Philosophy & Technology (2021) 34:447–461
https://doi.org/10.1007/s13347-020-00395-7

RESEARCH ARTICLE

Is Big Data the New Stethoscope? Perils of Digital Phenotyping to Address Mental Illness

Şerife Tekin [1]

Received: 7 January 2019 / Accepted: 13 February 2020/ Published online: 2 March 2020

Abstract

Advances in applications of artificial intelligence and the use of data analytics technology in biomedicine are creating optimism, as many believe these technologies will fill the need-availability gap by increasing resources for mental health care. One resource considered especially promising is smartphone psychotherapy chatbots, i.e., artificially intelligent bots that offer cognitive behavior therapy to their users with the aim of helping them improve their mental health. While a number of studies have highlighted the positive outcomes of using smartphone psychotherapy chatbots to handle various anxiety related problems no conclusive data illustrate their effectiveness or warrant their use in mental illness diagnosis and treatment settings. Yet smartphone psychotherapy is highly endorsed by experts in the field of mental health research. In this paper, I focus on the specific features of smartphone psychotherapy chatbots intended for the diagnosis and treatment of mental illness and criticize three popular promises; i.e., (i) they enable early diagnosis and intervention through digital phenotyping; (ii) they defy the stigma of mental illness diagnosis and treatment; (iii) they offer increased access to mental health treatment globally. Going against the popular enthusiasm, I argue smartphone psychotherapy chatbots have epistemic and ethical limitations in the diagnosis and treatment of illnesses. In light of these, I encourage researchers, clinicians, policy makers, patients, and caregivers to pause before jumping on the artificial intelligence bandwagon to seek solutions for mental illness on the grounds of these three promises.

Keywords Artificial intelligence · Big data · Digital phenotyping · Psychotherapy · Smartphone psychotherapy chatbot · Mental illness · Treatment

✉ Şerife Tekin
serife.tekin@utsa.edu

[1] Department of Philosophy and Classics, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249, USA

1 Introduction

The gulf between the needs of individuals with mental disorders and the resources dedicated to mental health research and care makes resource allocation an issue of distributive justice. In 2016, 18.3% of all US adults were diagnosed with a mental disorder, and of these, only 43.1% received some kind of treatment, e.g., inpatient or outpatient counseling or prescription medication (Substance Abuse and Mental Health Services Administration 2017). The reasons for not seeking professional help include the lack of available services, inability to recognize symptoms, cost of treatment, time constraints, and concerns about confidentiality and stigma (Gulliver et al. 2010; Kazdin and Rabbitt 2013). The gap between the need for and availability of mental health care services especially affects vulnerable populations with a high risk of developing mental health problems, such as veterans, victims of domestic abuse, those in rural areas, refugees, and immigrants.

Advances in applications of artificial intelligence and the use of data analytics technology in biomedicine are creating optimism, however, as many believe that these technologies will fill the need–availability gap by increasing resources for mental health care. For example, the ubiquitous smartphone is thought to potentially help researchers and therapists explain, predict, and intervene in human psychological phenomena by tracking its owner's mental states and behaviors. One resource considered especially promising is smartphone psychotherapy chatbots, i.e., artificially intelligent bots that offer cognitive behavior therapy to their users with the aim of helping them improve their mental health. Some frequently cited advantages of using smartphone psychotherapy chatbots for mental health problems include their comparatively low cost, wide accessibility through cell phones, and availability in different languages, making them an ideal tool, especially in areas where there is a shortage of therapists who speak the native language of individuals requiring mental health care, such as refugees (Luxton et al. 2011; Whittaker et al. 2012).

Although a number of studies have highlighted the positive outcomes of using smartphone psychotherapy chatbots to handle various anxiety-related problems (e.g., Fitzpatrick et al. 2017), no conclusive data illustrate their effectiveness to warrant their use in mental illness diagnosis and treatment settings. In addition, the focus on various ethical issues concerning the use of smartphone psychotherapy chatbots as possible treatments has been limited. Yet, smartphone psychotherapy is highly endorsed by experts in the field of mental health research. For example, Thomas Insel, a long-time champion of neuroscientific research to advance mental illness treatment and the former director of the National Institute of Mental Health—the agency that provides the largest public funding for mental health research—suggests "the rich, ongoing streams of data that a smartphone can provide" are more promising than research into the brain mechanisms associated with mental illness, because they can be used "to detect a deteriorating state of mind faster and more reliably than humans," and he encourages the dedication of significant resources to developing smartphone psychotherapy technology (Dobbs 2017). Other academics cite the increased use of cell phone technology and participation in digital culture among young people and argue that online psychological support may be more desirable than the typical face-to-face psychotherapeutic methods for these individuals (Kretzschmar et al. 2019). There also appears to be notable interest in psychotherapy chatbots from the public. For example,

Woebot, a psychotherapy chatbot that is currently available via Facebook Messenger has over 16 k likes, and its standalone mobile application has around 50 k downloads to date. Such public interest seems to have turned into a significant user base as well; according to the company's website, Woebot has more than 2 million conversations per week, across more than 120 countries (Woebot website. https://woebot.io. Accessed September 18, 2019). In the face of the increased interest coming from both academics and the public at large, a plethora of connected questions emerge: Are smartphone psychotherapy chatbots effective tools for mental disorder diagnosis and treatment? If they are in fact effective, which ethical standards should guide their development and use? Must research on the artificial intelligence-assisted behavioral intervention technology be prioritized in lieu of improving on other diagnostic and treatment strategies, such as in person psychotherapy? If carried out, should this technology be funded by public or private funding resources?

I engage with *some* of these questions by assessing three popular promises of smartphone psychotherapy chatbots in diagnosing and treating mental disorders, namely, that (a) they will enable early diagnosis and intervention through digital phenotyping; (b) they will enhance treatment by defying the stigma associated with mental disorders; and (c) they will offer increased access to mental health treatment globally. I argue that these are not genuine promises that warrant the development and use of smartphone psychotherapy chatbots. I argue that the digital phenotyping technology that is promoted in the first promise contain significant epistemic and ethical constraints. Smartphone psychotherapy chatbots should not be recommended as possible treatments for mental disorders before addressing these constraints. My criticisms of the second and third promises are more structural. I suggest that this technology is offered as a band-aid to the deeper problems in social and political environments that make the use of these technologies seems urgent and appealing. Even if the epistemic and ethical concerns I raise in the first promise are addressed, the criticisms I raise about the second and third promises remain relevant. Thus, going against the popular enthusiasm, I argue that smartphone psychotherapy chatbots have epistemic and ethical limitations in the diagnosis and treatment of illnesses. In light of these, I encourage researchers, clinicians, policy makers, patients, and caregivers to pause before jumping on the artificial intelligence bandwagon to seek solutions for mental illness on the grounds of these three promises. I conclude with two recommendations. First, at this stage, research funding should be allotted cautiously to develop and test the efficacy of this technology. Second, in this development phase, chatbots should not be used in lieu of existing person-level interventions. It is my hope that this article will stimulate philosophical and ethical debate among app developers, researchers, practitioners, patients, and their caregivers to inspire the development of ethically responsible research and practice in digital mental health.

2 Promise One: Early Diagnosis and Intervention Through Digital Phenotyping

A popular promise of smartphone psychotherapy chatbot technology is epistemic in nature; its proponents highlight the potential ability of artificial intelligence technology to detect the deterioration of an individual's mental states faster and more reliably than

clinicians currently can. This relies on the premise that people's patterns of smartphone use are indicative of their mental health. Tracking their smartphone use is considered to have advantages over existing forms of treatments where, by the time patients seek treatment, the symptoms and signs of their condition may have significantly progressed. Instead of detecting and then treating mental illness, the idea is to track mental states before anomalies become full-blown mental illness. In other words, the goal is to preempt mental illness and intervene right away.

The enabler of early detection is considered to be digital phenotyping, i.e., moment to moment quantification of the individual-level human phenotype in situ by using data from personal digital devices (Onnela and Rauch 2016; Jain et al. 2015). The data acquired through digital phenotyping are divided into the following two subgroups: active and passive. Active data require active input from users to be generated, whereas passive data, such as sensor data and phone usage patterns, are collected without requiring active user participation.

The main idea behind how digital phenotyping can detect or predict the onset of mental illness and quickly disseminate effective, affordable care to those who need it is the use of smartphones to track an individual's daily behavior, as this is presumed to be revelatory of mental health. Through digital phenotyping, a range of active and passive data can be acquired from the smartphone. Smartphones track active data, for example, how often individuals walk, how much they sleep, how long they talk on the phone. They also track passive data in the form of human–computer interactions, such as taps, scrolls, and clicks. If individuals start developing depression, for example, they may talk with fewer people, and when they talk, they may speak more slowly, say less, and use shorter sentences and a smaller vocabulary. They may return fewer calls, texts, emails, Twitter direct messages, and Facebook messages (Dobbs 2017). They may answer the phone more slowly, if they pick up at all, and they may spend more time at home and go fewer places (Torous et al. 2016). They may sleep differently. Someone slipping toward a psychotic state might show similar signs, as well as particular changes in syntax, speech rhythm, and movement. All these can be sensed by a phone's microphones, accelerometers, GPS units, and keyboards.

Although there is *ongoing* research to understand how these new sources of data can be turned into valuable clinical information, we do *not* have any conclusive evidence. For example, small-scale clinical trials of the use of the Beiwe app among patients with schizophrenia are under way (e.g., Torous et al. 2016). Motor disorders, such as decreased movement, are known markers for schizophrenia and considered essential to understanding the progression of the disease. The hypothesis is that digital markers like the ones listed previously could help recognize those patients who experience decreased movement indicating a risk of relapse and intervene before their symptoms worsen. Specifically, the Beiwe app uses GPS and accelerometer data to recognize the way that a person walks or holds their phone and identify any abnormal movements. Moreover, data regarding someone's call history and text messaging activity serve as a proxy measure for social engagement, which tends to be less frequent as schizophrenia progresses. The Beiwe app collects data, such as the time a call was placed, its length, or the number of characters on a text message. Whether the apps like Beiwe will be successful in developing effective therapeutic interventions remains to be seen.

Another argument in support of the diagnostic power of digital phenotyping has been made by Thomas Insel, former NIMH director. In his last years as director, he has

been upfront about how both psychiatry and NIMH were "failing to help the mentally ill" (Dobbs 2017). He stated his discomfort with the "pharmaceutical industry's failure to develop effective new drugs for depression, bipolar disorder, or schizophrenia," academic psychiatry's close relationship with Big Pharma, and "the paucity of treatments produced by the billions of dollars the NIMH had spent during his tenure" (Dobbs 2017). He compared medical advances with the advances in psychiatry, noting that "in the previous half century, [medical advances] had reduced mortality rates from childhood leukemia, heart disease, and AIDS by 50 percent or more, whereas psychiatry failed to reduce suicide or disability from depression or schizophrenia." In an interview, he recalled a conversation he had with the family member of a patient after one of his talks in which he was listing the significant neuroscientific discoveries of NIMH. This person told Insel: "Our house is on fire … and you're telling us about the chemistry of the paint. We need someone to focus on the fire" (Dobbs 2017). Insel pointed to the truth in the man's words: "It's not just that we don't know enough. The gap between what we know and what we do is unacceptable" (Dobbs 2017). Insel left NIMH and started working for Google's Verily to develop AI in a way that would meet the needs of the mentally ill. Shortly thereafter, he left Verily and started a company to develop smart psychotherapy chatbots, expressing his commitment to use AI technology to address the needs of those with mental disorders. In a recent article, Insel wrote that in 2050, "when psychiatrists look back at the first two decades of the 21st century," they will recognize the impact of the "revolution in genomics, which has given us new insights into the risk architecture of mental illness, and the revolution in neuroscience, which has given us a new view of mental illnesses as circuit disorders" (Insel 2018). But, perhaps more importantly, he continues, "the revolution in technology and information science will prove more consequential for global mental health" because almost everyone will have a cell phone, and, with these, those with mental health challenges can be helped. Insel continues, in an interview:

> Putting sensor data, speech and voice data, and human–computer interaction together might provide a digital phenotype that could do for psychiatry what HgbA1c or serum cholesterol has done for other areas of medicine, giving precision to diagnosis and accuracy to outcomes. (Dobbs 2017)

Insel is moving faster than the evidence available for the efficacy of digital phenotyping. He was once this optimistic about the power of neuroscience to fathom the etiology of mental disorders and successfully intervene, and his optimism resulted in significant changes in research funding to neuroscience-based projects (Insel 2013).[1] Yet, neuroscientific advances have not met his expectations in addressing the needs of the patients, and he himself admits this. So, he should be more cautious when he says artificial intelligence and big data technology are the keys to unlocking the mysterious door to treatment for mental disorders.

[1] In 2010, NIMH launched the Research Domain Criteria (RDoC) initiative aimed at developing, for research purposes, new ways of classifying mental disorders based on behavioral dimensions and neurobiological measures. The goal of RDoC is to create a new conceptual framework for psychiatric research that identifies domains of functioning that can be analyzed at several levels, thereby integrating resources from various basic sciences, especially neuroscience, and cognitive science.

Let me now turn to the epistemic and ethical concerns about the promise that smartphone psychotherapy chatbots, through AI and big data technology, will allow early detection, diagnosis, and intervention in mental disorders. I am skeptical that digital phenotyping will achieve all these outcomes for a number of reasons. To focus my criticism, I turn to a smartphone psychotherapy chatbot, MyCompass, which delivers cognitive behavioral therapy (CBT) to its users (Proudfoot et al. 2013). CBT is an evidence-based and widely used therapeutic approach developed by Aaron Beck (Beck 1975). It is a problem-focused, time-limited, and evidence-based approach that rests on the assumption that individuals feel bad not only because of events but also because of how they think about those events. CBT techniques help them understand that their perceptions of events can sometimes be exaggerated or false and aim to enable them to reframe their interpretation. For example, "I'm never going to make any friends" is all-or-nothing thinking, and people saying this usually believe it. Removing the (all-or-nothing) distortion leads to a more balanced thought. Rephrasing the thought as "I haven't made any friends yet" or "I'm sure I'll make one or two friends eventually" takes the sting out and helps people cope with the reality of the situation in a more productive way. For CBT to be effective, individuals need to repeatedly record their thoughts and challenge them, again and again, before the new thought becomes natural. This is especially difficult at the moment when they would benefit most from doing it—i.e., when they are experiencing strong emotions.

MyCompass builds on CBT. It is a self-guided psychological treatment delivered via mobile phone and computer, designed to reduce mild-to-moderate depression, anxiety, and stress and to improve work and social functioning. It encourages real-time self-monitoring of moods, mood triggers, and lifestyle behaviors using SMS text messaging and email prompts. MyCompass is thought to enable the collection of a more objective picture of an individual's life than data collected, say, in weekly sessions with a therapist. This is because, in the latter, the therapist only finds out about the patient's life based on the patient's subjective reports at the end of the week as opposed to a detailed report gathered on a daily basis made possible by the former. In addition, the acquired data are considered to be more textured or fine grained because it will record the details of patient's day-to-day activities as they happen, instead of a broad picture of the week summarized in one therapy session. Its proponents suggest that this objective and textured data could thus sense "the beginning of a crisis and trigger an appropriate response. Because this response would come earlier, it could be more measured, less jarring, and less medication-heavy … The earlier you intervene, the better the outcomes" (Dobbs 2017).

The first epistemic problem here is the assumption that individuals who use the MyCompass psychotherapy chatbot will track and report their moods, trigger, and lifestyle behavior and that these will reflect the actual/true state of affairs that the individual is experiencing. There is plethora of reasons why this assumption may not hold. First, not everyone is equally self-reflective; individuals may not be aware of their moods, the changes in those moods, or how various triggers may affect their moods and behavior. In fact, one advantage of in-person CBT is the psychotherapist's ability to challenge patients and encourage them to notice the connection between their moods and their behavior. Because psychotherapy chatbots like MyCompass are self-directed and the user is in charge of tracking and reporting, they may be limited in fully observing and tracking mental and behavioral phenomena.

The second epistemic issue with digital phenotyping is the risk of false positives: misdiagnosing anomalies in behavior as a sign of mental distress. For example, the reason that individuals respond less to the SMS and email prompts from the MyCompass app requesting them to record their moods, etc., may not be because they are depressed or have other mental illness experiences; rather, the information requests from these apps may simply be a notification that they choose to ignore.

The third epistemic issue pertains to the specific experiences of mental disorders. For example, some individuals with mental disorders suffer from anosognosia, which leads them to deny that they have a mental health problem (Amador and David 2004; Tekin 2016). If they do not think they have a problem, they will be less likely to monitor their moods and behaviors. For example, the Beiwe app developed to track the mental state of individuals with schizophrenia will not work with patients with schizophrenia.

Fourth and finally, there is increased awareness worldwide of the various concerns about the use of private data by businesses; this may lead users of MyCompass to self-censor and not report everything about their mental states and behavior. As I discuss subsequently, public awareness of various data businesses' (such as Facebook) manipulation and selling of private data may lead the potential users of these chatbots to lose trust in their potential effectiveness, thus hampering their ability to benefit.

There are also significant ethical problems involved in using digital phenotyping technology for early detection and diagnosis. The first big concern is the ethical challenge of data privacy. Smartphone psychotherapy chatbots collect a great amount of demographic and medical information by urging users to enter a lot of personally identifiable data, for example, name, phone number, email address, age, gender, and even photos. They frequently catalog lifestyle information, such as food consumption and exercise habits, or information related to diagnoses and treatments (e.g., chronic health/mental health problems, screening results, medication dosages). Moreover, when using the app, people usually create a record of their daily routines and practices (e.g., diet, exercise, moods). Even if there is a privacy policy issued by the developer, there are usually no regulations[2] to protect the privacy and security of personal health information. Second, there is a strong possibility that smartphone psychotherapy chatbots will lack reliable security; they might transmit unencrypted personal data over insecure network connections or allow ad networks to track users, raising serious concerns about their ability to protect the confidentiality of user information (Harris 2013; Njie 2013). Personal health information is of great value to cybercriminals and can be used to obtain medical services and devices or bill insurance companies for phantom services in the victim's name. As there are few legal protections, victims are forced to pay or risk losing their insurance and/or ruining their credit ratings (Dolan 2013). Fraudulent healthcare events can leave inaccurate data in medical records about tests, diagnoses, and procedures that could greatly affect future healthcare and insurance coverage

[2] Note that researchers working on developing and using this technology in the US must abide by the statues of the Health Insurance Portability and Accountability Act (HIPAA) legislation, which requires data privacy and security provisions for safeguarding medical information. However, if the researchers are not in the US, they may be exempt from such requirements. At this point, there are no universal guidelines.

(Dolan 2013). Erroneous mental health information could even influence a person's social life or work opportunities (Hoffman and Zachar 2017; Tekin 2014).[3]

My third and perhaps most important ethical concern is advertising smartphone psychotherapy chatbots as possible diagnosis and treatment tools for mental disorders despite the lack of research evidence on their potential efficacy. The number of tested evidence-based mental health apps in general is small, and studies usually rely on small, non-controlled, and non-randomized samples (Tomlinson et al. 2013; Buijink et al. 2012). Only a few report sustainable results for a period of more than three months, try to replicate these results, or test the effects of mobile interventions on everyday life, work, and social functions in general (Fiordelli et al. 2013; Donker et al. 2013). The data are even slimmer for CBT delivered by smartphone psychotherapy chatbots. Although CBT interventions are successful in a number of mental health problems in face-to-face therapy, evidence of the impact of CBT-based smartphone psychotherapy chatbots is limited (Hofmann et al. 2012; Aguilera and Muench 2012; Kretzschmar et al. 2019). The technology is advancing so fast that research seems unable to keep up.[4]

Given these concerns, it seems wise to rein in the enthusiasm about the promise of digital phenotyping. That said, most of the epistemic and ethical challenges associated with the digital phenotyping technology I raised here are empirical in nature. Thus, it is plausible to address them in future research. A responsible way of allocating research funding do further develop these technologies in order to address these challenges might be to follow these three steps: (a) gather evidence that digital phenotyping is in fact successful at tracking mental health; (b) use this evidence to further develop technologies that can potentially help treat mental illness; and (c) collect evidence on whether such technologies are effective in treating mental illness with a serious consideration of their ethical constraints. Even though these obstacles are overcome however, there are reasons to remain skeptical of the chatbot psychotherapy technology, the reasons for which can be found subsequently in my criticisms of the second and third promises.

3 Promise Two: Defying Stigma in Treatment

Stigma associated with mental illness is a major barrier to seeking treatment (Corrigan et al. 2014). Two kinds of stigma affect the decision to seek treatment (Corrigan and Watson 2004). The first is perceived public stigma; individuals with mental disorders do not seek treatment or they stop treatment prematurely because they want to avoid the scrutiny of others. The second is internalized stigma; individuals avoid seeking help

[3] More could be said here about the dangers of infosecurity: With increased issues of data breaches, we must be very concerned about these chatbot companies using and selling the data of their users. I merely scratched the surface of these issues here, for reasons of space. I hope that further ethical evaluations of the issue of privacy are raised by other philosophers and ethicists as these technologies become more widespread.

[4] In addition, there are further important questions about evidence. For example, some of these apps are marketed directly to consumers, which do not seem to be as vigilant in relaying the existing evidence for their effectiveness or limitations (for example, Woebot, an app that is discussed in Section 3). Whereas others, especially the ones that are developed by clinicians, seek endorsement by therapists for their patients and are thereby more forthcoming about their limitations (for example, MyCompass).

because they want to avoid personal feelings of shame and guilt. These two constructs manifest differently in individuals, but they tend to influence each other. For example, those who perceive public stigma as high are more likely to internalize negative stereotypes than those who perceive public stigma as low. One promise of smartphone psychotherapy chatbot technology is that it will enable people to seek treatment without fear of public stigma, which might, in turn, will lower the internalized stigma. Because the psychotherapy chatbot is available through cellphones, and the therapist is not a person but a chatbot, people can keep the fact of seeking treatment private and get the help they need without fear of judgment (Dobbs 2017; Kretzschmar et al. 2019).

Consider one smartphone psychotherapy chatbot, Woebot. Woebot was created by psychologists and AI experts who decided to leave the clinic to address the mental health needs of those with no access to basic health care. They built an AI bot to provide CBT to its users. Like MyCompass, Woebot uses brief daily chat conversations, a mood-tracking facility, curated videos, and word games to help people manage their mental health. The goal is for people to talk to Woebot when they are feeling badly. The ostensible advantage of using Woebot is its ability to guide people in challenging their thoughts. Woebot does not develop solutions to individual problems, but it asks questions, so users can find answers on their own. Woebot's prompts are modeled on CBT; it asks people to recast their negative thoughts in a more objective light, encouraging them to talk about their emotional responses to life events, and then to stop and identify the psychological traps causing their stress, anxiety, and depression. Its creators argue that Woebot is not only more affordable than seeing an actual therapist every week (or more frequently); it is also more effective because the person using it does not feel stigmatized. Alison Darcy, one of the psychologists who developed Woebot, said in an interview, "[T]here's a lot of noise in human relationships … Noise is the fear of being judged. That's what stigma really is" (Dobbs 2017). For Darcy, when users are talking to an anonymous algorithm, they will not fear judgment.

A similar argument comes from a recent article by Kretzschmar et al. that evaluates the attitudes of young people to chatbot psychotherapy (Kretzschmar et al. 2019). The authors argue that young people are reluctant to seek mental health treatment due to stigma. They cite research that shows that some young people express their preference for self-reliance when coping with emotional distress or obtaining support from people they feel close to, such as family members or friends, as opposed to receiving professional support (Rickwood and Braithwaite 1994; Rickwood et al. 2005). Kretzschmar et al. argue as follows:

> … as technology and digital culture become increasingly more present in young people's lives, young people may … prefer to look for support online rather than face-to-face. An online, mobile-based intervention is less likely to carry the stigma attached to formal mental health services and provides a self-reliant intervention platform for those who would otherwise be reluctant to seek support. (Kretzschmar et al. 2019, 2)

I have a number of concerns about the optimism that smartphone psychotherapy chatbots will encourage people to seek help by minimizing stigma. First, instead of advocating for the development of strategies to reduce the stigma of mental health, proponents of smartphone psychotherapy chatbots offer ways to sidestep it. This is

dangerous because it legitimizes and perpetuates the idea that mental disorders are phenomena that warrant stigma, and the most effective way to improve help-seeking behavior is to keep it secret, instead of taking a stance against the stigmatization of mental disorders. As research has well established, the effects of stigma are best moderated by increasing and disseminating accurate knowledge of mental illness, increasing mental health literacy, and creating family engagement programs where individuals are educated on various aspects of mental disorders (Stuart 2016). These, not chatbot technology, can counteract the effects of public, self, and structural stigma. I worry that the arguments used by proponents of smartphone psychotherapy chatbots may undermine policies designed to reduce stigma and promote mental health care.

Arguably, smartphone psychotherapy chatbots offer an immediate solution to the problem of not seeking help due to stigma. By the time we get rid of mental illness stigma through education, it might be too late for many people who have not sought help due to stigma. This can be overlooked if psychotherapy chatbots are genuinely helpful, but unfortunately there are reasons to be skeptical about how effective they can be. Aside from the observations I listed in the previous section about the limitations of smartphone technology in tracking mental states and behavior, the inconclusiveness of research about its efficacy, and the various ethical problems concerning privacy, a fundamental component of recovery or improvement facilitated by therapy is the therapeutic alliance between patient and therapist, which can be defined as the process in which they work together to determine the goals of treatment based on the patient's existing problems, identify the steps to achieve that goal, and form a bond in the process. Research suggests the therapeutic alliance is a strong predictor of successful outcomes (Ardito and Rabellino 2011; Capaldi et al. 2016). Building a therapeutic alliance is a relational process, in which the therapist gives uptake to the patient's concerns, and the patient feels recognized and cared for. I am skeptical that this type of alliance can be formed between a person and a bot.

Another important component of successful psychotherapies, or other healthcare treatments in medicine for that matter, is the patient's trust in healthcare professionals and the healthcare system at large (Collier 2012). Research indicates that both medical professionals and patients perceive trust to be the fundamental ingredient of a successful treatment program. Some even say, "Without trust, physician–patient interactions could become more like consumer transactions at a shopping mall" (Collier 2012). In light of this, it is hard to imagine a patient building a trusting relationship with a chatbot. And if they are aware of the various unethical ways private data are sold by businesses, such as Facebook, potential users of psychotherapy chatbots may be even less likely to trust this technology.

Finally, the nature of the relationship between the healthcare professional and the patient, not only in psychiatry but in all areas of medicine, is an important topic for proponents of the humanistic approaches to medicine. Rita Charon, the founder of narrative medicine, argues that the clinician must acquire the skills to listen, interpret, and reflect on the patient's stories with an "engaged concern" to achieve therapeutic outcomes because this is the fundamental way in which the patient learns to trust the clinician (Charon 2006). Giving uptake is necessary to build trust between clinicians and patients. In the field of mental health, this is crucial. I am skeptical that a bot can ever offer patients the crucial therapeutic experience of feeling that someone else, despite knowing their flaws and vulnerabilities, cares about them. Perhaps people will

seek help from a bot to avoid stigma, but I doubt such help will bring them the results they need or desire.

4 Promise 3: Increased Access to Mental Health Treatment

The third promise of smartphone psychotherapy chatbots is increased access to mental health treatment globally, especially for vulnerable populations, including refugees and veterans. Insel, for example, points out that one in seven of the world's 7.5 billion people is struggling with mental illness. We cannot "reach all those people by hiring more psychiatrists," Insel says, but we can reach them with smartphones (Dobbs 2017). By 2020, it is expected that six billion people will use smartphones with the capability of capturing mental health data and apps able to provide a form of treatment. In addition, smartphone psychotherapy chatbots can be made available in multiple languages, increasing their reach to vulnerable populations. To give only one example, according to a recent World Health Organization Report, more than one million Syrians have fled to Lebanon since the start of the conflict in Syria, and as many as one-fifth of these refugees may be suffering from mental disorders after losing loved ones, livelihood, and community. However, their mental health needs are unmet because Lebanon's mental health services are mostly private and thus are not available to refugees. The proponents of psychotherapy chatbot technology argue that while it may not be feasible to send clinicians who speak Arabic to help these refugees, an Arabic-speaking bot can be used to deliver psychological support.

For example, Karim, an AI bot created by the Silicon Valley startup X2AI, is enabled to engage in personalized text message conversations in Arabic to help people with their emotional problems. Like Woebot, the system uses natural language processing to analyze an individual's emotional state and returns appropriate comments, questions, and recommendations. Karim gets smarter as it interacts with the user. To disseminate Karim among refugees in Lebanon, X2AI teamed up with Field Innovation Team (FIT), a nongovernmental organization delivering tech-enabled disaster relief. Desi Matel-Anderson from FIT comments, "Psychosocial services create a bedrock in order to create learning outcomes and do something that helps. Exponential technology like X2AI's will let us reach people we wouldn't normally get to help" (Solon 2016). For now, Karim is being used cautiously, positioned as a friend rather than a therapist, and it is unclear whether and how much support Karim has given to refugees in Lebanon. Results are anecdotal. For example, a Syrian refugee who fled his home in Damascus to live in Lebanon and who now teaches at a school for refugee children was given the opportunity to trial Karim. He said he felt like he was talking to real person. A lot of Syrian refugees have trauma, and he thought this might help them overcome it (Solon 2016). He added that given the stigma of psychotherapy, people might feel more comfortable talking to a "robot" than to a human.

In addition to my overarching skepticism about the epistemic and ethical constraints of bots in addressing mental health challenges, I worry that motivating the development of this technology to address the growing needs of refugee populations medicalizes social and political problems. It does not encourage masking these problems instead of proposing solutions. Some might say that the medicalization problem will still be there even if opt for other medical intervention methods. My response is that while the

medicalization of social and political problems is never desirable, it is intrinsically more morally problematic to address social and political problems by using robots instead of persons. In other medical interventions, the methods of intervention are arguably more moral because they are humanistic in nature: doctors, nurses, therapists, and other medical professionals—all actual humans—are physically there to bear witness to individuals' suffering, listen to their stories, and offer help. Healthcare delivery is strong when healthcare professionals express "empathy, reflection, professionalism, and trustworthiness" when interacting with their patients (Charon 2006). These necessary humanistic ingredients for treatment will be missing in the context of psychotherapy chatbot intervention. Thus, if the concern for the mental health of the refugees is genuine, which seems to be one of the main motivators of chatbot psychotherapy technology development, the suggestion to replace human healthcare professionals with chatbots does not reflect it.

Another limitation of psychotherapy chatbots, such as X2AI's Karim, is the assumption of Western standards in mental health and treatment of mental health challenges. First of all, these chatbots focus on the way mental disorders manifest in the mostly white and privileged communities in the West yet impose these criteria on people in Middle Eastern communities. For example, there are many cultural differences in the way individuals may choose to share their struggles. Although people in Western societies are encouraged to verbalize, express, and share problems, the norms may be different in the Middle Eastern Cultures, and it may be harder to get people open up even if they wanted to, especially if they are not used to sharing their feelings. In these contexts, actual therapists may be helpful as they might be able to earn the refugee's trust more effectively than a chatbot and thereby help the persons in distress. Second, there are significant differences in the way mental health challenges are experienced and communicated by different genders, races, and ethnic groups (Bluhm 2011). For example, as Meri Nana-Ama Danquah writes in her memoir, *Willow Weep for Me: A Black Women's Journey Through Depression*, there is an expectation in some Black communities that women should be "strong," accept suffering as a typical aspect of living as a woman, recognize that "emotional hardship is *supposed* to be built into the structure" of their lives, and never ask for help (Danquah 1998). Unfortunately, the currently designed smartphone psychotherapy chatbots are rather standardized and are not sensitive to cultural variations in mental distress experience.

5 Conclusion

This paper evaluated the three popular promises of smartphone psychotherapy chatbots that are said to facilitate the diagnosis and treatment of mental illness. I showed that these promises contain various epistemic and ethical weaknesses, cautioning academics, mental health professionals, policy makers, and patients against jumping on the Artificial Intelligence bandwagon uncritically. The first promise is that (a) smartphone psychotherapy chatbots will enable early diagnosis and intervention through digital phenotyping. The epistemic concerns I raised here include the possibility of false negatives in mental disorder diagnosis, the (false) assumption that people's cell phone use is indicative of their mental health, and lack of empirical evidence that shows that digital phenotyping can indeed measure mental health status. In addition, I

laid out multiple ethical concerns, including, data privacy, and using technology that is not shown to be efficacious as a potential treatment for mental disorder. My criticisms of the first promise are empirical in nature: If they are addressed, it *may* be justifiable to use smartphone psychotherapy chatbots as potential methods of treatment. However, my broader criticism of promoting psychotherapy chatbots as potential treatment agents in challenging the second and third promises give us more reasons to remain skeptical about this new technology. These are that smartphone psychotherapy chatbots will (b) defy the stigma of mental illness diagnosis and treatment and (c) offer increased access to mental health treatment globally. Unfortunately, the second promise calls for strategies to *sidestep* the stigma associated with mental disorders as opposed to reduce it. Thus, it further perpetuates the idea that stigma about mental health problems is warranted by suggesting that the most effective way to improve help-seeking behavior is to keep it secret. The third promise is also problematic because it medicalizes social and political problems associated with flight from war zones and experiences of living as refugees and immigrants in non-native countries. In addition, these chatbots and the method of psychotherapy they endorse impose primarily white and Western standards to mental distress experience and treatment and lack the sensitivity to different forms of experiences.

Moving forward, I recommend that research funding should be allotted cautiously to develop and test the efficacy of this technology at this stage, and during this developmental phase, chatbots should not be used in lieu of existing person-level interventions. I hope that this article stimulates philosophical and ethical debates among app developers, researchers, practitioners, patients, and their caregivers to inspire the development of ethically responsible research and practices in digital mental health.

Acknowledgments I would like to acknowledge the three anonymous reviewers for providing tremendously helpful feedback on the earlier drafts of this manuscript. I am also grateful for the questions raised by the participants of the Mellon Foundation Sawyer Seminar on "Human Plasticity and Human–Machine Interface" at Boston University where I presented my early thoughts on the topic.

References

Aguilera, A., & Muench, F. (2012). There's an app for that: information technology applications for cognitive behavioral practitioners. *The Behavior Therapist, 35*, 65–73.

Amador, X. F., & David, A.S. (2004). *Insight and psychosis*. New York: Oxford University Press.

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. *Frontiers in Psychology, 2*.

Beck, A. T. (1975). *Cognitive therapy and the emotional disorders*. Madison: International Universities Press, Inc..

Bluhm, R. (2011). Gender differences in depression: explanations from feminist ethics. *International Journal of Feminist Approaches to Bioethics, 4*(1), 69.

Buijink, A. W. G., Visser, B. J., & Marshall, L. (2012). Medical apps for smartphones: lack of evidence undermines quality and safety. *Evidence-Based Medicine, 18*, 90–92.

Capaldi, S., Asnaani, A., Zandberg, L. J., Carpenter, J. K., & Foa, E. B. (2016). Therapeutic alliance during prolonged exposure versus client-centered therapy for adolescent posttraumatic stress disorder. *Journal of Clinical Psychology, 72*(10), 1026–1036.

Charon, R. (2006). *Narrative medicine: honoring the stories of illness*. New York: Oxford University Press.

Collier, R. (2012). Professionalism: the importance of trust. *Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne, 184*(13), 1455–1456.

Corrigan, P. W., & Watson, A. C. (2004). At issue: stop the stigma: call mental illness a brain disease. *Schizophrenia Bulletin, 30*, 477–479.

Corrigan, P. W., Druss, B. G., & Perlick, D. A. (2014). The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest, 15*(2), 37–70.

Danquah, M. (1998). *Willow weep for me: a black woman's journey through depression.* New York: WW. Norton&Co.

Dobbs, D. (2017). The smartphone psychiatrist. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2017/07/the-smartphone-psychiatrist/528726/. Accessed 20 Feb 2019.

Dolan, P.L. (2013) Health data breaches usually aren't accidents anymore. http://www.amednews.com/article/20130729/business/130729953/4/. Accessed 20 Feb 2019.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of Medical Internet Research, 15*, e247.

Fiordelli, M., Diviani, N., & Schulz, P. J. (2013). Mapping mHealth research: a decade of evolution. *Journal of Medical Internet Research, 15*, e95.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health, 4*(2), e19.

Gulliver, A., Griffiths, K., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help seeking in young people: a systematic review. *BMC Psychiatry, 10*, 113.

Harris, K.D. (2013). Privacy on the go: recommendations for the mobile ecosystem. *California Department of Justice*. http://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/privacy_on_the_go.pdf. Accessed 20 Feb 2019.

Hoffman, G., & Zachar, P. (2017). RDoC's metaphysical assumptions: problems and promises. In J. In Poland & Ş. Tekin (Eds.), *Extraordinary science and psychiatry: responses to the crisis in mental health research* (pp. 59–86). Cambridge: MIT Press.

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognitive Therapy and Research, 36*, 427–440.

Insel, T. (2013) *Director's blog: transforming diagnosis*. https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml. Accessed 20 Feb 2019.

Insel, T. (2018). Digital phenotyping: a global tool for psychiatry. *World Psychiatry, 17*(3), 275–277.

Jain, S. H., Powers, B. W., Hawkins, J. B., et al. (2015). The digital phenotype. *Nature Biotechnology, 33*, 462–463.

Kazdin, A. E., & Rabbitt, S. M. (2013). Novel models for delivering mental health services and reducing the burdens of mental illness. *Clinical Psychological Science, 1*, 170–191.

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*.

Luxton, D. D., McCann, R. A., Bush, N. E., Mishkind, M. C., & Reger, G. M. (2011). mHealth for mental health: integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice, 42*, 505–512.

Njie, C.M.L. (2013) Technical analysis of the data practices and privacy risks of 43 popular mobile health and fitness applications. *Privacy Rights Clearinghouse*. https://www.privacyrights.org/mobile-medical-apps-privacy-technologist-research-report.pdf. Accessed 20 Feb 2019.

Onnela, J., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41*(7), 1691–1696.

Proudfoot, J., Clarke, J., Birch, M. R., Whitton, A. E., Parker, G., Manicavasagar, V., et al. (2013). Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. *BMC Psychiatry, 13*, 312.

Rickwood, D. J., & Braithwaite, V. A. (1994). Social-psychological factors affecting help-seeking for emotional problems. *Social Science & Medicine, 39*, 563–572. https://doi.org/10.1016/0277-9536(94)90099-X.

Rickwood, D., Deane, F. P., Wilson, C. J., & Ciarrochi, J. (2005). Young people's help-seeking for mental health problems. *Aust e-Journal Adv Ment Heal., 4*, 218–251. https://doi.org/10.5172/jamh.4.3.218.

Solon, O. (2016). Karim the AI delivers psychological support to Syrian refugees. *The Guardian*. and Mental Health Services Administrationhttps://www.theguardian.com/technology/2016/mar/22/karim-the-ai-delivers-psychological-support-to-syrian-refugees. Accessed 20 Feb 2019.

Stuart, H. (2016). Reducing the stigma of mental illness. *Global Mental Health, 3*, e17.

Substance Abuse and Mental Health Services Administration. (2017). *Key substance use and mental health indicators in the United States: results from the 2016 national survey on drug use and health*. HHS publication no. SMA 17-5044, NSDUH series H-52. https://www.samhsa.gov/data/. Accessed 20 Feb 2019.

Tekin, Ş. (2014). Self-insight in the time of mood disorders: after the diagnosis, beyond the treatment. *Philosophy, Psychiatry, and Psychology, 21*(2), 139–155.

Tekin, Ş. (2016). Are mental disorders natural kinds? A plea for a new approach to intervention in psychiatry. *Philosophy, Psychiatry, and Psychology, 23*(2), 147–163.

Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling up mHealth: where is the evidence? *PLoS Medicine, 10*, e1001382.

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J. P. (2016). New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health, 3*(2), e16. https://doi.org/10.2196/mental.5165.

Whittaker, R. A., Merry, S., Stasiak, K., McDowell, H., Doherty, I., Shepherd, M., Dorey, E., Ameratunga, S., & Rodgers, A. (2012). Universal depression prevention via mobile phones. *Journal of Mobile Technology in Medicine, 1*, 4S.

Woebot website. https://woebot.io. Accessed 18 Sept 2019.