<u>Original Paper</u>

Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy

Amelia Fiske[1], BA, PhD; Peter Henningsen[2], PhD, MD; Alena Buyx[1], MPhil, PhD, MD

[1]Institute for History and Ethics of Medicine, Technical University of Munich School of Medicine, Technical University of Munich, Munich, Germany
[2]Department of Psychosomatic Medicine and Psychotherapy, Klinikum rechts der Isar at Technical University of Munich, Munich, Germany

Corresponding Author:
Amelia Fiske, BA, PhD
Institute for History and Ethics of Medicine
Technical University of Munich School of Medicine
Technical University of Munich
Ismaninger Straße 22
Munich, 81675
Germany
Phone: 49 8941404041
Email: a.fiske@tum.de

Abstract

Background: Research in embodied artificial intelligence (AI) has increasing clinical relevance for therapeutic applications in mental health services. With innovations ranging from 'virtual psychotherapists' to social robots in dementia care and autism disorder, to robots for sexual disorders, artificially intelligent virtual and robotic agents are increasingly taking on high-level therapeutic interventions that used to be offered exclusively by highly trained, skilled health professionals. In order to enable responsible clinical implementation, ethical and social implications of the increasing use of embodied AI in mental health need to be identified and addressed.

Objective: This paper assesses the ethical and social implications of translating embodied AI applications into mental health care across the fields of Psychiatry, Psychology and Psychotherapy. Building on this analysis, it develops a set of preliminary recommendations on how to address ethical and social challenges in current and future applications of embodied AI.

Methods: Based on a thematic literature search and established principles of medical ethics, an analysis of the ethical and social aspects of currently embodied AI applications was conducted across the fields of Psychiatry, Psychology, and Psychotherapy. To enable a comprehensive evaluation, the analysis was structured around the following three steps: assessment of potential benefits; analysis of overarching ethical issues and concerns; discussion of specific ethical and social issues of the interventions.

Results: From an ethical perspective, important benefits of embodied AI applications in mental health include new modes of treatment, opportunities to engage hard-to-reach populations, better patient response, and freeing up time for physicians. Overarching ethical issues and concerns include: harm prevention and various questions of data ethics; a lack of guidance on development of AI applications, their clinical integration and training of health professionals; 'gaps' in ethical and regulatory frameworks; the potential for misuse including using the technologies to replace established services, thereby potentially exacerbating existing health inequalities. Specific challenges identified and discussed in the application of embodied AI include: matters of risk-assessment, referrals, and supervision; the need to respect and protect patient autonomy; the role of non-human therapy; transparency in the use of algorithms; and specific concerns regarding long-term effects of these applications on understandings of illness and the human condition.

Conclusions: We argue that embodied AI is a promising approach across the field of mental health; however, further research is needed to address the broader ethical and societal concerns of these technologies to negotiate best research and medical practices in innovative mental health care. We conclude by indicating areas of future research and developing recommendations for high-priority areas in need of concrete ethical guidance.

XSL•FO
RenderX

Introduction

Research in embodied artificial intelligence (AI) has increasing clinical relevance for therapeutic applications in mental health services, that is, in psychiatry, psychology, and psychotherapy. Innovations range from 'virtual psychotherapists' [1] to social robots in dementia care and autism disorder [2] and robots for sexual disorders [3]. Increasingly, artificially intelligent virtual and robotic agents are not only available for relatively low-level elements of mental health support, such as comfort or social interaction, but also perform high-level therapeutic interventions that used to be offered exclusively by highly trained, skilled health professionals such as psychotherapists [4]. Importantly, such 'virtual' or 'robotic therapists' include an artificially intelligent algorithm that responds independently of any expert human guidance to the client or patient through a virtually embodied presence, such as a face icon, or a physically embodied presence, such as a robotic interface. As such, these emerging applications are distinct from the many varieties of Web-based therapy, which usually involve either a human therapist, albeit remotely (telemedicine), or the patient herself, working independently with manuals, questionnaires, or other self-help materials [5].

Embodied AI applications in mental health care carry hopes of improving quality of care and controlling expenditure [6]. In addition, they also hold the promise of reaching underserved populations in need of mental health services and improving life opportunities for vulnerable groups. However, there is a persistent gap between current, rapid developments in AI mental health and the successful adoption of these tools into clinical environments by health professionals and patients. In addition, it has been demonstrated that the interventions are often designed without any explicit ethical considerations [7]. Furthermore, although studies often examine the effectiveness or ethical use of a single application, rarely do they consider the implications for the integration of AI across the field of mental health more broadly. In this paper, we argue that virtually and physically embodied artificially intelligent agents and applications have great potential in mental health care. However, their societal and ethical implications require further probing to identify pertinent concerns surrounding trust, privacy, and autonomy, as well as to anticipate concerns that may arise in the future. Identifying the broader ethical and societal implications of embodied AI is crucial for negotiating best research and medical practices in innovative mental health care. We conclude by indicating areas of future research and identifying points in need of ethical caution.

Overview: Existing Embodied Intelligent Applications

Although AI-enabled virtual and robot therapy has long been used across a number of medical fields [8-10], the integration of AI through the use of embodied agents is still at an early stage in mental health care; it is arguably the most recent addition to psychotherapeutic practice, supporting a host of emotional, cognitive, and social processes [11]. In what follows, we have sketched a range of applications with the aim of characterizing some of the embodied artificially intelligent innovations across the field of mental health. To maintain focus amid a broad and growing field, we have chosen to exclude from our analysis applications that are not intended to interact with patients, or that have no virtual presence or robotic interface; this includes AI-supported scanning and diagnostic tools. We have also excluded applications that may have a virtual or robotic interface but do not employ AI, such as telemedicine therapy (for further scholarship on this topic, please see [12-19]).

Virtually Embodied Artificially Intelligent Agents

AI-supported virtually embodied psychotherapeutic devices are currently developing at a rapid speed. For example, therapeutic apps such as Tess and other "chatbots" such as Sara, Wysa, and Woebot, which work over short message service text messaging, WhatsApp, or internet platforms, are being explored for addressing depression and anxiety. These applications come with interactive screen presences. Woebot and other programs engage with the patient like a virtual psychotherapist, with the aim of helping patients to recognize their emotions and thought patterns and to develop skills such as resilience or techniques for reducing anxiety. For example, using natural language processing, Tess is programed to flag expressions that indicate emotional distress. Often cited as a digital tool to reach underserved populations across the world that lack mental health services, the bots can explain to users the clinical terms for what they are experiencing—such as cognitive distortions—or provide concrete advice for recognizing and dealing with difficult situations [20]. Initial studies found that depression symptoms decreased with the use of Woebot more than groups who relied on electronic book resources [21], and another study found that Tess helped to reduce depression and anxiety among users [20].

A similar approach involves the use of avatars, such as the Avatar Project, for addressing persistent auditory hallucinations for patients with psychosis [22]. These usually involve computer-generated images of faces on computer screens or tablets that interact with a patient via intelligent algorithms. Avatars are also being explored in treatment of schizophrenia, for example, to improve medication adherence [23]. Similar to the Avatar Project, virtual reality–assisted therapy for schizophrenia often encourages patients to engage with the voices they hear through the use of an AI avatar. Initial studies found that the therapy could help in developing therapeutic targets [24] and also in particularly difficult cases of schizophrenia [25]. Another study found improvements in auditory visual hallucinations, symptoms of depression, and overall quality of life following therapy sessions for treatment-resistant schizophrenia patients [26]. 'Avatar coaches' have also been employed as part of an immersive virtual reality situation for treating the fear of heights [27] or as 'virtual patients' to provide medical students with lifelike interviewing practice [28]. Finally, avatars are also being implemented in risk prevention education, such as the Kognito program, which

uses an avatar to help college students and faculty identify others at risk for suicide [29].

Artificially Intelligent Robot Therapy

In addition to these virtually embodied therapeutic applications, clinicians and scientists are exploring the translation of innovations at the intersection of AI and robotics into the clinic. For example, intelligent animal-like robots such as Paro, a fuzzy harp seal, are increasingly being used to help patients with dementia. Paro, along with the large furry eBear, is part of a class of 'companion bots,' engaging individuals as at-home health care assistants, responding to speech and movement with dynamic 'dialog', or seeking to help elderly, isolated, or depressed patients through companionship and interaction. Several studies have examined the role of such robots in reducing stress, loneliness, and agitation and in improving mood and social connections [30,31]. Thus far, the outcomes are promising [32,33].

AI robots also provide opportunities for different forms of engagement with children suffering from autism spectrum disorders (ASDs) [34]. Children with autism have been found to react positively to robots, even in cases where they have trouble interacting with others [35]. The Kaspar robot has demonstrated potential for integration in current education and therapy interventions [36] and is being investigated for the potential to improve social skills among children [37]. Similarly, RoboTherapy is an example of socially assistive robotics designed to help children with ASDs to develop social skills, and the robot Nao is designed to improve facial recognition and appropriate gaze response. The aim of such robotic interaction is to learn appropriate social skills (eg, imitation, taking turns, staying engaged, and empathy), with the hope that children can then apply the skills learned with the robot peer to their relationships with human peers. Initial studies are promising; individuals with ASDs performed better with their robot partners than human therapists, responded with social behaviors toward robots, and improved spontaneous language during therapy sessions [38]. However, the devices are still being developed and are not yet in wider therapeutic use.

AI-enabled robots are also being explored across a variety of other mental health areas including mood and anxiety disorders, children with disruptive behavior, and patients who may not have a specific diagnosis but who would benefit from assistance with mental health concerns [39]. Perhaps, most controversially, artificially intelligent robots have entered the field of human sexuality. Companies are now offering adult sex robots such as Roxxxy, which can speak, learn their human partners' preferences, register touch, and provide a form of intimate companionship. Although the range of medical applications that sex robots can reportedly address remains debated, these include meeting the sexual needs of disabled and elderly individuals or as part of therapy for concerns such as erectile dysfunction, premature ejaculation, and anxiety surrounding sex [40]. Furthermore, some researchers have asked if sex robots could help to reduce sex crimes such as rape and assault or be used for treatment of paraphilia, such as pedophilia [3,41].

Methods

Based on a thematic literature analysis and established principles of medical ethics, an analysis of the ethical and social aspects of currently embodied AI applications was conducted across the fields of Psychiatry, Psychology, and Psychotherapy. To enable a comprehensive evaluation, the analysis was structured around the following three steps: assessment of potential benefits; analysis of overarching ethical issues and concerns; discussion of specific ethical and social issues of the interventions.

Results

Ethical and Social Implications and Concerns

The devices and applications described above have yet to be integrated into widespread clinical use. However, in view of the speed of research and development trajectories of these applications, it is reasonable to expect that therapeutic chatbots, avatars, socially assistive devices, and sex robots will soon translate into broader clinical applications in earnest. In some cases, initial ethical assessments are already available [42,43]; however, most of these studies focus on a single application. In general, for most of the applications we are discussing, large-scale rigorous research studies have not yet been conducted or are still in pilot stages [44-47]. Even in Web-based non-AI applications, evidence of patient acceptance and treatment outcomes in routine care is still limited and mixed [48,49]; there has so far been very little research on patient acceptance and contingent treatment outcomes of embodied AI applications in mental health fields. As with any medical innovation, the effects, impacts, and clinical utility of the applications can only be fully assessed once evidence has improved [50,51].

To enable responsible and responsive innovation and clinical translation into the field of mental health, further and more in-depth analysis of the ethical and social implications of embodied AI is necessary to flag areas of concern. Early identification of ethical issues can help researchers, designers, and developers consider these concerns in the design and construction of the next generation of AI agents and robots for mental health. In the following sections, we have provided an analysis of benefits, challenges, and risks of embodied AI in mental health from an ethical perspective. Beginning with a discussion of potential benefits, we have then turned to risks and challenges, followed by immediate concerns in clinical application and long-term effects.

Anticipated Benefits

All of the aforementioned intelligent applications promise significant benefits for the field of mental health, satisfying many aspects of the ethical principle of beneficence [52]. From a clinical point of view, the use of embodied AI applications holds the potential to open new avenues for intervention in places where there are still significant unmet health needs. AI interventions might be particularly well placed for detecting mental health concerns early on, for reaching high-risk groups such as veterans, or for those who are concerned about the social

stigma associated with psychotherapy [53]. In some cases, patients may respond positively and productively to the fact that the counterpart is *not* a human therapist [54-56]. In one study, subjects overwhelmingly preferred the virtual agent over the human counterpart when being discharged from the hospital because they could self-direct the pace of information—something that is especially important for low-literacy patients [57]. Thus, in mental health services, nonhuman virtual or robotic applications might be preferable for some patients, reducing embarrassment when asking for specific information or services or feelings of shame when admitting noncompliance with a treatment plan. Embodied AI in mental health could also help to empower particular patient groups (such as those who are less familiar with the medical system), thereby helping to improve trust and openness between patients and the medical system. Another important advantage of AI applications is that many of them are low-threshold and self-administered, such that people who do not have an acute condition can elect services without going through the time-consuming process of being screened and admitted into the health care system.

Arguably, the greatest benefit of AI applications is structural, namely the potential to reach populations that are difficult to treat via traditional routes of provision. The provision of some mental health services, for example, through low-threshold, convenient therapeutic interventions via chatbots or avatars may be particularly beneficial for populations living in resource-poor settings. For those living in remote or rural locations or in settings where on-site mental health services are scarce, intelligent applications can increase geographical access and provide some minimal mental health care services where they are otherwise absent. The same may also be true for individuals living in higher income countries who do not have insurance or whose insurance does not cover therapy. Furthermore, it is likely that there are individuals who, for various reasons, do not respond to more traditional clinical services and might prefer low-threshold interventions that can be conducted in the privacy of their homes or on the go. For all of these patients, AI applications could complement existing services or constitute an entry point for pursuing more standard clinical interventions in the future.

In sum, embodied AI interventions may offer entirely new modes of treatment that are potentially more successful than traditional modalities either because they address hard-to-reach populations or because patients respond better to them. Whether, and for which conditions this is the case, requires further investigation. However, given that broadly speaking, conditions such as ASD and sexual dysfunction are increasing in incidence and patient populations with these and many other mental health conditions continue to have unmet health needs [58-62], increased exploration of embodied AI in these fields is promising.

Finally, there are also clear benefits of having a virtual or robotic therapist that is always accessible, has endless amounts of time and patience, never forgets what a patient has said, and does not judge [63,64], thus potentially offering a service that is highly reliable and particularly well-suited to certain patient populations. If integrated into a scaled provision of services,

AI-enabled applications could provide support for mild cases of depression and other nonacute conditions [65], therefore helping health professionals to devote more time to the most severe cases. In view of overall increasing burden of illness in mental health and against a background of limited resources, these are important benefits to consider. However, it is likely that embodied AI may not be warmly received by all mental health care professionals, and some may even have serious misgivings about its use because of ethical or clinical concerns. Thus far, there has been no substantial review of the reception of AI across or within specific mental health fields, marking an area in need of further research.

Overarching Ethical Concerns

Harm Prevention and Data Ethics Issues

To satisfy the well-established ethical principle of nonmaleficence, more robust research is needed on embodied AI applications in mental health to prevent harm both within therapeutic encounters and in cases where robots could malfunction or operate in unpredictable ways. For instance, in interviewing respondents working with AI robotic technologies, Cresswell et al discuss an example of a woman who was stuck in an elevator with a robot and another who was run over by a robot [6]. Chatbots and avatars could also stop working or malfunction. Hence, it needs to be discussed if embodied AI devices—potentially including virtual agents and freely available mental health applications—should require the same kind of rigorous risk assessment and regulatory oversight that other medical devices are subject to before they are approved for clinical use.

Similar to other devices employed in medical settings, the use of any AI applications in mental health care requires careful consideration surrounding data security of devices that communicate personal health information, the ways that the data generated is used, and the potential for hacking and nonauthorized monitoring [66,67]. Clear standards are needed on issues surrounding confidentiality, information privacy, and secure management of data collected by intelligent virtual agents and assistive robots as well as their use for monitoring habits, movement, and other interactions [68,69]. Concerns around privacy may be amplified as the amount of data collected continues to expand; for example, we anticipate that applications that integrate video data would need to have specific privacy protections in place for the communication of sensitive information, or information pertaining to individuals other than the consenting patient.

Lack of Guidance on Development, Clinical Integration, and Training

With embodied AI being one of the newest and most rapidly changing areas of psychological and psychiatric research and treatment, existing legal and ethical frameworks are often not closely attuned to these changes. Rather than providing regulatory guidance, there is the risk that the 'gaps' between application and ethical frameworks would only be addressed once harm had already occurred [6]. Again, this is the case with many forms of emerging medical technologies; however, in view of the rapid pace of translation of embodied AI into

practice in settings where traditional health technology assessment and medical oversight systems are not fully applicable—for example, through freely available therapy chatbots—this is an important concern. Although anticipating the ethical and legal questions that will emerge alongside future developments is difficult, active reflection on the 'regulatory fit' for embodied mental health AI is necessary. Initiatives for establishing guidelines are emerging, including the online collaboratively generated document "Moral Responsibility for Computing Artifacts: The Rules," or the recent "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations" [70]. However, thus far, no guidance exists that is specific to the field of mental health services; pointing to the need for the development of further recommendations to better guide advances in this area.

In addition to a lack of guidance on the development of these interventions for design, use, and regulatory questions, so far, there are also no frameworks available on how medical professionals can effectively engage with and train for increased use of embodied AI in the clinic; that is, although there is an increasing body of both academic and popular literature on how embodied AI can be integrated into clinical practice in mental health, there remains a lack of high-level guidance from professional bodies on the best use of AI in mental health services [15,71-73]. There are also no recommendations available on how to train and prepare young doctors for a mental health field in which such tools will increasingly be used by patients. Thus, further ethical guidelines are needed that are specific to assisting mental health professionals who will be supervising patients who have, or possibly will, engage with AI services.

Potential for Misuse to Reduce Service Provision

An ethically informed integration of AI should also consider questions of a just provision of mental health care [52]. There is the worry that the incorporation of embodied AI in mental health could be a justification for replacing established services, resulting in fewer available health resources or principally AI-driven services, thereby potentially exacerbating existing health inequalities. Many proponents insist that although informed by evidence-based psychotherapeutic approaches, chatbots, for instance, are not intended to replace therapists entirely. In some cases, forms of 'blended' care involving both in-person and virtual forms of therapy are being explored [74], which might also be appropriate for intelligent applications. Blended care models potentially offer the opportunity to draw on the strengths of both AI applications and in-person clinical supervision. However, whether or not it is appropriate to implement AI applications in mental health care depends in part on the availability of other resources in that area. As noted, in cases with limited mental health services, AI applications could provide a needed resource that is decidedly better than no services at all. However, at this point, AI mental health services are not a substitute or a stand-in for the kind of robust, multitiered mental health care available in high-resource health care systems. Appropriately considering the status quo of mental health resources in each context is thus highly relevant from an ethical perspective [75]. Otherwise, AI tools in mental health could be used as an excuse for reducing the provision of

high-quality, multilayered care by trained mental health professionals in low-resource settings.

Discussion

Specific Challenges in Application

Risk-Assessment, Referrals, and Supervision

Considering the application of embodied AI tools in mental health practice, a host of specific challenges need to be kept in mind: mental health professionals have an ethical responsibility to inform other service providers as well as third parties or authorities if a patient indicates that they are a threat to themselves or to another individual. How this would work in artificially intelligent interventions, particularly when there is no supervision of the interaction between the AI agent and the patient by a qualified health professional, remains to be determined. It is unclear when, and how, assistive robots that patients have in their homes, or freely available virtual agents and chatbots, would effectively connect at-risk individuals with appropriate services, including hospitalization and other protections. This scenario is particularly relevant in the aforementioned situation of using AI mental health applications to extend access to rural, hard to reach, or uninsured populations. In these cases, some provision of service is arguably better than nothing. However, what should be done if, for example, a therapy bot detects through speech patterns that an individual is at higher risk for self-harm, yet appropriate referral services are not available in the area?

AI applications engaged in therapeutic relationships with clients will likely also need to be bound by similar ethical guidelines as those that bind mental health professionals. However, so far, how an AI duty of care or a code of practice on reporting harm should be operationalized is entirely unclear. An obvious suggestion would be to always mandate supervision through a qualified mental health clinician—when a human therapist evaluates a patient's expression of self-harm, she also considers contextual information in her interpretation of the level of risk. Whether, and to what degree, robotic therapists are able to do this remains unclear. However, many AI applications are available outside established mental health settings; in addition, the capacity of computerized methods to identify and predict psychiatric illness are increasing [12,13], as are their skills of therapeutic interaction and communication. Thus, the question of whether supervision of embodied AI in health should always be provided, and how such a requirement could be successfully implemented, remains a subject for further debate.

Respecting and Protecting Patient Autonomy

Another concern for the application of embodied AI in mental health practice centers on enabling and respecting patient autonomy [52]. These are novel technologies that require assessment to guarantee that patients fully understand how the application or avatar works in order to ensure that a patient does not misunderstand or mistake the intelligent system for a human-driven application. For instance, it would be problematic if a patient were to assume that 'at the other end' of the chatbot there is a doctor communicating or reviewing her messages. Furthermore, obtaining consent for applications used outside

of medical systems raises thorny concerns. For instance, an elderly person or a person with intellectual disabilities may not be able to understand what a robot is or what a robot does when it is installed at home to monitor the patient's activities, risking privacy infringement [42], manipulation, and even coercion if the conditions for informed consent are not satisfied. Such questions make consenting to surveillance, interaction, and data collection with the robot challenging matters. Distinctions could be drawn between interventions that are seen as helping and monitoring, as opposed to those that run the risk of manipulation and coercion; however, these lines are often blurred both in theory and in practice [76]. How AI applications should evaluate if a patient has fully understood the information provided when giving consent, and how to proceed in cases where it is not possible for individuals to provide consent, such as children, patients with dementia, those with intellectual disabilities, or those in acute phases of schizophrenia, needs to be addressed.

Another area of particular concern in relation to matters of promoting autonomy in the use of AI in mental health care is the engagement of vulnerable populations. People have been shown to be more compliant when a robot asks them to do something as compared with a person [9]. Although this could lead to better results when helping patients with autism or those needing to make difficult behavioral changes, the concern exists that people could be manipulated or coerced into doing things that they should not or that they have not fully thought through, either because of the novelty of the device or because of a lack of companions with whom to discuss alternatives. Some studies have made a distinction between a "suspension of disbelief" when anthropomorphizing a robot caregiver and deception *per se* [42,77], but this is a line that requires further investigation in practice.

Nonhuman Therapy?

In general, the question remains as to whether there are aspects of the therapeutic encounter that cannot be achieved through AI. Some therapeutic benefits may be difficult to anticipate, or highly specific to a particular individual's relationship to his or her therapist. One study found that embodied conversational agents had difficulty evaluating a user's emotional state in a real-time dialog and that the absence of a human therapist in Web-based mental health interventions for treating depression and anxiety had a negative effect on user adherence to the programs [78]. In the treatment of insomnia, some patients indicated that they missed having a human therapist [79]. Relatedly, because robots and artificially intelligent systems blur previously assumed boundaries between reality and fiction, this could have complex effects on patients. Similar to therapeutic relationships, there is the risk of transference of emotions, thoughts, and feelings to the robot. In particular, given that many of the target populations are vulnerable because of their illness, age, or living situation in a health care facility, there is the additional concern that patients would be vulnerable in their engagements with the robot because of their desire for company or to feel cared for [80]. Unlike with a therapist, however, there is no person on the other side of this transference. Whether robot therapists will ever be able to deal adequately with such transference remains to be seen. Further concerns are likely to emerge in practice; thus, embodied AI therapeutic aids

need to also be evaluated carefully for unanticipated differences with standard therapy modalities.

Ethical Issues in Algorithms

It is necessary to note that AI mental health interventions work with algorithms, and algorithms come with ethical issues. It has been well-established that existing human biases can be built into algorithms, reinforcing existing forms of social inequality [81]. This raises the concern that AI-enabled mental health devices could also contain biases that have the potential to exclude or harm in unintended ways, such as data-driven sexist or racist bias or bias produced by competing goals or endpoints of devices [82,83]. Following other calls for transparency [84], the algorithms used in artificially intelligent applications for mental health purposes could be similarly open to scrutiny. This may require investing additional time in explaining to patients (and their families) what an algorithm is and how it works in relation to the therapy provided [85]. However, how to best do this, in particular with patients with compromised mental capacities, requires further consideration.

Concerns Regarding Long-Term Effects

Apart from these more immediate concerns, the implementation of embodied AI into mental health services also raises a number of broader questions regarding long-term impacts on patients, the mental health community, and society more widely. For instance, it has been noted that long-term use of AI interventions could lead to some patients or patient groups becoming overly attached to these applications. A study by Cresswell et al noted that robots that aim to alleviate loneliness or provide emotional comfort carry the risk that the patients they work with could become dependent on them [6]. More broadly, others have raised questions about ways that robots could contribute to changing social values surrounding care or situations in which caregiving is increasingly 'outsourced' to robotic aids. The impact of intelligent robots on relationships, both human-robot and human-human relationships, is an area that requires further probing, as do potential effects on identity, agency, and self-consciousness in individual patients. Specifically, research into the effectiveness of these applications needs to cover not only if the social skills of children with ASD are improved by working with robots but also their ability to apply these skills to relationships with other humans. Similarly, if a sex robot is provided therapeutically to an individual with paraphilia, the effects of this on the targeted behaviors with other humans also needs to be evaluated. The risk exists that if robotic interventions are not translatable to improving human interaction, that they merely remain a way of improving human relations with machines, or worse, an outlet that further limits human-to-human relationships. Similarly, engagement with embodied intelligent devices could also have important effects on the individual, such as on personal sense of identity or agency.

The integration of AI devices into our everyday lives and medical care is undoubtedly changing social expectations and practices of communication. There are essential differences between communicating with an AI device and communicating with another human. Anecdotal findings suggest that some users often speak to assistive devices such as Siri or Alexa in a curter or ruder manner than they would to a human [86]. Importantly,

perceptions of the devices can vary by users: children often understand these devices differently than adults, sometimes attributing human characteristics to the device or believing that the device has a real individual inside [87]. Extrapolating from this example, it is clear that the ways that individuals interact with the AI applications in their lives can have implications for communication and social interaction. How this will evolve as more patients have the opportunity to interact with AI applications as part of their mental health care requires further empirical investigation to catch problematic trends early and correct for future development.

A related concern of objectification exists for some areas of AI applications, such as sex robots. The use of 'sexbots' has already been notably controversial, with scholars objecting that sexual dysfunction depends on a range of physical, psychological, and sociocultural factors that are profoundly relational and reciprocal. Rather than addressing issues of isolation associated with sexual dysfunction, robots might aggravate it or contribute to reductionist understandings of sexual violence [88]. It has been cautioned that the use of sex robots—also available in childlike models or programed with personalities such as "Frigid Farrah" to resist sexual advances—could instead increase the occurrence of sex crimes, normalize the production of social inequalities surrounding the male gaze [89], and contribute to unwanted sexual encounters. Furthermore, the creation of humanoid robots for use in sexual dysfunction raises concerns that it could reinforce or even legitimize the objectification of humans, in particular women and children [3,88]. As the use of AI in many therapeutic applications has not yet been validated in randomized controlled trials (RCTs), there is the risk that particular applications might make problems such as sexual violence worse. More broadly, embodied AI applications necessarily involve a relatively narrow understanding of illness. For instance, sex robots may help with some medical concerns but do not address other determinants of illness that would have to be taken into account from a bio-psycho-social understanding of mental health illness. Widespread AI use could thus exacerbate trends of reductionism in mental health.

Ideas around embodied AI are culturally and historically shaped. Whether providing motivational interviews in therapy [64], acting as embodied conversational agents for mental disorders [47], or working with populations with intellectual disabilities [90], discussion of embodied AI often turns to worries surrounding the limits of human control over technology. Conjuring images of the Terminator or other depictions of the nonhuman in science fiction or cinema, such tools can carry with them negative or scary associations that bring the issue of trust in medical practice into new light [6]. However, exposure to robotic devices, or living in places with positive or caring associations with robots, can influence the adoption of AI devices in different settings [91]. Initiatives that integrate embodied AI into health care practices need to be duly attuned to existing cultural understandings of the role of technology in social lives, and work to ensure that trust between patient and provider, or patient and the health care system, is not eroded.

Finally, AI agents for mental health raise fundamental questions about what it is to be human [6]. One of the principal contributions of science and technology studies scholarship has been to show how humans do not simply act upon objects but rather our relationships with objects also alters, transforms, and imposes limits upon human activity [92]. Interaction with embodied AI agents, just like interaction with other individuals or a therapist, alters behaviors and understandings of the world. Although social relationships are characterized by reciprocity, relationships with intelligent devices are neither mutual nor symmetric. In particular, some have raised the concern that interacting more with artificial agents may lead some individuals to engage less with other people around them or to develop forms of intimacy with intelligent robots [93], raising concerns specific to the use of robots with children or those with intellectual disabilities. As mentioned, people develop attachments to objects and have been shown to also develop attachments to simpler robotic systems such as AIBO. Thus, it is likely that as more intelligent and autonomous devices are developed, human relationships with them will become even more complicated [94].

Conclusions

In light of the demonstrated benefits and potential, such as expanding the reach of services to underserved populations or enhancing existing services provided by mental health professionals, embodied AI has emerged as an exciting and promising approach across the field of mental health. At present, the quality of research on embodied AI in psychiatry, psychology, and psychotherapy is varied, and there is a marked need for more robust studies including RCTs on the benefits and potential harms of current and future applications.

This is still an emerging field, and any analysis of ethical implications can only be preliminary at this point. However, a few conclusions and recommendations are warranted, based on the considerations presented in this paper:

1. It is necessary to develop clear guidance on whether (and which) embodied AI applications should be subject to standard health technology assessment and require regulatory approval. This should include a set of broader provisions for the use of AI services outside the supervision of a health care professional.
2. Professional associations in mental health should develop guidelines on the best use of AI in mental health services as well as recommendations on how to train and prepare young doctors for wide-spread use of embodied AI in mental health, including blended care models.
3. AI tools in mental health should be treated as an additional resource in mental health services. They should not be used as an excuse for reducing the provision of high-quality care by trained mental health professionals, and their effect on the availability and use of existing mental health care services will need to be assessed.
4. To satisfy duties of care and reporting of harm, ideally embodied AI should remain under the supervision of trained mental health professionals. Any applications offered outside of mental health care settings, such as apps and bots, should be required to demonstrate reliable pathways of risk-assessment and referral to appropriate services.
5. Embodied AI should be used transparently. Guidance on how to implement applications in a way that respects patient

autonomy needs to be developed, for example, regarding when and how consent is required and how to best deal with matters of vulnerability, manipulation, coercion, and privacy.

6. AI algorithms in mental health need to be scrutinized, for example, for bias. Ideally, health professionals should be trained in communicating to their patients the role of the algorithms used in different applications they might be using or consider using, and such algorithms should be open for public debate and shaping.

7. Increased use of embodied AI should be accompanied by research that investigates both direct and indirect effects on the therapeutic relationship, other human-human relationships, and effects on individual self-consciousness, agency, and identity. Long-term effects, ranging from health reductionism to increased objectification and impacts on our understandings of what it means to be human, need to be monitored.

Conflicts of Interest

None declared.

References

1. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health 2018 Apr 23;5(2):e32 [FREE Full text] [doi: 10.2196/mental.9423] [Medline: 29685865]

2. Góngora Alonso S, Hamrioui S, de la Torre Díez I, Motta Cruz E, López-Coronado M, Franco M. Social robots for people with aging and dementia: a systematic review of literature. Telemed J E Health 2018 Aug 23 Epub ahead of print(forthcoming). [doi: 10.1089/tmj.2018.0051] [Medline: 30136901]

3. Torjesen I. Sixty seconds on . . . sex with robots. Br Med J 2017 Dec 11;358:j3353. [doi: 10.1136/bmj.j3353] [Medline: 28698234]

4. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]

5. Mehrotra S, Kumar S, Sudhir P, Rao GN, Thirthalli J, Gandotra A. Unguided mental health self-help apps: reflections on challenges through a clinician's lens. Indian J Psychol Med 2017;39(5):707-711 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_151_17] [Medline: 29200577]

6. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]

7. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Sci Eng Ethics 2018 Dec;24(4):1035-1055. [doi: 10.1007/s11948-017-9976-1] [Medline: 28940133]

8. Calderita LV, Manso LJ, Bustos P, Suárez-Mejías C, Fernández F, Bandera A. THERAPIST: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. JMIR Rehabil Assist Technol 2014 Oct 07;1(1):e1 [FREE Full text] [doi: 10.2196/rehab.3151] [Medline: 28582242]

9. Broadbent E. Interactions with robots: the truths we reveal about ourselves. Annu Rev Psychol 2017 Jan 03;68:627-652. [doi: 10.1146/annurev-psych-010416-043958] [Medline: 27648986]

10. Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Using artificial intelligence (Watson for Oncology) for treatment recommendations amongst Chinese patients with lung cancer: feasibility study. J Med Internet Res 2018 Sep 25;20(9):e11087 [FREE Full text] [doi: 10.2196/11087] [Medline: 30257820]

11. Eichenberg C, Küsel C. Deutsches Ärzteblatt. 2018. Roboter in der Psychotherapie: Intelligente artifizielle Systeme URL: https://www.aerzteblatt.de/pdf.asp?id=199391 [accessed 2019-04-24] [WebCite Cache ID 77rn5Im3V]

12. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015;1:15030 [FREE Full text] [doi: 10.1038/npjschz.2015.30] [Medline: 27336038]

13. LaFrance A. The Atlantic. 2015 Aug 26. Computers Can Predict Schizophrenia Based on How a Person Talks URL: https://www.theatlantic.com/technology/archive/2015/08/speech-analysis-schizophrenia-algorithm/402265/ [accessed 2019-04-22] [WebCite Cache ID 77pavCxgw]

14. Venturi D. Medium Corporation. 2017 Jan 26. How AI is revolutionizing mental health care URL: https://medium.com/@davidventuri/how-ai-is-revolutionizing-mental-health-care-a7cec436a1ce [accessed 2019-04-24] [WebCite Cache ID 77rnD5WjU]

15. Luxton D. Artificial Intelligence in Behavioral and Mental Health Care. Amsterdam, Boston: Academic Press; 2015:308.

16. Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, et al. Predicting the risk of suicide by analyzing the text of clinical notes. PLoS One 2014;9(1):e85733 [FREE Full text] [doi: 10.1371/journal.pone.0085733] [Medline: 24489669]

17. Kuwamura K, Nishio S, Sato S. Can we talk through a robot as if face-to-face? Long-term fieldwork using teleoperated robot for seniors with Alzheimer's disease. Front Psychol 2016;7:1066 [FREE Full text] [doi: 10.3389/fpsyg.2016.01066] [Medline: 27486416]

18. Yamazaki R, Nishio S, Ishiguro H, Nørskov M, Ishiguro N, Balistreri G. Acceptability of a teleoperated android by senior citizens in Danish society. Int J of Soc Robotics 2014 Jul 15;6(3):429-442. [doi: 10.1007/s12369-014-0247-x]

19. Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, et al. Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study. J Med Internet Res 2018 Sep 21;20(9):e264 [FREE Full text] [doi: 10.2196/jmir.9460] [Medline: 30249588]

20. Sachan D. Self-help robots drive blues away. Lancet Psychiatry 2018 Jul;5(7):547. [doi: 10.1016/S2215-0366(18)30230-X]

21. Fitzpatrick KK, Darcy A, Vierhile M. 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 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

22. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):31-40. [doi: 10.1016/S2215-0366(17)30427-3]

23. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e18 [FREE Full text] [doi: 10.2196/mhealth.7030] [Medline: 28223265]

24. Dellazizzo L, Percie du Sert O, Phraxayavong K, Potvin S, O'Connor K, Dumais A. Exploration of the dialogue components in Avatar Therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. Clin Psychol Psychother 2018 Nov;25(6):878-885. [doi: 10.1002/cpp.2322] [Medline: 30221431]

25. Dellazizzo L, Potvin S, Phraxayavong K, Lalonde P, Dumais A. Avatar therapy for persistent auditory verbal hallucinations in an ultra-resistant schizophrenia patient: a case report. Front Psychiatry 2018;9:131 [FREE Full text] [doi: 10.3389/fpsyt.2018.00131] [Medline: 29713292]

26. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. Schizophr Res 2018 Feb 24 Epub ahead of print(forthcoming). [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]

27. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry 2018 Aug;5(8):625-632. [doi: 10.1016/S2215-0366(18)30226-8]

28. Ohio State University. 2015 Feb 25. Virtual patient: Avatar shows emotions as he talks to med students URL: https://www.sciencedaily.com/releases/2015/02/150225151639.htm [accessed 2019-04-22] [WebCite Cache ID 77pci9R3P]

29. Rein BA, McNeil DW, Hayes AR, Hawkins TA, Ng HM, Yura CA. Evaluation of an avatar-based training program to promote suicide prevention awareness in a college setting. J Am Coll Health 2018 Jul;66(5):401-411. [doi: 10.1080/07448481.2018.1432626] [Medline: 29461940]

30. Wada K, Shibata T. Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. IEEE Trans Robot 2007 Oct;23(5):972-980. [doi: 10.1109/TRO.2007.906261]

31. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. 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 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]

32. Griffiths A. The Guardian. 2014 Jul 08. How Paro the robot seal is being used to help UK dementia patients URL: https://www.theguardian.com/society/2014/jul/08/paro-robot-seal-dementia-patients-nhs-japan [accessed 2019-04-22] [WebCite Cache ID 77pdxW9XI]

33. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. J Am Med Dir Assoc 2012 Feb;13(2):114-120.e1. [doi: 10.1016/j.jamda.2010.10.002] [Medline: 21450215]

34. Grossard C, Palestra G, Xavier J, Chetouani M, Grynszpan O, Cohen D. ICT and autism care: state of the art. Curr Opin Psychiatry 2018 Nov;31(6):474-483. [doi: 10.1097/YCO.0000000000000455] [Medline: 30256263]

35. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012;14:275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]

36. Huijnen CA, Lexis MA, Jansens R, de Witte LP. How to implement robots in interventions for children with autism? A co-creation study involving people with autism, parents and professionals. J Autism Dev Disord 2017 Oct;47(10):3079-3096 [FREE Full text] [doi: 10.1007/s10803-017-3235-9] [Medline: 28685411]

37. Mengoni SE, Irvine K, Thakur D, Barton G, Dautenhahn K, Guldberg K, et al. Feasibility study of a randomised controlled trial to investigate the effectiveness of using a humanoid robot to improve the social skills of children with autism spectrum disorder (Kaspar RCT): a study protocol. BMJ Open 2017 Dec 22;7(6):e017376 [FREE Full text] [doi: 10.1136/bmjopen-2017-017376] [Medline: 28645986]

38. Pennisi P, Tonacci A, Tartarisco G, Billeci L, Ruta L, Gangemi S, et al. Autism and social robotics: a systematic review. Autism Res 2016 Feb;9(2):165-183. [doi: 10.1002/aur.1527] [Medline: 26483270]

39. Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use. Clin Psychol Rev 2015 Feb;35:35-46. [doi: 10.1016/j.cpr.2014.07.001] [Medline: 25462112]

XSL•FO
RenderX

40. Sharkey N, van Wynsberghe A, Robbins S, Hancock E. Foundation for Responsible Robotics. 2017. Our Sexual Future with Robots URL: https://responsible-robotics-myxf6pn3xr.netdna-ssl.com/wp-content/uploads/2017/11/FRR-Consultation-Report-Our-Sexual-Future-with-robots-.pdf [accessed 2019-04-24] [WebCite Cache ID 77rnzVe17]

41. Torjesen I. Society must consider risks of sex robots, report warns. Br Med J 2017 Dec 05;358:j3267. [doi: 10.1136/bmj.j3267] [Medline: 28679604]

42. Vandemeulebroucke T, Dierckx de Casterlé B, Gastmans C. The use of care robots in aged care: a systematic review of argument-based ethics literature. Arch Gerontol Geriatr 2018 Jan;74:15-25. [doi: 10.1016/j.archger.2017.08.014] [Medline: 28926749]

43. Coeckelbergh M, Pop C, Simut R, Peca A, Pintea S, David D, et al. A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: ethical acceptability, trust, sociability, appearance, and attachment. Sci Eng Ethics 2015 Apr 17;22(1):47-65. [doi: 10.1007/s11948-015-9649-x]

44. Riek L. Robotics technology in mental health care. In: Artificial Intelligence in Behavioral Health and Mental Health Care. Cambridge, Massachusetts: Academic Press; 2015:185-203.

45. Piette JD, Krein SL, Striplin D, Marinec N, Kerns RD, Farris KB, et al. Patient-centered pain care using artificial intelligence and mobile health tools: protocol for a randomized study funded by the US Department of Veterans Affairs Health Services Research and Development Program. JMIR Res Protoc 2016 Apr 07;5(2):e53 [FREE Full text] [doi: 10.2196/resprot.4995] [Medline: 27056770]

46. Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. JMIR Ment Health 2018 Jul 31;5(3):e10454 [FREE Full text] [doi: 10.2196/10454] [Medline: 30064969]

47. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 Dec 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]

48. Dockweiler C, Kupitz A, Hornberg C. Acceptance of online-based therapy by patients with light to moderate depressive disorders. Gesundheitswesen 2018 Nov;80(11):1013-1022. [doi: 10.1055/s-0043-119287] [Medline: 29126341]

49. Rojas G, Guajardo V, Martínez P, Castro A, Fritsch R, Moessner M, et al. A remote collaborative care program for patients with depression living in rural areas: open-label trial. J Med Internet Res 2018 Dec 30;20(4):e158 [FREE Full text] [doi: 10.2196/jmir.8803] [Medline: 29712627]

50. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 2012 Jan;6(1):249-262 [FREE Full text] [doi: 10.1016/j.rasd.2011.05.006] [Medline: 22125579]

51. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: a review. Gerontechnology 2009 Apr;8(2):94-103. [doi: 10.4017/gt.2009.08.02.002.00]

52. Beauchamp T, Childress J. Principles of Biomedical Ethics. New York: Oxford University Press; 2012.

53. Stix C. World Economic Forum. 2018 Mar 05. 3 ways AI could help our mental health URL: https://www.weforum.org/agenda/2018/03/3-ways-ai-could-could-be-used-in-mental-health/ [accessed 2019-04-24] [WebCite Cache ID 77roplA75]

54. Pugh A. The New Yorker. 2018 May 22. Automated Health Care Offers Freedom from Shame, But Is It What Patients Need? URL: https://www.newyorker.com/tech/annals-of-technology/automated-health-care-offers-freedom-from-shame-but-is-it-what-patients-need [accessed 2019-04-22] [WebCite Cache ID 77pfV94W6]

55. Costescu CA, Vanderborght B, David DO. Reversal learning task in children with autism spectrum disorder: a robot-based approach. J Autism Dev Disord 2015 Nov;45(11):3715-3725. [doi: 10.1007/s10803-014-2319-z] [Medline: 25479815]

56. Warren Z, Zheng Z, Das S, Young EM, Swanson A, Weitlauf A, et al. Brief report: development of a robotic intervention platform for young children with ASD. J Autism Dev Disord 2015 Dec;45(12):3870-3876 [FREE Full text] [doi: 10.1007/s10803-014-2334-0] [Medline: 25503680]

57. Bickmore TW, Pfeifer LM, Jack BW. Taking the Time to Care: Empowering Low Health Literacy Hospital Patients with Virtual Nurse Agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2009; New York p. 1265-1274.

58. Balon R. Burden of sexual dysfunction. J Sex Marital Ther 2017 Jan 02;43(1):49-55. [doi: 10.1080/0092623X.2015.1113597] [Medline: 26683616]

59. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR Surveill Summ 2018 Dec 27;67(6):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6706a1] [Medline: 29701730]

60. McCabe MP, Sharlip ID, Lewis R, Atalla E, Balon R, Fisher AD, et al. Incidence and prevalence of sexual dysfunction in women and men: a consensus statement from the Fourth International Consultation on Sexual Medicine 2015. J Sex Med 2016 Feb;13(2):144-152. [doi: 10.1016/j.jsxm.2015.12.034] [Medline: 26953829]

61. Messias EL, Chen CY, Eaton WW. Epidemiology of schizophrenia: review of findings and myths. Psychiatr Clin North Am 2007 Sep;30(3):323-338 [FREE Full text] [doi: 10.1016/j.psc.2007.04.007] [Medline: 17720026]

62. Brown AS, Lau FS. Review of the Epidemiology of Schizophrenia. In: Pletnikov MV, Waddington JL, editors. A Handbook of Behavioral Neuroscience. Netherlands: Elsevier; 2016:17-30.

XSL•FO

RenderX

63. Gionet K. The Guardian. 2018 Apr 25. Meet Tess: the mental health chatbot that thinks like a therapist URL: https://www.theguardian.com/society/2018/apr/25/meet-tess-the-mental-health-chatbot-that-thinks-like-a-therapist [accessed 2019-04-22] [WebCite Cache ID 77pgIMS3h]

64. Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J. Experiences of a motivational interview delivered by a robot: qualitative study. J Med Internet Res 2018 Dec 03;20(5):e116 [FREE Full text] [doi: 10.2196/jmir.7737] [Medline: 29724701]

65. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. Cogn Ther Res 2017 Apr 22;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]

66. Deutscher Ethikrat. 2017 Nov 30. Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung URL: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf [accessed 2019-04-24] [WebCite Cache ID 77rpJezPa]

67. Nuffield Council on Bioethics. 2014 Jul 01. The collection, linking and use of data in biomedical research and health care: ethical issues URL: http://nuffieldbioethics.org/wp-content/uploads/Biodata-a-guide-to-the-report-PDF.pdf [accessed 2019-04-24] [WebCite Cache ID 77rpMHI9y]

68. Feil-Seifer D, Mataric M. Socially assistive robotics. IEEE Robot Automat Mag 2011 Mar;18(1):24-31. [doi: 10.1109/MRA.2010.940150]

69. Körtner T. Ethical challenges in the use of social service robots for elderly people. Z Gerontol Geriatr 2016 Jun;49(4):303-307. [doi: 10.1007/s00391-016-1066-5] [Medline: 27220734]

70. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People - an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach (Dordr) 2018;28(4):689-707 [FREE Full text] [doi: 10.1007/s11023-018-9482-5] [Medline: 30930541]

71. Luxton DD. Recommendations for the ethical use and design of artificial intelligent care providers. Artif Intell Med 2014 Sep;62(1):1-10. [doi: 10.1016/j.artmed.2014.06.004] [Medline: 25059820]

72. Oliveira T, Novais P, Neves J. Development and implementation of clinical guidelines: an artificial intelligence perspective. Artif Intell Rev 2013 Mar 13;42(4):999-1027. [doi: 10.1007/s10462-013-9402-2]

73. Fulmer R. The Medical Futurist. 2018 May 24. 5 Reasons Why Artificial Intelligence Won't Replace Physicians URL: https://medicalfuturist.com/5-reasons-artificial-intelligence-wont-replace-physicians [accessed 2019-04-22] [WebCite Cache ID 77pgpQRMQ]

74. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]

75. Ienca M, Jotterand F, Vică C, Elger B. Social and assistive robotics in dementia care: ethical recommendations for research and practice. Int J of Soc Robotics 2016 Jun 22;8(4):565-573. [doi: 10.1007/s12369-016-0366-7]

76. Anderson S. Stanf Encycl Philos Arch. 2017. Coercion URL: https://plato.stanford.edu/entries/coercion/ [accessed 2019-04-24] [WebCite Cache ID 77rpcS2YC]

77. Sharkey A, Sharkey N. Children, the elderly, and interactive robots. IEEE Robot Automat Mag 2011 Mar;18(1):32-38. [doi: 10.1109/MRA.2010.940151]

78. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Dec 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]

79. Heim E, Rötger A, Lorenz N, Maercker A. Working alliance with an avatar: how far can we go with internet interventions? Internet Interv 2018 Mar;11:41-46 [FREE Full text] [doi: 10.1016/j.invent.2018.01.005] [Medline: 30135758]

80. Johnston A. KALW Local Public Radio. 2015 Aug 17. Robotic seals comfort dementia patients but raise ethical concerns URL: https://www.kalw.org/post/robotic-seals-comfort-dementia-patients-raise-ethical-concerns [accessed 2019-04-22] [WebCite Cache ID 77phfBLmf]

81. Tett G. Financial Times. 2018 Feb 09. When algorithms reinforce inequality URL: https://www.ft.com/content/fb583548-0b93-11e8-839d-41ca06376bf2 [accessed 2019-04-22]

82. Corea F. Machine Ethics and Artificial Moral Agents. In: Applied Artificial Intelligence: Where AI Can Be Used In Business. Switzerland: Springer; 2019:33-41.

83. Hammond K. TechCrunch. 2016. 5 unexpected sources of bias in artificial intelligence URL: https://techcrunch.com/2016/12/10/5-unexpected-sources-of-bias-in-artificial-intelligence/?guccounter=1&guce_referrer_us=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_cs=oDTg6AcZIiq57jAEze2jlw [accessed 2019-04-22]

84. Powles J. The New Yorker. 2017 Dec 20. New York City's Bold, Flawed Attempt to Make Algorithms Accountable URL: https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable [accessed 2019-04-22] [WebCite Cache ID 77pieWwCj]

85. Fiske A, Buyx A, Prainsack B. Health information counselors: a new profession for the age of big data. Acad Med 2019 Jan;94(1):37-41 [FREE Full text] [doi: 10.1097/ACM.0000000000002395] [Medline: 30095453]

86. Calfee I. Center for Mobile Communication Studies. 2017 Nov 16. Growing up with Alexa, Siri, and other A. I. technology URL: https://sites.bu.edu/cmcs/2017/11/16/growing-up-with-alexa-siri-and-other-a-i-technology/ [accessed 2019-04-22] [WebCite Cache ID 77piq3K2W]

87. Doucleff M, Aubrey A. National Public Radio. 2017 Oct 30. Alexa, Are You Safe For My Kids? URL: https://www.npr.org/sections/health-shots/2017/10/30/559863326/alexa-are-you-safe-for-my-kids [accessed 2019-04-22] [WebCite Cache ID 77pj2oDUT]

88. Facchin F, Barbara G, Cigoli V. Sex robots: the irreplaceable value of humanity. Br Med J 2017 Dec 15;358:j3790. [doi: 10.1136/bmj.j3790] [Medline: 28811272]

89. Scheutz M, Arnold T. Are we ready for sex robots? 2016 Presented at: 11th ACM/IEEE International Conference on Human-Robot Interaction; 2016; Christchurch, New Zealand p. 351-358.

90. Hall V, Conboy-Hill S, Taylor D. Using virtual reality to provide health care information to people with intellectual disabilities: acceptability, usability, and potential utility. J Med Internet Res 2011 Nov 14;13(4):e91 [FREE Full text] [doi: 10.2196/jmir.1917] [Medline: 22082765]

91. Misselhorn C, Pompe U, Stapleton M. Ethical considerations regarding the use of social robots in the fourth age. GeroPsych 2013 Jan;26(2):121-133. [doi: 10.1024/1662-9647/a000088]

92. Latour B, Woolgar S. Laboratory Life. Princeton: Princeton University Press; 1986.

93. Melson G, Kahn P, Beck A, Friedman B, Roberts T, Garrett E. Robots as dogs?: children's interactions with the robotic dog AIBO and a live australian shepherd. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems. 2005 Apr Presented at: CHI EA '05; April 02-07, 2005; Portland, OR, USA p. 1649-1652. [doi: 10.1145/1056808.1056988]

94. Dodig Crnkovic G, Çürüklü B. Robots: ethical by design. Ethics Inf Technol 2011 Aug 24;14(1):61-71. [doi: 10.1007/s10676-011-9278-2]

Abbreviations

AI: artificial intelligence
ASD: autism spectrum disorder
RCT: randomized controlled trial