through empirically validating the technology's functioning, and safety. Given the role of confidentiality in mental healthcare, privacy mechanisms should be implemented, in line with AI4SG#5 (privacy protection and data subject consent). Users should, for example, be given the ability to delete personal data in an easy manner, and robots should be designed with robust security measures. These measures should prevent unauthorized access and ensure that the conversations are not recorded or exposed to third parties without explicit consent.

- In the context of young adults' mental health, SARs potential impact on human autonomy is highly contextual, which underscores the importance of personalisation strategies, as dictated by AI4SG#3 (receiver-contextualised intervention). By supporting skill development and psycho-education, through adaptive, personalised communication features, SARs can contribute to users' competencies, decision-making, and understandings, effectively operationalising AI4SG#7 (human-friendly semanticisation). SARs non-judgemental presence could also negatively impact users' autonomy, creating over-reliance. This points to the importance of human oversight and continuous evaluation. Lastly, designing for privacy, in accordance to AI4SG#5 (privacy protection and data subject consent), is essential to promote human autonomy.

- Explainability can contribute to empowering users through enabling transparent and effective user education, in line with AI4SG#4, receiver-contextualised explanation and transparent purposes. The principle emphasizes the need for technology to be explained in a manner tailored to the user's context, enhancing its relevance and usability. This approach ensures that explanations are meaningful and accessible to different groups of users, considering, for example, their backgrounds and education levels. Similarly, AI4SG#7 (human-friendly semanticisation) underscores the importance of fostering users' 'semantic capital', supporting their capacity to make sense of things. This principle advocates for SARs communication features that enhance users' power to give meaning and understand. Unraveling strategies

to operationalize this principle should be supported by collaboration with experts and stakeholders.

- Attentiveness can be supported by endowing SARs with situational awareness, enabling them to respond to users' needs through detection and communication features. Collaboration during the design process is crucial to support competence, ensuring that SARs interventions align with the field's best practices. Responsibility can be supported through addressing accountability issues. Communication features aimed at building rapport could support reciprocity.

- Promoting well-being in the design and application of SARs requires an operationalisation and evaluation of all values discussed in this paper. The reccomendations are not exhaustive, as is the list of values. Promoting well-being requires a continuous effort to explore and evaluate potential avenues for research and design.

6 Limitations and future research

This research initiated a first iteration of the AI4SG-VSD methodology applied to SARs for supporting the well-being of young adults with emotional complaints, yielding a context analysis, conceptual investigation of values, and proposed design requirements. To the author's best knowledge, it is the first research applying Value-Sensitive design to SARs for mental healthcare. This methodology is an iterative process, to be continued throughout the design phase: by providing a first account of context, values, and design requirements, the study at hand aims at laying a foundation of the concepts and understandings relevant to this design process. First iteration in this case means starting from abstract notions: instead of commencing the design process for a particular application, identifying potential applications was one of the study's goals. This translates into yielding broader directions and recommendations, instead of specific steps toward robotic design and prototyping. Future research should re-iterate the research methods within a specific application amongst the ones proposed. Narrowing the scope allows for a deeper and contextualised exploration of context, values, and design requirements. Additionally, as previously mentioned, efforts towards facilitating inter-disciplinary collaborations should be brought forward, as should efforts towards understanding and addressing the systemic barriers posed by Academia, such as funding strategies and publication incentives. Future research should focus on the integration of technology into existing practices. Lastly, methods for tailoring technological interventions to individual needs and preferences should be researched.

The study's limitations include the wide range of concepts encountered, highlighting the need for expert collaboration across various fields for effective exploration. Originally, the study aimed to contrast the views of the two different stakeholder groups. However, the similarity in backgrounds between the groups and recruitment challenges led to the decision of analysing the data as a whole.

Additionally, the study excluded young adults with emotional complaints from the empirical investigation. While this was determined by the ethical feasibility of a Master's thesis, it also means the research may lack some viewpoints that could be critical for a more complete understanding.

7 Conclusion

This thesis explored the potential of SARs as a tool for supporting the mental health of young adults with emotional complaints. The mental health state of young adults is critical, with a high incidence of disorders and low access to services, which justifies the investigation of potential new ways of delivering care. Through a value-oriented design process, following the AI4SG-VSD framework, the study provided an analysis of the context, including young adults' mental health state, tools and practices adopted to support it, and the role of SARs in the field. The context analysis was expanded through an analysis of qualitative interviews with psychologists and HRI researchers, selected as stakeholders, revealing their needs, perspectives, and values. Participants identified professional challenges of a systemic nature, indicating difficulties and limitations posed by Academia and mental health institutionalisation. They also highlighted the importance of integrating the technology in current practices, of the therapeutic relationship and personalisation in therapy, and issues of inclusivity and accessibility of mental health services. The mental health crisis was described as a societal issue, highlighting the need for shared responsibility towards the issue. The identified challenges were in line with previous research, with the risk of dependence and of negative effects on users' social spheres being a few of them; so were the benefits, which included accessibility and availability. In accordance with the AI4SG-VSD framework, values relevant to the design of SARs in the field were identified and conceptualised. These included values from the EU HLEG on AI: autonomy, fairness, prevention of harm and explicability; UN SDG#3, 'Ensuring healthy lives and promoting well-being at all ages'; care values: attentiveness, responsibility, competence, and reciprocity; values identified through the context analysis:collaboration and legitimacy. The context analysis and values conceptualisation wereused to identify potential applications and design recommendations. Therapy support, prevention, and positioning SARs as an entry point to mental health services were the potential applications presented in this research. Design recommendations were provided for each identified value. Design recommendations included features contributing to equity, such as affordability, accessibility and ease of use such as personalised, adaptive communication, transparent user education strategies, situational awareness, maintaining human oversight and facilitating feedback mechanisms. This research carried out a first iteration of the AI4SG-VSD framework, effectively gathering insights, conceptual understanding and recommendations for future

research. These serve as a stepping stone for future investigations on the application of SARs for young adults' mental health support.

References

[1] Arielle Aj Scoglio, Erin D Reilly, Jay A Gorman, and Charles E Drebing. Use of social robots in mental health and well-being research: Systematic review.

[2] Imane Guemghar, Paula Pires de Oliveira Padilha, Amal Abdel-Baki, Didier Jutras-Aswad, Jesseca Paquette, and Marie-Pascale Pomey. Social robot interventions in mental health care and their outcomes, barriers, and facilitators: Scoping review.

[3] Samira Rasouli, Garima Gupta, Elizabeth Nilsen, and Kerstin Dautenhahn. Potential applications of social robots in robot-assisted interventions for social anxiety. *International Journal of Social Robotics 2022 14:5*, 14:1–32, 1 2022.

[4] Elias Aboujaoude, Lina Gega, Michelle B. Parish, and Donald M. Hilty. Editorial: Digital interventions in mental health: Current status and future directions. *Frontiers in Psychiatry*, 11:111, 2 2020.

[5] Chin Kuo Chang, Richard D. Hayes, Matthew Broadbent, Andrea C. Fernandes, William Lee, Matthew Hotopf, and Robert Stewart. All-cause mortality among people with serious mental illness (smi), substance use disorders, and depressive disorders in southeast london: A cohort study. *BMC Psychiatry*, 10:1–7, 9 2010.

[6] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. *JMIR Ment Health 2017;4(2):e19 https://mental.jmir.org/2017/2/e19*, 4:e7785, 6 2017.

[7] Amelia Fiske, Peter Henningsen, and Alena Buyx. Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Internet Res 2019;21(5):e13216 https://www.jmir.org/2019/5/e13216*, 21:e13216, 5 2019.

[8] Emily G. Lattie, Rachel Kornfield, Kathryn E. Ringland, Renwen Zhang, Nathan Winquist, and Madhu Reddy. Designing mental health technologies that support the social ecosystem of college students. *Conference on Human Factors in Computing Systems - Proceedings*, 4 2020.

[9] Sandra Garrido, Chris Millington, Daniel Cheers, Katherine Boydell, Emery Schubert, Tanya Meade, and Quang Vinh Nguyen. What works and what doesnât work? a systematic review of digital mental health interventions for depression and anxiety in young people. *Frontiers in Psychiatry*, 10:759, 11 2019.

[10] Rachel Kenny, Barbara Dooley, and Amanda Fitzgerald. Ecological momentary assessment of adolescent problems, coping efficacy, and mood states using a mobile phone app: An exploratory study. *JMIR Ment Health 2016;3(4):e51 https://mental.jmir.org/2016/4/e51*, 3:e6361, 11 2016.

[11] Sarah M. Rabbitt, Alan E. Kazdin, and Brian Scassellati. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology Review*, 35:35–46, 2 2015.

[12] Katarzyna Kabacińska, Tony J Prescott, and Julie M Robillard. Socially assistive robots as mental health interventions for children: A scoping review. *International Journal of Social Robotics*, 13:919–935, 2021.

[13] Nicole Lee Robinson, Timothy Vaughan Cottier, and David John Kavanagh. Psychosocial health interventions by social robots: Systematic review of randomized controlled trials. *J Med Internet Res 2019;21(5):e13203 https://www.jmir.org/2019/5/e13203*, 21:e13203, 5 2019.

[14] Cristina A. Costescu, Bram Vanderborght, and Daniel O. David. The effects of robot-enhanced psychotherapy: A meta-analysis. *Review of General Psychology*, 18:127–136, 6 2014.

[15] Sooyeon Jeong, Laura Aymerich-Franch, Kika Arias, Sharifa Alghowinem, Agata Lapedriza, Rosalind Picard, Hae Won Park, and Cynthia Breazeal. Deploying a robotic positive psychology coach to improve college studentsâ psychological well-being. *User Modeling and User-Adapted Interaction*, pages 1–45, 7 2022.

[16] Ashwin Sadananda Bhat, Christiaan Boersma, Max Jan Meijer, Maaike Dokter, Ernst Bohlmeijer, and Jamy Li. Plant robot for at-home behavioral activation therapy reminders to young adults with depression. *ACM Transactions on Human-Robot Interaction (THRI)*, 10, 7 2021.

[17] Marco Solmi, Joaquim Radua, Miriam Olivola, Enrico Croce, Livia Soardo, Gonzalo Salazar de Pablo, Jae Il Shin, James B. Kirkbride, Peter Jones, Jae Han Kim, Jong Yeob Kim, AndrÃ¨ F. Carvalho, Mary V. Seeman, Christoph U. Correll, and Paolo Fusar-Poli. Age at onset of mental disorders worldwide: large-scale meta-analysis ofÂ 192 epidemiological studies. *Molecular Psychiatry 2021 27:1*, 27:281–295, 6 2021.

[18] Gavin Andrews, Kristy Sanderson, Justine Corry, and Helen M. Lapsley. Using epidemiological data to model efficiency in reducing the burden of depression. *The Journal of Mental Health Policy and Economics*, 3:175–186, 12 2000.

[19] Kathleen Vanheusden, Cornelis L. Mulder, Jan van der Ende, Frank J. van Lenthe, Johan P. Mackenbach, and Frank C. Verhulst. Young adults face major barriers to seeking help from mental health services. *Patient Education and Counseling*, 73:97–104, 10 2008.

[20] Vikram Patel, Alan J. Flisher, Sarah Hetrick, and Patrick McGorry. Mental health of young people: a global public-health challenge. *The Lancet*, 369:1302–1313, 4 2007.

[21] Terhi Aalto-Setälä, Mauri Marttunen, Annamari Tuulio-Henriksson, Kari Poikolainen, and Jouko Lönnqvist. Psychiatric treatment seeking and psychosocial impairment among young adults with depression. *Journal of Affective Disorders*, 70:35–47, 6 2002.

[22] Nisha Mehta, Tim Croudace, and Sally C. Davies. Public mental health: evidenced-based priorities. *The Lancet*, 385:1472–1475, 4 2015.

[23] Derek Richards. Prevalence and clinical course of depression: A review. *Clinical Psychology Review*, 31:1117–1125, 11 2011.

[24] Sidney Zisook, Ira Lesser, Jonathan W. Stewart, Stephen R. Wisniewski, G. K. Balasubramani, Maurizio Fava, William S. Gilmer, Timothy R. Dresselhaus, Michael E. Thase, Andrew A. Nierenberg, Madhukar H. Trivedi, and A. John Rush. Effect of age at onset on the course of major depressive disorder. *American Journal of Psychiatry*, 164:1539–1546, 10 2007.

[25] Aimee van Wynsberghe. Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics*, 19:407–433, 6 2013.

[26] Bernd Carsten Stahl and Mark Coeckelbergh. Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems*, 86:152–161, 2016.

[27] Steven Umbrello, Marianna Capasso, Maurizio Balistreri, Alberto Pirni, and Federica Merenda. Value sensitive design to achieve the un sdgs with ai: A case of elderly care robots. *Minds and Machines*, 31:395–419, 9 2021.

[28] Steven Umbrello and Ibo van de Poel. Mapping value sensitive design onto ai for social good principles. *AI and Ethics 2021 1:3*, 1:283–296, 2 2021.

[29] Laurie A. Manwell, Skye P. Barbic, Karen Roberts, Zachary Durisko, Cheolsoon Lee, Emma Ware, and Kwame McKenzie. What is mental health? evidence towards a new definition from a mixed methods multidisciplinary international survey. *BMJ Open*, 5:e007079, 6 2015.

[30] World Health Organization et al. *Promoting mental health: Concepts, emerging evidence, practice: Summary report*. World Health Organization, 2004.

[31] George L. Engel. The need for a new medical model: A challenge for biomedicine. *Science*, 196:129–136, 1977.

[32] Silvana Galderisi, Andreas Heinz, Marianne Kastrup, Julian Beezhold, and Norman Sartorius. Toward a new definition of mental health. *World Psychiatry*, 14:231, 6 2015.

[33] Markus Jokela. Why is cognitive ability associated with psychological distress and wellbeing? exploring psychological, biological, and social mechanisms. *Personality and Individual Differences*, 192:111592, 7 2022.

[34] Sylvaine Artero, Jacques Touchon, and Karen Ritchie. Disability and mild cognitive impairment: a longitudinal population-based study. *International Journal of Geriatric Psychiatry*, 16:1092–1097, 11 2001.

[35] Karny Gigi, Nomi Werbeloff, Shira Goldberg, Shirly Portuguese, Abraham Reichenberg, Eyal Fruchter, and Mark Weiser. Borderline intellectual functioning is associated with poor social functioning, increased rates of psychiatric diagnosis and drug use â a cross sectional population based study. *European Neuropsychopharmacology*, 24:1793–1797, 11 2014.

[36] James J. Gross and Ricardo F. Mu noz. Emotion regulation and mental health. *Clinical Psychology: Science and Practice*, 2:151–164, 1995.

[37] Sam Wren-Lewis and Anna Alexandrova. Mental health without well-being. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 46:684–703, 12 2021.

[38] Rebecca B. Price and Mary L. Woody. Emotional disorders in development. *Encyclopedia of Behavioral Neuroscience: Second Edition*, 3-3:364–368, 1 2022.

[39] World Health Organization. Regional Office for Europe. *The European Mental Health Action Plan 2013â2020*. World Health Organization. Regional Office for Europe, 2015.

[40] Tom Bschor and Mazda Adli. Treatment of depressive disorders. *Deutsches Ãrzteblatt International*, 105:782, 11 2008.

[41] Sarah C. Cook, Ann C. Schwartz, and Nadine J. Kaslow. Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics*, 14:537, 7 2017.

[42] K. Hoagwood, B. J. Burns, L. Kiser, H. Ringeisen, and S. K. Schoenwald. Evidence-based practice in child and adolescent mental health services. *https://doi.org/10.1176/appi.ps.52.9.1179*, 52:1179–1189, 9 2001.

[43] Sidney J. Blatt, David C. Zuroff, Lance L. Hawley, and John S. Auerbach. Predictors of sustained therapeutic change. *Psychotherapy Research*, 20:37–54, 1 2010.

[44] Michael J. Lambert and Dean E. Barley. Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy*, 38:357–361, 2001.

[45] Ronald C. Kessler, G. Paul Amminger, Sergio Aguilar-Gaxiola, Jordi Alonso, Sing Lee, and T. Bedirhan Üstün. Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, 20:359–364, 7 2007.

[46] Anu E. Castaneda, Annamari Tuulio-Henriksson, Mauri Marttunen, Jaana Suvisaari, and Jouko Lönnqvist. A review on cognitive impairments in depressive and anxiety disorders with a focus on young adults. *Journal of Affective Disorders*, 106:1–27, 2 2008.

[47] Paul Rohde, Peter M. Lewinsohn, and John R. Seeley. Are adolescents changed by an episode of major depression? *Journal of the American Academy of Child Adolescent Psychiatry*, 33(9):1289–1298, 1994.

[48] Ronald C. Kessler, Patricia Berglund, Olga Demler, Robert Jin, Kathleen R. Merikangas, and Ellen E. Walters. Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. *Archives of general psychiatry*, 62:593–602, 6 2005.

[49] D L Newman, T E Moffitt, A Caspi, L Magdol, P A Silva, and W R Stanton. Psychiatric disorder in a birth cohort of young adults: prevalence, comorbidity, clinical significance, and new case incidence from ages 11 to 21. *J. Consult. Clin. Psychol.*, 64(3):552–562, June 1996.

[50] Marco Colizzi, Antonio Lasalvia, and Mirella Ruggeri. Prevention and early intervention in youth mental health: is it time for a multidisciplinary and transdiagnostic model for care? *International Journal of Mental Health Systems 2020 14:1*, 14:1–14, 3 2020.

[51] Fiona M. Gore, Paul J.N. Bloem, George C. Patton, Jane Ferguson, Véronique Joseph, Carolyn Coffey, Susan M. Sawyer, and Colin D. Mathers. Global burden of disease in young people aged 10-24 years: A systematic analysis. *The Lancet*, 377:2093–2102, 2011.

[52] World Health Organization. *Suicide worldwide in 2019: global health estimates*. World Health Organization, 2021.

[53] Linda M. Richter. Studying adolescence. *Science*, 312:1902–1905, 6 2006.

[54] Michael Rutter and David John Smith. Psychosocial disorders in young people : time trends and their causes. page 843, 1995.

[55] David Wood, Tara Crapnell, Lynette Lau, Ashley Bennett, Debra Lotstein, Maria Ferris, and Alice Kuo. Emerging adulthood as a critical stage in the life course. *Handbook of Life Course Health Development*, pages 123–143, 1 2017.

[56] Katerina Koutra, Varvara Pantelaiou, and Georgios Mavroeides. Why donrsquo;t young people seek help for mental illness? a cross-sectional study in greece. *Youth*, 3(1):157–169, 2023.

[57] Joanna K. Anderson, Emma Howarth, Maris Vainre, Peter B. Jones, and Ayla Humphrey. A scoping literature review of service-level barriers for access and engagement with mental health services for children and young people. *Children and Youth Services Review*, 77:164–176, 6 2017.

[58] Jonathan G. Perle, Leah C. Langsam, Allison Randel, Shane Lutchman, Alison B. Levine, Anthony P. Odland, Barry Nierenberg, and Craig D. Marker. Attitudes toward psychological telehealth: Current and future clinical psychologistsâ opinions of internet-based interventions. *Journal of Clinical Psychology*, 69:100–113, 1 2013.

[59] G. Andrews, A. Basu, P. Cuijpers, M. G. Craske, P. McEvoy, C. L. English, and J. M. Newby. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders*, 55:70–78, 4 2018.

[60] Mark A. Reger and Gregory A. Gahm. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. *Journal of Clinical Psychology*, 65:53–75, 1 2009.

[61] Jesse H. Wright and Matthew Mishkind. Computer-assisted cbt and mobile apps for depression: Assessment and integration into clinical care. *https://doi.org/10.1176/appi.focus.20190044*, 18:162–168, 4 2020.

[62] Arfan Ahmed, Nashva Ali, Sarah Aziz, Alaa A Abd-alrazaq, Asmaa Hassan, Mohamed Khalifa, Bushra Elhusein, Maram Ahmed, Mohamed Ali Siddig Ahmed, and Mowafa Househ. A review of mobile chatbot apps for anxiety and depression and their self-care features. *Computer Methods and Programs in Biomedicine Update*, 1:100012, 1 2021.

[63] Andrea B. Temkin, Jennifer Schild, Avital Falk, and Shannon M. Bennett. Mobile apps for youth anxiety disorders: A review of the evidence and forecast of future innovations. *Professional Psychology: Research and Practice*, 51:400–413, 8 2020.

[64] Pooja Chandrashekar. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *mHealth*, 4:6–6, 3 2018.

[65] Nannan Long, Yongxiang Lei, Lianhua Peng, Ping Xu, Ping Mao, Nannan Long, Yongxiang Lei, Lianhua Peng, Ping Xu, and Ping Mao. A scoping review on monitoring mental health using smart wearable devices. *Mathematical Biosciences and Engineering 2022 8:7899*, 19:7899–7919, 2022.

[66] D. Freeman, S. Reeve, A. Robinson, A. Ehlers, D. Clark, B. Spanlang, and M. Slater. Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 47:2393–2400, 10 2017.

[67] Caroline J. Falconer, Aitor Rovira, John A. King, Paul Gilbert, Angus Antley, Pasco Fearon, Neil Ralph, Mel Slater, and Chris R. Brewin. Embodying self-compassion within virtual reality and its effects on patients with depression. *British Journal of Psychiatry Open*, 2:74–80, 1 2016.

[68] Giuseppe Riva. Virtual environment for body image modification: virtual reality system for the treatment of body image disturbances. *Computers in Human Behavior*, 14:477–490, 9 1998.

[69] Katharina Meyerbröker and Paul M.G. Emmelkamp. Virtual reality exposure therapy in anxiety disorders: a systematic review of process-and-outcome studies. *Depression and Anxiety*, 27:933–944, 10 2010.

[70] Cynthia Breazeal. Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3527–3538, 12 2009.

[71] Kim Baraka, Patrícia Alves-Oliveira, and Tiago Ribeiro. An extended framework for characterizing social robots. *ArXiv*, pages 21–64, 2019.

[72] Eric Deng, Bilge Mutlu, and Maja J. Matarić. Embodiment in socially interactive robots. *Foundations and Trends in Robotics*, 7(4):251–356, 2019.

[73] Maja J. Matarić and Brian Scassellati. Socially assistive robotics. *Springer Handbook of Robotics*, pages 1973–1993, 1 2016.

[74] Roger Bemelmans, Gert Jan Gelderblom, Pieter Jonker, and Luc de Witte. Effectiveness of robot paro in intramural psychogeriatric care: A multicenter quasi-experimental study. *Journal of the American Medical Directors Association*, 16:946–950, 11 2015.

[75] Geoffrey W. Lane, Delilah Noronha, Kathy Craig, Christina Yee, Alexandra Rivera, Brent Mills, and Eimee Villanueva. Effectiveness of a social robot, âparo,â in a va long-term care setting. *Psychological Services*, 13:292–299, 2016.

[76] Hayley Robinson, Bruce MacDonald, Ngaire Kerse, and Elizabeth Broadbent. The psychosocial effects of a companion robot: A randomized controlled trial. *Journal of the American Medical Directors Association*, 14:661–667, 9 2013.

[77] Ruby Yu, Elsie Hui, Jenny Lee, Dawn Poon, Ashley Ng, Kitty Sit, Kenny Ip, Fannie Yeung, Martin Wong, Takanori Shibata, and Jean Woo. Use of a therapeutic, socially assistive pet robot (paro) in improving mood and stimulating social interaction and communication for people with dementia: Study protocol for a randomized controlled trial. *JMIR Res Protoc 2015;4(2):e45 https://www.researchprotocols.org/2015/2/e45*, 4:e4189, 5 2015.

[78] Kazuyoshi Wada, Takanori Shibata, Tomoko Saito, and Kazuo Tanie. Psychological and social effects in long-term experiment of robot assisted activity to elderly people at a health service facility for the aged. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3:3068–3073, 2004.

[79] Casey C. Bennett, Selma Sabanovic, Jennifer A. Piatt, Shinichi Nagata, Lori Eldridge, and Natasha Randall. A robot a day keeps the blues away. *Proceedings - 2017 IEEE International Conference on Healthcare Informatics, ICHI 2017*, pages 536–540, 9 2017.

[80] Kentarou Kurashige, Eriko Sakurai, Rainer Knauf, Setsuo Tsuruta, Yoshitaka Sakurai, and Ernesto Damiani. Context respectful counseling agent integrated with robot nodding for dialog promotion. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, 2017-January:1540–1545, 11 2017.

[81] Minoo Alemi, Ashkan Ghanbarzadeh, Ali Meghdari, and Leila Jafari Moghadam. Clinical application of a humanoid robot in pediatric cancer interventions. *International Journal of Social Robotics*, 8:743–759, 11 2016.

[82] Molly K. Crossman, Alan E. Kazdin, and Elizabeth R. Kitt. The influence of a socially assistive robot on mood, anxiety, and arousal in children. *Professional Psychology: Research and Practice*, 49:48–56, 2 2018.

[83] Cory Ann Smarr, Akanksha Prakash, Jenay M. Beer, Tracy L. Mitzner, Charles C. Kemp, and Wendy A. Rogers. Older adultsâ preferences for and acceptance of robot assistance for everyday living tasks. *http://dx.doi.org/10.1177/1071181312561009*, pages 153–157, 9 2012.

[84] Juan Fasola and Maja J Matarić. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2, 6 2013.

[85] Joana Galv ao Gomes Da Silva, David J. Kavanagh, Tony Belpaeme, Lloyd Taylor, Konna Beeson, and Jackie Andrade. Experiences of a motivational interview delivered by a robot: Qualitative study. *Journal of medical Internet research*, 20, 5 2018.

[86] Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *https://doi.org/10.1146/annurev-bioeng-071811-150036*, 14:275–294, 7 2012.

[87] Elizabeth S. Kim, Lauren D. Berkovits, Emily P. Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, 43:1038–1049, 5 2013.

[88] Hoang Long Cao, Pablo G. Esteban, Madeleine Bartlett, Paul Baxter, Tony Belpaeme, Erik Billing, Haibin Cai, Mark Coeckelbergh, Cristina Costescu, Daniel

David, Albert De Beir, Daniel Hernandez, James Kennedy, Honghai Liu, Silviu Matu, Alexandre Mazel, Amit Pandey, Kathleen Richardson, Emmanuel Senft, Serge Thill, Greet Van De Perre, Bram Vanderborght, David Vernon, Kutoma Wakanuma, Hui Yu, Xiaolong Zhou, and Tom Ziemke. Robot-enhanced therapy: Development and validation of supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robotics and Automation Magazine*, 26:49–58, 6 2019.

[89] Sofia Pliasa and Nikolaos Fachantidis. Can a robot be an efficient mediator in promoting dyadic activities among children with autism spectrum disorders and children of typical development? *ACM International Conference Proceeding Series*, 9 2019.

[90] Efstathia Karakosta, Kerstin Dautenhahn, Dag Sverre Syrdal, Luke Jai Wood, and Ben Robins. Using the humanoid robot kaspar in a greek school environment to support children with autism spectrum condition. *Paladyn*, 10:298–317, 1 2019.

[91] Ramona E. Simut, Johan Vanderfaeillie, Andreea Peca, Greet Van de Perre, and Bram Vanderborght. Children with autism spectrum disorders make a fruit salad with probo, the social robot: An interaction study. *Journal of Autism and Developmental Disorders*, 46:113–126, 1 2016.

[92] Júlia Pareto Boada, Bego na Román Maestre, and Carme Torras Genís. The ethical issues of social assistive robotics: A critical literature review. *Technology in Society*, 67:101726, 11 2021.

[93] Aimee van Wynsberghe and Shuhong Li. a paradigm shift for robot ethics: from hri to humanndash;robotndash;system interaction (hrsi) . *Medicolegal and Bioethics*, 9:11–21, 9 2019.

[94] Selma Šabanović. Robots in society, society in robots: Mutual shaping of society and technology as a framework for social robot design. *International Journal of Social Robotics*, 2:439–450, 10 2010.

[95] Batya Friedman, Peter Kahn, and Alan Borning. Value sensitive design: Theory and methods. *University of Washington technical report*, 2:12, 2002.

[96] B. Friedman, D. C. Howe, and E. Felten. Informed consent in the mozilla browser: Implementing value-sensitive design. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2002-January:10–19, 2002.

[97] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. Value sensitive design and information systems. *Philosophy of Engineering and Technology*, 16:55–95, 2013.

[98] Alessandra Cenci, Susanne Jakobsen Ilskov, Â· Nicklas, Sindlev Andersen, and Marco Chiarandini. The participatory value-sensitive design (vsd) of a mhealth app targeting citizens with dementia in a danish municipality. *AI and Ethics 2023*, 1:1–27, 4 2023.

[99] Anders Albrechtslund. Ethics and technology design. *Ethics and Information Technology*, 9(1):63–72, December 2006.

[100] Noëmi Manders-Huits. What values in design? the challenge of incorporating moral values into design. *Science and Engineering Ethics*, 17(2):271–287, 2011.

[101] Luciano Floridi, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. How to design ai for social good: Seven essential factors. *Science and Engineering Ethics*, 26:1771–1796, 6 2020.

[102] Basil Varkey. Principles of clinical ethics and their application to practice. *Medical Principles and Practice*, 30:17–28, 2 2021.

[103] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. Ai4people - an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28:689–707, 12 2018.

[104] Rachel Kornfield, Jonah Meyerhoff, Hannah Studd, Ananya Bhattacharjee, Joseph Jay Williams, Madhu Reddy, and David C. Mohr. Meeting users where they are: User-centered design of an automated text messaging tool to support the mental health of young adults. *Conference on Human Factors in Computing Systems - Proceedings*, 4 2022.

[105] Gesine Schwan. Sustainable development goals: A call for global partnership and cooperation. *GAIA - Ecological Perspectives for Science and Society*, 28(2):73–73, January 2019.

[106] Batya Friedman, David G. Hendry, and Alan Borning. A survey of value sensitive design methods. *Foundations and Trends in HumanâComputer Interaction*, 11(2):63–125, 2017.

[107] Lisa P. Nathan, Predrag V. Klasnja, and Batya Friedman. Value scenarios: A technique for envisioning systemic effects of new technologies. *Conference on Human Factors in Computing Systems - Proceedings*, pages 2585–2590, 2007.

[108] Ole R Holsti. *Content analysis for the social sciences and humanities*. Longman Higher Education, Harlow, England, October 1969.

[109] Niek Mouter and Diana Vonk Noordegraaf. Intercoder reliability for qualitative research. *TRAIL Research School*, 12, 2012.

[110] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

[111] Laila Burla, Birte Knierim, Jurgen Barth, Katharina Liewald, Margreet Duetz, and Thomas Abel. From text to codings: intercoder reliability assessment in qualitative content analysis. *Nursing research*, 57(2):113–117, 2008.

[112] Benedikt Leichtmann, Verena Nitsch, and Martina Mara. Crisis ahead? why human-robot interaction user studies may have replicability problems and directions for improvement. *Frontiers in Robotics and AI*, 9, 3 2022.

[113] Frauke Zeller and Lauren Dwyer. Systems of collaboration: challenges and solutions for interdisciplinary research in ai and social robotics. *Discover Artificial Intelligence 2022 2:1*, 2:1–12, 6 2022.

[114] S. Sabanovic, Eric Meisner, Linnda Caporael, Volkan Isler, and J.C. (Jeff) Trinkle. 'outside-in" design for interdisciplinary hri research. pages 41–48, 01 2009.

[115] S. Sabanovic, Marek Michalowski, and Linnda Caporael. Making friends: Building social robots through interdisciplinary collaboration. pages 71–77, 01 2007.

[116] Patrícia Alves-Oliveira, Dennis KÃ¼ster, Arvid Kappas, and Ana Paiva. Psychological science in hri: Striving for a more integrated field of research. 11 2016.

[117] Piers Gooding and Timothy Kariotis. Ethics and law in research on algorithmic and data-driven technology in mental health care: Scoping review. *JMIR Ment Health 2021;8(6):e24668 https://mental.jmir.org/2021/6/e24668*, 8:e24668, 6 2021.

[118] Scott T. Ronis, Amanda K. Slaunwhite, and Kathryn E. Malcom. Comparing strategies for providing child and youth mental health care services in canada, the united states, and the netherlands. *Administration and Policy in Mental Health and Mental Health Services Research*, 44:955–966, 11 2017.

[119] Keith C. Herman, Wendy M. Reinke, Aaron M. Thompson, Kristin M. Hawley, Kelly Wallis, Melissa Stormont, and Clark Peters. A public health approach to reducing the societal prevalence and burden of youth mental health problems: Introduction to the special issue. *School Psychology Review*, 50:8–16, 11 2020.

[120] John R. Weisz, Irwin N. Sandler, Joseph A. Durlak, and Barry S. Anton. Promoting and protecting youth mental health through evidence-based prevention and treatment. *American Psychologist*, 60:628–648, 9 2005.

[121] Leah S. Steele, Carolyn S. Dewa, Elizabeth Lin, and Kenneth L.K. Lee. Education level, income level and mental health services use in canada: Associations and policy implications. *Healthcare Policy*, 3:96, 8 2007.

[122] John C. Norcross. The therapeutic relationship. *The heart and soul of change: Delivering what works in therapy (2nd ed.).*, pages 113–141.

[123] Elsa Marziali and Leslie Alexander. The power of the therapeutic relationship. *American Journal of Orthopsychiatry*, 61:383–391, 1991.

[124] Joshua K. Swift, Rhett H. Mullins, Elizabeth A. Penix, Katharine L. Roth, and Wilson T. Trusty. The importance of listening to patient preferences when making mental health care decisions. *World Psychiatry*, 20:316–317, 10 2021.

[125] Mei Yi Ng and John R. Weisz. Annual research review: Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry*, 57:216–236, 3 2016.

[126] Silvan Hornstein, Kirsten Zantvoort, Ulrike Lueken, Burkhardt Funk, and Kevin Hilbert. Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. *Frontiers in Digital Health*, 5:1170002, 5 2023.

[127] Patrícia Alves-Oliveira, Tanya Budhiraja, Samuel So, Raida Karim, Elin Björling, and Maya Cakmak. Robot-mediated interventions for youth mental health. *Design for Health*, 6:138–162, 5 2022.

[128] Charles F. Reynolds. Optimizing personalized management of depression: the importance of realâworld contexts and the need for a new convergence paradigm in mental health. *World Psychiatry*, 19:266, 10 2020.

[129] Audrey Chapman, Carmel Williams, Julie Hannah, and Dainius Pūras. Reimagining the mental health paradigm for our collective well-being. *Health and Human Rights*, 22:1, 6 2020.

[130] Mike Slade, Michaela Amering, Marianne Farkas, Bridget Hamilton, Mary O'Hagan, Graham Panther, Rachel Perkins, Geoff Shepherd, Samson Tse, and Rob Whitley. Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. *World Psychiatry*, 13(1):12–20, 2014.

[131] Hatice Gunes, Â· Frank Broz, Chris S Crawford, Astrid Rosenthal-Von Der Pütten, Megan Strait, and Â· Laurel Riek. Reproducibility in human-robot interaction: Furthering the science of hri. *Current Robotics Reports 2022 3:4*, 3:281–292, 10 2022.

[132] *The 2030 agenda and the sustainable development goals: An opportunity for Latin America and the Caribbean.* 2018.

[133] Angel De-Juanas, Teresita Bernal Romero, and Rosa Goig. The relationship between psychological well-being and autonomy in young people according to age. *Frontiers in Psychology*, 11:559976, 12 2020.

[134] Carolina M. Henn, Carin Hill, and Lené I. Jorgensen. An investigation into the factor structure of the ryff scales of psychological well-being. *SA Journal of Industrial Psychology*, 42:12, 11 2016.

[135] Carol D. Ryff. Psychological well-being revisited: Advances in science and practice. *Psychotherapy and psychosomatics*, 83:10, 12 2014.

[136] Evgeny Morozov. *To save everything, click here : the folly of technological solutionism*. PublicAffairs, New York, 2013.

[137] Tamar Sharon. Blind-sided by privacy? digital contact tracing, the apple/google api and big techâs newfound role as global health policy makers. *Ethics and Information Technology*, 23:45–57, 11 2021.

[138] Julian Sheather. Patient autonomy. *BMJ*, 342:d680, 3 2011.

[139] Carol D. Ryff. Happiness is everything, or is it? explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57:1069–1081, 12 1989.

[140] Jennifer H. Pfeifer and Elliot T. Berkman. The development of self and identity in adolescence: Neural evidence and implications for a value-based choice perspective on motivated behavior. *Child Development Perspectives*, 12:158–164, 9 2018.

[141] Paul Formosa. Robot autonomy vs. human autonomy: Social robots, artificial intelligence (ai), and the nature of autonomy. *Minds and Machines*, 31:595–616, 12 2021.

[142] Carlos Gómez-Vírseda, Yves De Maeseneer, and Chris Gastmans. Relational autonomy: What does it mean and how is it used in end-of-life care? a systematic review of argument-based ethics literature. *BMC Medical Ethics*, 20:1–15, 10 2019.

[143] Vikki A. Entwistle, Stacy M. Carter, Alan Cribb, and Kirsten McCaffery. Supporting patient autonomy: The importance of clinician-patient relationships. *Journal of General Internal Medicine*, 25:741–745, 7 2010.

[144] Elli Zey and Sabine Windmann. Grassroots autonomy: A laypersons' perspective on autonomy. *Frontiers in Psychology*, 13:871797, 4 2022.

[145] Jan Keenan. A concept analysis of autonomy. *Journal of Advanced Nursing*, 29:556–562, 3 1999.

[146] Coralie J. Wilson and Frank P. Deane. Brief report: Need for autonomy and other perceived barriers relating to adolescentsâ intentions to seek professional mental health care. *Journal of Adolescence*, 35:233–237, 2 2012.

[147] Paul Benson. Free agency and self-worth. *The Journal of Philosophy*, 91:650–668, 12 1994.

[148] Catriona MacKenzie. Relational autonomy, normative authority and perfectionism. *Journal of Social Philosophy*, 39:512–533, 12 2008.

[149] Peter A. Ubel, Karen A. Scherr, and Angela Fagerlin. Autonomy: What's shared decision making have to do with it? *The American Journal of Bioethics*, 18:W11–W12, 2 2018.

[150] A. van Wynsberghe. Service robots, care ethics, and design. *Ethics and Information Technology*, 18:311–321, 12 2016.

Appendices

A Interview Form and Interview Questions

Research project title: Value-centered approach to Socially Assistive Robots design for young adults' mental health
Research investigator:
Research Participant Name:

Thank you for participating in this research. This document will outline the research goals and ask for explicit consent on data collection and analysis to ensure that you understand the purpose of your involvement and agree to the conditions of participation.

This research aims at critically investigating the potential use of socially assistive robots to support young adults' with mental health complaints related to anxiety and depression. This research aims to contribute to the field of Responsible Innovation through an investigation of values and practices of stakeholders, with the goal of supporting design practices that respect and promote human values.
Interviews with professionals in mental health and Human-Robot Interaction will be conducted and analysed. The data will be used to understand the context and values in current practices, attitudes and opinions on technology and robotics for mental health support.

The interview will take between 20 and 45 minutes. There are no anticipated risks associated with your participation. You do not have to answer any questions you do not wish to answer. Your participation is voluntary and you have the right to withdraw from the research at any moment without consequences by contacting the researcher. After the interview, you have 1 week to consider withdrawing your data from the project. You can review the interview topics and questions in the appendix of this document.

This research project involves making audio recordings of an interview with you. The audio recordings will be deleted after transcription. The transcriptions will be carried out by the primary investigator and will be de-identified before analysis. After de-identification it will no longer be possible to withdraw your data.
The transcription will be used for the purpose of analysis and will be kept confidential, stored in a secure server owned by Utrecht University. In the case of publication, your words may be quoted directly and pseudonyms will be used. Forms, and other documents created or collected as part of this study will be stored in a secure location on a server owned by Utrecht University.

If you decide to stop taking part in the study, want to access or withdraw your data, if you have questions, concerns, or complaints, please contact the primary investigator (b.ghedi@students.uu.nl), the UU's privacy department (privacy@uu.nl) or the Data Protection Officer of the UU (fg@uu.nl).

By signing this form I agree that:

- I am voluntarily taking part in this project. I understand that I don't have to take part, and I can stop the interview at any time;
- After 1 week from the interview it will no longer be possible to withdraw my data;
- The transcribed interview or extracts from it may be used as described above;
- I have read the Information sheet;

- I don’t expect to receive any benefit or payment for my participation;
- I have been able to ask any questions I might have, and I understand that I am free to contact the researcher with any questions I may have in the future.

Participant Signature Date (DD-MM-YY)

Researcher Signature Date (DD-MM-YY)

Appendix - Interview Guide

TOPIC 1: Background
How did you become involved with Human-Robot Interaction?
What motivates you in your work?
What are some of the difficulties or limitations you encounter in your profession?
Do you have experience with robots?

TOPIC 2: Values in mental health support practices
Many reports indicate that mental health-care resources are insufficient in meeting the requirements of people in need of assistance [1,2]. With regards to the focus population of this research, despite the significant frequency of mental health problems among young people, in particular emotional disorders, such as depression and anxiety, only one third of them seeks professional help [3,4]. Reasons are negative perceptions of help-seeking, a lack of mental health knowledge, and mental health stigma and embarrassment [5]. For those seeking help, some of the barriers are poor accessibility, high costs and long waiting lists [1].

Let's imagine a situation in which there is no money and time constraint. What do you think would be a desirable way to bridge this gap?

What are important characteristics of someone able to provide mental health support for young adults?

TOPIC 3: Technology
Technology-mediated interventions offer a possible solution to mitigate this issue by reaching under-served populations, reducing costs and improving accessibility to support [1].
Examples are teletherapy, self-help computer delivered programs, chatbots, smartphone applications, wearable devices.
Do you have experience with any of them?
What is your attitude towards technology for mental health?
Could technology be a solution?
What do you think is important to consider when deploying technology in the mental healthcare domain?

TOPIC 4: Robots
Robots that are made to support people socially and emotionally are known as "socially assistive robotics" or SARs. SARs are designed to interact with people in a way that is considerate of their needs and emotions. They have been gaining attention in research and clinical settings due to their ability to interact socially with patients, building affective relationships [1]. Their social presence, which is the extent to which they are perceived as a social entity, is influenced by their embodiment and social capabilities and it has a positive effect on motivation and engagement of users [6].

What do you think about the potential use of robots for addressing gaps in mental healthcare provision for young adults?
What features and capabilities would you implement?
What do you think are risks and benefits related to deploying mental health support robots which foster affective relationships?

What are important factors to keep in mind when designing and developing these robots?

Let’s imagine a world in which, to tackle the mental health crisis, the government of a country hands out personal mental support robots to young adults with emotional complaints. The robot is able to understand the environment through sensory inputs, communicate, build rapport with their users, provide comfort, companionship and deliver some forms of therapy.
Do you think this is a desirable scenario?
What do you think would happen?
Would you feel comfortable with someone close to you using it?
Could the situation cause harm? Impact autonomy?
Do you think it is necessary to understand how the robot works for it to be acceptable?
Do you think it is fair to address the issue in this way?

References

[1] Arielle Aj Scoglio, Erin D Reilly, Jay A Gorman, and Charles E Drebing. Use of social robots in mental health and well-being research: Systematic review.
[2] Elias Aboujaoude, Lina Gega, Michelle B. Parish, and Donald M. Hilty. Editorial: Digital interventions in mental health: Current status and future directions. Frontiers in Psychiatry, 11:111, 2 2020.
[3] Terhi Aalto-Setälä, Mauri Marttunen, Annamari Tuulio-Henriksson, Kari Poikolainen, and Jouko Lönnqvist. Psychiatric treatment seeking and psychosocial impairment among young adults with depression. Journal of Affective Disorders, 70:35–47, 6 2002.
[4] Vikram Patel, Alan J. Flisher, Sarah Hetrick, and Patrick McGorry. Mental health of young people: a global public-health challenge. The Lancet, 369:1302–1313, 4 2007.
[5] Koutra K, Pantelaiou V, Mavroeides G. Why Don’t Young People Seek Help for Mental Illness? A Cross-Sectional Study in Greece. Youth. 2023; 3(1):157-169.
[6] Eric Deng, Bilge Mutlu, and Maja J Mataric. Embodiment in socially interactive robots. Foundations and Trends in Robotics, 7(4):251–356, 2019.

B Coding Manual

Project Name: Value-centered approach to Socially Assistive Robots design for young adults' mental health

Coding Manual for Qualitative Interview Coding

This coding manual serves as a guide for systematically analyzing qualitative interview data collected as part of the research critically investigating the potential use of socially assistive robots to support young adults' with mental health complaints related to anxiety and depression. The purpose of this coding manual is to establish a standardized set of codes and definitions that correspond to the key topics discussed during the interviews.

Instructions:

- Carefully review each interview transcript and apply the relevant codes to segments of text that align with the identified themes or topics.
- If a segment of text pertains to multiple codes, assign all appropriate codes to ensure comprehensive analysis.

Code List:

1. Challenges in Research and Practice for HRI Professionals
   Definition: This code encompasses the obstacles, difficulties, and issues that HRI researchers encounter in their work, requiring attention and resolution to enhance the conditions and outcomes of their research and practice in the field of mental health.

   1.1. Collaboration: Collaboration between researchers, psychologists, and other professionals from diverse fields can be challenging due to differing perspectives and expectations.
   1.2. Systemic Issues: Issues arising from the way Academia functions.
   1.3. Incongruences: research trends going in a certain direction without proper motivation.

2. Challenges in Research and Practice for Psychology Professionals
   Definition: This code encompasses the obstacles, difficulties, and issues that psychology professionals encounter in their work, requiring attention and resolution to enhance the conditions and outcomes of their research and practice in the field of mental health.

   2.1. Therapy related issues: difficulties arising in therapy such as mitigating the conflicting wished of patients and their families.
   2.2. Institutionalisation and Funding: The funding system for mental health can be clunky and unable to provide professionals and institutions with the support and resources they need. No enough funds for training

3. Challenges of Applying Socially Assistive Robots in Mental Health Support

   Definition: This code refers to the potential issues, risks, and obstacles associated with the implementation of socially assistive robots for mental health support among young adults.

   3.1. Feasible integration : discussions about identifying useful cases for the technology and the target population or lack thereof. Considerations about what would make it acceptable and functional within one's life, society and mental healthcare.
   3.2. Social consequences: interacting with robots can lead to isolation from people and can lead to a reduced social skill set.
   3.3. Dependence: Relying heavily on robots for mental health support could lead to decreased human interaction skills and overreliance on technology.
   3.4. Limited Human Connection: Robots might lack the emotional depth and empathy that human interactions provide.
   3.5. Regulation and safety: Discussion about regulation and safety issues, such as hacking. The regulation of this technology is non-trivial. There is uncertainty on who should be liable for it and how its use woud translate in law and justice.
   3.6. Misinterpretations: Technical glitches or malfunctions could disrupt the support sessions.Robots might misinterpret verbal or nonverbal cues, leading to inappropriate responses or interventions. Users can misinterpret robots' intentions, projecting their own and misinterpreting behaviours.
   3.7. Technical limitations: current technical limitations translate to a limited scope of interaction, making them unable to provide appropriate mental health support.
   3.8. Assessment: the dynamic and complex nature of mental health and well-being makes the assessment of interventions 'in the wild' challenging. There may be unintended consequences.

4. Scenarios and Roles for Socially Assistive Robots in Mental Health Support
   Definition: This code pertains to the various scenarios and roles in which socially assistive robots could play a helpful and supportive role in enhancing the mental health and well-being of young adults.

   4.1. Therapy Support: Robots could assist in delivering therapy.
   4.2. In-between: Robots ca be used to direct of facilitate a connection with health proessionals or institutions.
   4.3. Therapist, Friend, Coach: discussions about SARs potential role as a therapist, friend or coach.
   4.4. Skills Building: Robots could aid users in practicing social or coping skills and build confidence in various social scenarios.
   4.5. Complementary: Robots and humans have complementary roles.
   4.6. Prevention: Robots could assist in prevention-related activities.

4.7. Data Collection: SARs data collection capabilities could be used to improve or support care.
4.8. Crisis Intervention: Robots could identify signs of distress and initiate appropriate crisis interventions.

5. Benefits of Socially Assistive Robots in Mental Health Support

Definition: This code pertains to the positive impacts and advantages that the application of socially assistive robots can have in the context of mental health support.

5.1. Accessibility: SARs can improve accessibility for mental health practices and resources.
5.2. Availability: SARs' availability can prove to useful in aiding mental health practices.
5.3. Reduction of Provider Burden: Robots can handle routine tasks, reducing the burden on mental health professionals and allowing them to focus on more complex cases.
5.4. Influence of Robots on Individuals: discussions about the impact, influence, and effects that robots have on individuals' thoughts, emotions, behaviors.

6. Features and Design Requirements for Socially Assistive Robots in Mental Health Support

Definition: This code refers to the essential features, functionalities, and design considerations that socially assistive robots should possess to effectively support mental health among young adults.

6.1. Communication: SARs features and capabilities which contribute to being able to communicate naturally with users.
6.2. Collaboration: collaboration between experts of different expertise is an important design requirement.
6.3. Personalization: Robots should be able to personalize their interactions and interventions to cater to each user's unique mental health goals and challenges.
6.4. User Education: Robots should educate users about their capabilities and limitations, fostering a clear understanding of how the technology can assist them.
6.5. Appearance: discussions about the looks of SARs
6.6. Privacy: discussion about privacy mechanisms.
6.7. Accessibility: features which ensure SARs are easy to use and accessible to various people.
6.8. Support and Monitoring: support and monitoring throughout the design, deployment and use fases is an important design requirement.

7. Considerations about Technology in Mental Health

   Definition: This code pertains to discussions surrounding various opinions, roles, and personal experiences related to the use of technology in the context of mental health support.

   7.1. Role: Opinions on the role that technology can play in promoting overall well-being.
   7.2. Accessibility: technology needs to be easy to use.
   7.3. Different tools work for different people: there is no one-size fit all.

8. Considerations about Mental Health Support Practices

   Definition: This code encompasses discussions about specific mental health practices, important factors to consider, personal experiences, and insights related to mental health interventions and support.

   8.1. Prevention: Conversations about the benefits of preventative initiatives.
   8.2. Features of a Helper: characteristics of someone able to help young adults with emotional complaints.
   8.3. Therapeutic Relationship: discussion about the role of the therapeutic relationship in mental health practices.
   8.4. Goal of therapy
   8.5. Societal issue: the mental health crisis exists at a societal level. It needs to be addressed through societal restructuring.
   8.6. Personalisation: discussion about the role of personalisation in mental health practices.
   8.7. Accessibility: discussions about the accessibility of mental health practices.

9. EU HLEG Values

   Definition: This code takes into considerations explicability, fairness, prevention of harm, and autonomy as they are understood in this context.

   9.1. Autonomy: SARs should respect and support users' autonomy in their mental health journey. The robot's interventions should not undermine users' agency.
   9.2. Prevention of Harm: SARs should contribute to preventing harm. SARs design and use should be focused on preventing harm arising from various sources.
   9.3. Explainability: tools and processes should be understandable and accountability should be ensured.
   9.4. Fairness: SARs’ potential contribution to the distribution of resources, impact on discrimination and prevention of the creation of new harms.

10. Care Values

Definition: This code takes into considerations care values as they are understood in this context.

10.1. Competence: the skill of providing good and successful care
10.2. Reciprocity:the care-receiver’s capacity to guide the caregiver and the instauration of a reciprocal interaction.
10.3. Responsibility: a willingness to respond and take care of need.
10.4. Attentiveness: proclivity to become aware of need.