

## Pooling Mental Health Data with Chatbots

*Michael Mattioli*<sup>1</sup>

### 3.1 INTRODUCTION

For an invisible condition, clinical anxiety casts a long shadow. In the United States, about 40 million adults – a full 18.1 percent of the population – suffer from an anxiety disorder, making it the most common type of mental health problem in the country [1]. Anxiety is insidious; it can crop up like a weed at any point in life, coloring one’s days and nights with unaccountable feelings of dread. Anxiety is expensive. A recent study estimated that in the United States, the total costs of treatment, lost productivity, and lost wages are about \$42 billion every year (Du Pont et al., 1996) [2]. This amount exceeds the most recent annual revenues of Facebook, Coca-Cola, Nike, and Morgan Stanley. Anxiety is thriving. There is a shortage of mental health workers (clinicians) in the United States, and experts believe that this problem of supply and demand will worsen substantially by 2025. Even when help is available, high prices and perceptions of social stigma prevent many people from seeking help [1].

This chapter explores a new type of technology that relies in part upon large sets of patient data to relieve anxiety symptoms (Dobbs, 2017). The technology is software that can simulate certain kinds of conversations with human therapists. Referred to as “conversational agents” or “therapy chatbots,” the idea might sound like science fiction, but it is a commercial reality. Recent studies indicate that conversational agents are useful in helping to reduce the symptoms of anxiety and depression. A recent peer-reviewed study showed that the program highlighted in this chapter, “Woebot,” can significantly reduce depressive symptoms in two weeks of regular use [3]. Some therapy chatbots have also been able to predict the onset of panic attacks or depressive episodes based on patterns in user behavior (Mathotaarachchi, et al., 2017). A key feature of therapy chatbots is their ability to improve over time by drawing insights from the ever-

<sup>1</sup> Professor of Law, Indiana University Maurer School of Law (Bloomington). Affiliated Faculty and Member of Advisory Board and Governance Committee, The Vincent and Elinor Ostrom Workshop in Political Theory and Policy Analysis, Indiana University Bloomington; Affiliated Faculty, Center for IP Research, Indiana University Maurer School of Law; J.D., Penn Law; B.S., Tufts University; Previously held fellowships at Berkeley Law and Michigan Law.

growing pools of information they receive from their conversations with users.<sup>2</sup> The potential for good seems significant.

An even more tantalizing possibility is that therapy chatbots offer can solve the problem of collecting, pooling, and drawing helpful insights from large sets of mental health data. In the past, only individual therapists or scientists have been able to collect such information in the course of their work. Patient confidentiality, funding constraints, a lack of technological infrastructure and expertise, and limited numbers of patients have made it infeasible to build useful pools of mental health data in treatment or research settings. One such problem is the fact that many people regard their mental health conditions as private. As mentioned, the desire for privacy often stems from a sense of social stigma that many attach to mental health care. Individuals struggling with, say, anxiety, might prefer to benefit from the insights of a data pool without contributing to it themselves. If everyone felt this way, there would be no data to collect in the first place. The effort would fail – a classic free-rider dilemma.<sup>3</sup>

Woebot and conversational agents like it address this problem in a unique way. The system obscures its most valuable asset – broad knowledge or wisdom about how to help people – behind an automated chat interface that the company tightly controls. Unlike a database, this information cannot easily be copied or otherwise expropriated. The asset, so to speak, can be accessed only indirectly through conversing with the software. Chatting in this way requires users to share information with the system. (Importantly, the company keeps all user data private and has pledged not to monetize any user data.) Woebot shows how a chat interface might be able to regulate the flow of mental health data, and in doing so, ameliorate the collective action problems that make aggregating mental health data difficult.

Students of commons governance could regard chatbots like Woebot as technologies that provide an “imposed pattern” for governing the flow of private information. Like other “imposed commons” that scholars have examined in this series of volumes, therapy chatbots are affected by a vast number of exogenous factors, such as laws. The Governing Knowledge Commons (“GKC”) helps situate these new technologies in their broader cultural, medical, legal, and technological contexts.

Drawing upon the GKC framework, this chapter presents an ethnographic study of Woebot – a therapy chatbot designed to administer a form of Cognitive-Behavioral Therapy (“CBT”). Section 3.1 explains the methodology of this case study. Section 3.2 describes the background contexts that relate to anxiety as a public health problem. These include the nature of anxiety and historical approaches to diagnosing and treating it, the ascendancy of e-Mental Health therapy provided through apps, and relevant laws and regulations. Section 3.3 describes how Woebot was developed and

<sup>2</sup> See, e.g., “Digital Me,” created by Patients Like Me.

<sup>3</sup> As Elinor Ostrom wrote in her pathbreaking book, *Governing the Commons*, “Whenever one person cannot be excluded from the benefits that others provide, each person is motivated not to contribute to the joint effort, but to free-ride on the efforts of others. If all participants choose to free-ride, the collective benefit will not be produced.”

what goals its designers pursued. Section 3.4 describes the kinds of information that users share with Woebot. Section 3.5 describes how the designers of the system seek to manage this information in a way that benefits users without disrupting their privacy.

### 3.2 METHODOLOGY

This chapter's approach follows the GKC framework developed by Katherine Strandburg, Brett Frischmann, and Michael Madison (2014). The framework is an adaptation of Elinor Ostrom's Institutional Analysis and Development (IAD) (Ostrom, 1990; Strandburg, Frischmann, and Madison, 2014). I followed the following process.

- A literature review. To gather general information about the state of e-mental health apps and services, I surveyed recently published books, newspaper articles, and academic works related to this topic. This research also covered general interest publications on anxiety and treatments such as Cognitive-Behavior Therapy. From these sources, I identified software and services (apps) designed to help people manage their anxiety. I then contacted individuals who led the companies that produce these apps.
- Semi-structured interviews. I interviewed ten experts with knowledge of either Woebot or the e-mental health landscape more generally. I selected some of these individuals because they were cited or quoted commonly in news articles, books, and academic works. I learned of others in the interview group by asking for contacts and referrals in the first interviews I conducted for this project. Interview subjects included professors, technologists, and CEOs. In keeping with the GKC framework, these interviews were semi-structured and focused on the following topics: (1) the scientific, technological, and social contexts in which this form of information-sharing takes place; (2) the various types of data and related informational assets this group seeks to aggregate and organize access to; (3) the "default" status of these assets; (4) the players involved, including corporations and health-care institutions; (5) the community's goals; (6) rules and related internal governance mechanisms; (7) the technological infrastructure supporting the community.

I conducted all interviews by telephone and recorded them through handwritten notes. The average duration of the interviews was 45 minutes. Some interviews were supplemented with brief follow-up email exchanges. In keeping with Internal Review Board procedures, I furnished each interview subject with an information sheet describing the goals of this study.

### 3.3 BACKGROUND ENVIRONMENT: CONTEXTS

This section focuses on the most prominent landmarks in the anxiety landscape: how the disorder has been understood and treated historically, how anxiety sufferers

typically encounter and experience the problem, the evolving nexus between mental health and technology, and relevant laws and regulations in the United States.

### 3.3.1 A Brief History of Anxiety

Culture has shaped how people have understood and treated anxiety throughout history (Horwitz, 2013). The very language we use to describe the problem is rooted in cultural lore: The word “panic” derives from the Greek god Pan, whose battle cry was said to fill his enemies with uncontrollable fear. In ancient Greece, soldiers who experienced panic in battle were often regarded as cowards and imprisoned or executed. Some historians credit Aristotle (384–320 BCE) for taking the first step toward understanding anxiety as a pathological condition. In exploring the idea of fear, he wrote, “For the man who is by nature apt to fear everything, even the squeak of a mouse, is cowardly with a brutish cowardice, while the man who feared a weasel did so in consequence of disease.” In other words, fear can sometimes be healthy and warranted and other times a sign of illness. (We can surmise that in Aristotle’s time, mice were generally considered a rational source of fear, whereas weasels were not.)

Other Greek philosophers made some of the earliest and most enduring contributions toward Western understandings of anxiety. Hippocrates (460–367 BCE), for example, sought to understand medical disorders by observing his patients in a systematic, scientific manner. This process led him to develop a system for classifying mental disorders, including anxiety. Today, a similar classification system called the *Diagnostic and Statistical Manual (DSM)* is widely used in the medical profession to identify mental health disorders (Ghinassi, 2010). Hippocrates conjectured that the key to treating anxiety was in achieving a kind of harmony between the body and mind (Kleisiaris, Sfakianakis, and Papathanasiou, 2014).

Socrates (470–399 BCE) expanded the set of tools that experts had to investigate anxiety. The Socratic Method, a process of deeply questioning one’s underlying assumptions, is perhaps his most lasting methodological contribution. Like Hippocrates, Socrates also believed that the key to mental health had to do with the relationship between the body and the mind. Understanding this relationship, he believed, required not only physical examination but also introspection on the part of the patient. The idea is captured well in his famous injunction to his followers, “know thyself” (Ghinassi, 2010). Plato (427–347 BCE) believed that anxiety was rooted in incorrect or distorted beliefs about the world. His solution was straightforward: educating sufferers about their erroneous beliefs, and through persistent reminding and reasoning, attempts to alter those beliefs.

Arguably, even more nuanced insights about mental health appear in Buddhist teachings from the fifth century BCE. In his teachings, Siddhartha Gautama (The Buddha) explains that people fundamentally misperceive the true nature of the world. The path away from anxiety and other afflictions, he explains, lies in gaining an accurate view of reality as it is, rather than how we imagine it to be. Buddhist

traditions usually focus on better aligning the body and mind to gain this clearer view. As Thich Nhat Hanh, a widely known Vietnamese Buddhist master, explains, “Wrong-thinking causes us to see the world in an upside-down way. Our mind is often thinking about one thing while our body is doing another. Mind and body are not unified.”

The Middle Ages, which lasted roughly from the year 476 to 1000 BCE, saw a retreat from the scientific methods developed in ancient Greece in favor of the supernatural. In some cultures, the belief emerged that mental disorders are evidence of demonic possession. This idea is likely the historical root of the stigma that many sufferers of anxiety and depression experience today. In some cultures, the notion of demonic possession persists.

The European renaissance sparked a return to ancient insights. In the fifteenth century, for instance, Descartes developed the idea of “mind-body dualism” – the notion, as Descartes described it, that the mind and the body “compose a certain unity.” This insight led to the modern recognition of the fact that the mind and the body influence each other – “bidirectionality” – a core principle of contemporary treatment. These ideas were not new, of course. The Latin phrase “mens sana in corpore sano” (“a healthy mind in a healthy body”) dates to the second century.

Psychology, the modern scientific study and treatment of behavior and the mind, took root and flourished in the late nineteenth and early twentieth centuries. Some leaders in the field, such as Sigmund Freud and Carl Jung, focused heavily on the subjective experiences of patients. Freud believed, for instance, that the roots of anxiety, depression, and related mental health disorders lay in repressed sexual energy. Although this view is not widely credited today, Freud made a lasting contribution to the modern understanding of anxiety by classifying it into different forms: generalized anxiety, panic disorders, phobias, and obsessions (Horwitz, 2013).

Others in the then-emerging field searched for objective sources of information on the disorder. Ivan Pavlov (1849–1936) illustrated that anxiety responses could be conditioned in animals – a discovery that suggested people might develop anxiety disorders as a learned response to life experiences. Pavlov’s work inspired the field of behavioral psychology, which explores the theory that people acquire mental conditions entirely through learning. John Watson (1878–1959), Rosalie Rayner (1899–1935), and B.F. Skinner (1904–1990) are widely known names in this field.

In the 1950s, researchers challenged the notion that anxiety is *solely* a learned behavior. These experts argued that cognition, or how one sees and appraises the world, must play a role too. George Kelly (1905–67) advanced the idea that people perceive the world as a series of mental constructs. Mental disorders may arise when these constructs are inaccurate. This view seems to align remarkably well with the core teachings of Buddhist philosophy, as well as Plato’s teachings, both mentioned earlier. Recent research indicates that life experiences, including stressful environments and incidents, particularly in childhood, can “activate” the disorder.

Cognitive-behavioral therapy (CBT), one of the most effective techniques to treat anxiety today, is grounded in the idea that anxiety is a self-reinforcing behavior that stems from our misunderstanding of the world. Patients who undergo CBT treatment learn about common types of cognitive distortions, and then learn to spot these distortions in their thinking. For instance, some people who suffer from anxiety tend to think about future events in binary or black-and-white terms. In the mind of a person experiencing anxiety, the results of a routine blood test for cholesterol, for instance, will reveal either perfect health or imminent death. The simple act of recognizing this form of thinking and placing a label on it has been shown to drain it of its power.

In the 1950s, the American Psychiatric Association published the first edition of the *DSM*. Now in its fifth edition, the *DSM* is the primary tool that medical professionals use to diagnose mental health disorders, including anxiety. Some critics of the modern psychological establishment argue that the *DSM's* focus on classification implies a level of clarity and precision to the diagnosis of mental health disorders that does not exist. This line of criticism is mostly academic, however, and removed from the day-to-day business of treating people with mental health disorders.

Alongside CBT, another important set of tools for treating anxiety are drugs. People have used chemicals derived from plants to manage stress and anxiety for millennia. Alcohol, one of the earliest such substances, remains one of the most popular. Opium, a drug derived from a type of poppy plant, has been used as a sedative at least since 4000 BCE. In the late nineteenth century, a chemist working at the Bayer Company successfully diluted the active molecule in opium – diacetylmorphine. Bayer branded the chemical as “heroin” and commercialized it in 1898. The drug’s dangerousness and addictive qualities were evident within just a few years of public use. Opiates were eventually replaced by barbiturates, another depressant, which were developed chiefly by a German chemist and Nobel Prize winner named Adolf von Baeyer.

In the 1950s, the pharmaceutical company Smith-Kline began selling a drug it called Thorazine. The drug’s popularity for treating mental health disorders fueled research into the development of new drugs designed to address specific disorders. The first drug that was developed specifically to target anxiety was called Miltown. Although highly effective, it was, unfortunately, also highly addictive. Continued research led to the creation of benzodiazepines, including diazepam (Valium) and alprazolam (Xanax). These drugs are often prescribed to treat acute episodes of panic, but like earlier drugs, they can be highly addictive. In the 1970s, Eli Lilly and Company introduced Fluoxetine (Prozac), a new kind of medicine called a selective serotonin reuptake inhibitor (SSRI). Approved by the FDA in 1987, the drug remains widely used and highly regarded, along with other SSRIs, such as Zoloft, Paxil, Luvox, Celexa, and Lexapro. SSRIs are the most commonly prescribed drugs in the United States and by far the frontline defense against anxiety. Despite the significant

advances in drug development that took place in the twentieth century, very few new drugs have been developed to target anxiety since the 1990s.

Recent studies have shown that today, drugs (serotonin inhibitors) and CBT are the two most effective and standard tools to manage clinical anxiety [4]. Several complementary and integrated health treatments have also been shown to reinforce the benefits offered by conventional medical treatment. These include stress and relaxation techniques, yoga, meditation, and acupuncture [5]. Getting adequate sleep, regular exercise, maintaining a healthy diet, having a robust social support system, and living with low levels of stress are also important [6]. For people who have access to adequate mental health care, educational resources, and the ability to follow a beneficial lifestyle, anxiety can be highly manageable. Unfortunately, as discussed in the following section, there is a shortfall in access to mental health care.

### 3.3.2 *The Patient Context*

Nearly all anxiety disorders involve intense feelings of fear and the anticipation of imminent harm (Hilty and Mucic, 2016; Rachman, 2020). A short review of several types of anxiety shows subtler aspects of the problem. Under the heading of Anxiety, the *DSM* identifies the following: (i) generalized anxiety disorder, (ii) panic disorder, (iii) social anxiety disorder, (iv) obsessive-compulsive disorder, (v) post-traumatic stress disorder, and (vi) specific phobias:

- (i) In generalized anxiety disorder, feelings of worrying and fear permeate everyday life. Sometimes, these feelings are tied to real causes for worry, but they are disproportionately intense. (Recall Aristotle's mouse.) Other times, the feelings may be disconnected from any cause that the sufferer can identify – a kind of free-floating dread. These feelings are often accompanied by physical discomfort, such as muscle tension, restlessness, queasiness, and nausea. Those with generalized anxiety disorder are focused heavily on the future, both near- and long-term, and unable to tolerate feelings of uncertainty (Rachman, 2020).
- (ii) Panic disorders, another class of anxiety disorder, are typified by intense physical and mental episodes of uncontrollable fear brought on by a perceived threat that typically does not exist. These episodes are different from the chronic worrying and fear that generalized anxiety disorder brings on: they are more physically intense and often mistaken by sufferers for heart attacks.
- (iii) Social anxiety is identified by unwarranted feelings of embarrassment and self-consciousness. Unsurprisingly, people who suffer from this problem often isolate themselves.
- (iv) Obsessive-compulsive disorder is characterized by repetitive or compulsive behaviors, such as hoarding, arranging objects in specific patterns, and extreme avoidance of germs.

- (v) Another manifestation of anxiety, post-traumatic stress disorder (“PTSD”), has received a wealth of attention in the media because soldiers often experience it. Unlike generalized anxiety disorder or panic disorders, which inherently focus on the future, this problem is heavily focused on the past. Often a traumatic experience gives rise to nightmares and related, intensely upsetting memories.
- (vi) Lastly, are specific phobias – the experience of anxiety symptoms brought on by certain situations or objects.

As mentioned in the introduction, anxiety is the most common mental health disorder in the United States. According to the Anxiety and Depression Association of America, 40 million adults age 18 or older, or about 18.1 percent of the population, is affected by anxiety every year [1]. Although the United States has the highest rate of anxiety in the world, the disorder is prevalent elsewhere. Interestingly, there does not seem to be a correlation between anxiety and industrialization: rates are very high in France, Colombia, and Lebanon. China and Nigeria, meanwhile, have relatively low rates. Moderate evidence suggests that culture and context impact how often the problem arises. Poverty correlates with higher rates of generalized anxiety disorder, for instance.

Recent research indicates that life experiences, including stressful environments and incidents, particularly in childhood, can “activate” the disorder (Ghinassi, 2010). Studies have shown that the development of anxiety can be mitigated, however, by positive protective factors. Chief among these are parents who are warm and sensitive to the feelings of a child, and who impart upon the child an internal “locus of control” – a feeling that the child has control over his or her life. Although childhood experiences seem to be foundational, anxiety disorders can arise at any age in a person’s life. The average age of onset for generalized anxiety disorder, for instance, is thirty-one. As mentioned in the introduction, it is estimated that the total annual cost of anxiety runs in the neighborhood of \$47 billion annually.

One of the reasons that anxiety rates are soaring in the United States could be the fact that the problem is widely under-treated (U.S. Department of Health and Human Services, 2018). One reason for this is stigma: Many people with anxiety are reluctant to seek treatment because they believe others (family members, neighbors, etc.) will judge them poorly if they learn they suffer from anxiety. As with so many aspects of anxiety, the perception of stigma varies greatly by age, ethnicity, religious, and other cultural lines.

In addition to stigma, recent studies have shown that in many areas of the United States, there is a workforce shortage in child psychiatry and in other behavioral health practitioners (Health Resources and Services Administration/National Center for Health Workforce Analysis, 2015). The problem of access is particularly acute in rural areas: A recent study found that 65 percent of non-metropolitan counties do not have a single psychologist, and 47 percent of non-metropolitan areas do not have a single psychologist (Andrilla, et al., 2018). This problem is likely

to worsen. The U.S. Department of Health and Human Services has projected that by the year 2025 there will be shortages of between 6,080 and 15,400 psychiatrists; between 8,220 and 57,490 clinical, counseling, and school psychologists; and as many as 26,930 mental health counselors (Health Resources and Services Administration/National Center for Health Workforce Analysis, 2015). Closely tied to the access problem is the cost of mental health care: Many people with anxiety cannot afford treatment.

### 3.3.3 *Technological Contexts*

#### 3.3.3.1 E-Mental Health Services

Alongside CBT and drugs, several new technologies have been developed in the past twenty years that have helped anxiety sufferers. In the medical world, many of these technologies are referred to generally as “e-Mental Health Services.” Since the early 2000s, online services have helped connect patients to therapists online, for instance. Today, several companies provide remote talk therapy sessions carried out over smartphone apps. Similarly, there are a number of services that have made it possible for trained practitioners to administer CBT remotely. Two such services are Joyable and Talkspace. Joyable delivers an eight-week CBT program administered by a trained practitioner, and Talkspace connects users to licensed mental health professionals.<sup>4</sup>

Recent studies have shown that online mental health interventions can be as effective as traditional in-person therapy for treating both depression and anxiety. In one study, for instance, the recovery rates of individuals treated with online CBT for panic disorder and agoraphobia were comparable to people treated in-person. Based on a thirty-month follow-up study for the treatment of social phobia, the long-term effects of CBT delivered over the internet was comparable to CBT carried out over the internet (Mucic, Milty, Parish, and Yellowlees, 2016).

Sometimes, e-Mental Health Services can be even more effective than traditional care. As mentioned earlier, the stigma that some patients associate with visiting a clinician can discourage them from seeking help. Because patients can access e-Mental Health Services from their homes, this barrier disappears (Shoemaker and Hilty, 2016). Recent studies reveal that LGBTQ+ youth, for example, prefer to seek mental health care online (Dreyfuss, 2019). As discussed later in this chapter, the privacy of user data could either reinforce or erode this perception of anonymity.

In addition to linking patients to traditional mental health providers, many apps facilitate complementary services and treatments that may be helpful. These include thousands of apps and online services related to yoga, mindfulness, eating disorders, tracking mental health conditions, and general stress reduction (Mucic,

<sup>4</sup> Joyable was acquired by a company called AbleTo in early 2019 [7].

Milty, Parish, and Yellowlees, 2016). A website called “Psyberguide” provides a helpful catalog of these services, along with descriptions and ratings. The website is maintained by Dr. Stephen Schueller, an associate professor of psychological science at the University of California at Irvine [8].

### 3.3.3.2 Social Networking

Social media marks another nexus between anxiety and technology. At least in theory, an online social network could help anxiety sufferers learn more about their problems and potentially to connect with others who are navigating this complex and challenging disorder. This could help to alleviate feelings of isolation and helpfully establish new social support systems for sufferers.

An emerging body of research has shown that mainstream social networks such as Facebook and Twitter can contribute to anxiety (e.g. Hoge, Bickham, and Cantor, 2017).<sup>5</sup> The chief symptoms of this disorder are anxiety symptoms experienced by someone who is unable to check their social media notifications for even a short time [9]. Separately, the use of Facebook and Twitter has been linked to symptoms of depression. As of this writing, this body of research is still developing, and it is unknown how significant or widespread these problems are, however (Pantic, 2014).

Beyond the most widely visited social networks, a number of online communities exist for people who suffer from anxiety. Health Unlocked, one such service based in London, administers a collection of chat forums where patients can connect, share advice and encouragement with one another online. Interviewed for this book chapter, a person with knowledge of the company’s services described it as “a complement to clinical treatment” that “breaks down the isolation” and possibly provides access to “helpful information you wouldn’t always find from a doctor.” According to the individual interviewed, the service is free, and the company generates revenue by selling anonymized and aggregated data to partners, including pharmaceutical companies.

### 3.3.3.3 Therapy Chatbots

Therapy chatbots mark another nexus between mental health and technology. The development of such apps could be traced back to the 1950s, when Alan Turing, an English mathematician, considered whether a computer could simulate a human conversation. To analyze the question, Turing developed a thought experiment that involved two people and a machine designed to mimic a person. One of the two human participants, whom Turing called the “interrogator,” types questions to a subject located in a different room. The subject responds through typed messages. The interrogator must then guess whether the subject is the other human or the

<sup>5</sup> Notably, the American Medical Association has not yet recognized this disorder.

machine. Turing used this hypothetical setup to show that, in principle, there is nothing to stop a cleverly programmed computer from fooling the interrogator. Turing's work solidified his place as the grandfather of modern AI.

Turing's work also inspired generations of researchers to build chatbots. The first, a chatbot named "Eliza" and developed in 1966, simulated a psychotherapist. Although the system could not comprehend what was typed to it, Eliza was able to simulate human-like understanding through simple pattern matching. In 1972, a chatbot named "Parry" cleverly fooled many human interrogators by simulating a person suffering from paranoid schizophrenia. In 1992, a chatbot "therapist" named "Dr. Sbaitso" was developed for MS-DOS-based computers. This "doctor" was notable because it was one of the first chatbots that were widely available to the public. Ultimately, a limitation of the early chatbots was the fact that their behavior was scripted through rules. Given a certain input, they would always produce a pre-programmed output. Programming all of the possible conversations that a person might wish to have with a machine would, in theory, require providing the machine with an internal model of the universe.<sup>6</sup>

Since the late 2000s, chatbots or conversational agents have grown more sophisticated thanks to a new approach called "machine learning" ("ML"). In contrast to rule-based systems like Eliza, machine learning is a process by which a computer can identify and match patterns in massive sets of data. Provide an ML-based chatbot with enough examples of typical two-sided human conversations, and over time, the software will learn, through statistics and probability, what a good response to a new question it has never seen should look like. Many people saw the technology on display for the first time when IBM's Watson system defeated the Jeopardy Champion, Ken Jennings, before a national audience in 2011. For many of us, examples closer to home include personal assistants built into our phones and smart speakers. Apple's *Siri*, Google's *Assistant*, Amazon's *Alexa*, and Microsoft's *Cortana* have (for better or worse) integrated themselves into billions of people's daily routines.

Against this backdrop, a new generation of therapy chatbots has recently been developed and provided to the public (Hernandez, 2018). The first such service, Woebot, is the focus of this study. Tess, a chatbot developed by San Francisco-based X2AI, also delivers CBT interventions, along with similar techniques, including mindfulness therapy, acceptance and commitment therapy, self-compassion therapy, and interpersonal psychotherapy (Fulmer et al., 2018; Gionet, 2018; Green et al., 2019; Joerin, Rauws, and Ackerman, 2019; Stephens et al., 2019) [14]. The service works entirely through a text-like exchange. Wysa, a startup based in London and Bangalore, similarly delivers "CBT, mindfulness techniques, and other techniques via in-app text changes" (Wallach, 2018). When a patient uses Wysa for the first time,

<sup>6</sup> "But as those projects unfolded, researchers hit a major problem: there were simply too many rules that needed to be encoded for a system to do anything useful. This jacked up costs and significantly slowed ongoing efforts" (Pantic, 2014).

the app asks a series of questions designed to help it understand what sorts of problems the user is grappling with. Some chatbot designers are moving beyond text in order to better diagnose patients. AiME (Chin, 2018), a therapy chatbot developed by a team of researchers at UCLA, aims to detect risks of depression, anxiety, and addiction based on several minutes of video footage of a user's face (Condliffe, 2017; Garg and Glick, 2018; McCaney, 2019; Morris, Kouddous, Kshirsagar, and Schueller, 2018; Rieland, 2018; Rosso, 2018). To capture video and audio, the app requests access to the microphone and camera on a user's device.

Early evidence, while limited, suggests that these tools are effective. A peer-reviewed study shows that Tess can significantly reduce symptoms of anxiety and depression, for instance. Youper, another entrant in the therapy chatbot industry, claims that more than 80 percent of its users "experience a reduction in negative moods after just one conversation." Studies show that Wysa has promise in reducing anxiety symptoms, particularly in assisting children [10].

Because therapy chatbots need to collect private information from users in order to converse with them, these apps also carry privacy risks. These risks may seem especially serious today, as the technology industry is facing a widespread reckoning with user privacy. The privacy problem has played out in the form of countless user data leaks, hacks, and unauthorized disclosures by large technology companies – most notably, Facebook. The privacy issue is explored more deeply later in this chapter.

### 3.3.4 *The Legal and Regulatory Landscape*

The emerging therapy chatbot industry is taking form against a complicated patchwork of law and regulations. This policy framework significantly influences how these technologies are developed, funded, and provided. It is important to note that the areas highlighted here are simply the most visible and may not provide a complete picture of the laws and regulations relevant to every therapy chatbot, however.

#### 3.3.4.1 Privacy Laws

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) forbids healthcare providers such as hospitals from disclosing specific types of patient data. These include patient names, zip codes, and dates of treatment.<sup>7</sup> A number of other privacy laws enacted at the state and federal levels impose civil liabilities for wrongful disclosure of such data by health-care providers.<sup>8</sup> Interestingly, HIPAA rules pertain only to so-called covered entities, which include health-care providers

<sup>7</sup> 18 USC 1836 et seq.

<sup>8</sup> 45 C.F.R. 164.514(e) (2013).

(e.g., doctors, clinics, psychologists), health plans (e.g., health insurance companies, HMOs), and health clearinghouses (entities that process health information into standard formats).

Therapy chatbots appear to operate outside the purview of HIPAA by billing themselves as “wellness apps” rather than new forms of traditional psychotherapy. This distinction seems murky, however, especially in light of the fact that patients are encouraged to share the same kinds of information with these apps as they might traditionally share with a therapist. John Torous, the chair of the American Psychiatric Association’s smartphone app evaluation group, has said that the “gap” in HIPAA coverage raises privacy concerns (De Salvo and Samules, 2016; Nutt, 2017).<sup>9</sup> A recent government report on this coverage gap made the same conclusion, stating, “As the electronic sharing and storage of health information increases, and as individuals become more engaged in sharing personal health information online, organizations that are not regulated by HIPAA, the FTC, or state law may collect, share, or use health information about individuals in ways that may put such data at risk of being shared improperly” (Department of Health and Human Services, 2016).

Depending on its geographic reach, a therapy chatbot could also be subject to the requirements of the EU’s General Data Protection Regulation (GDPR). The GDPR ensures that EU citizens are able to access their personal data, as well as information about how their data is processed. It also provides individuals the right to request that their personal data is deleted promptly upon request. The regulation also requires that companies collecting personal data design their products around user privacy. This requirement could include encrypting all personal data, making sure such data isn’t transferred unnecessarily and ensuring that encryption and decryption of data aren’t conducted on remote servers.

### 3.3.4.2 The FDA

The Food and Drug Administration (“FDA”) has had the authority to oversee the safety of medical devices since the 1930s. It does so by requiring device manufacturers to demonstrate the safety of their products. (If they cannot do so, the devices may not be made available for sale.) Traditionally, the FDA’s definition of “medical device” offered a fairly predictable and clear indication of what sorts of products might receive oversight. The recent rise of smartphones has presented a new and challenging question, however: Can a smartphone running an app constitute a medical device?

Under the FDA Act’s broad definition of “medical device,” the agency has the authority to regulate medical software before it is released to the public. Although the FDA has declined to oversee most types of mobile health apps carefully, it has published guidance documents that leave open the possibility that it could regulate

<sup>9</sup> Snowdy, Christopher E., Erica Z. Shoemaker, Steven Chan, and Donald M. Hilty. “Social Media and Clinical Practice: What Stays the Same, What Changes, and How to Plan Ahead?” In *e-Mental Health*, pp. 151–170. Springer, Cham, 2016.

mental health apps. The FDA has said it intends to exercise enforcement discretion over apps that (i) help patients (i.e., users) self-manage their disease or conditions without providing specific treatment or treatment suggestions; (ii) provide patients with simple tools to organize and track their health information; (iii) provide easy access to information related to patients' health conditions or treatments; (iv) help patients document, show, or communicate potential medical conditions to health-care providers; (v) automate simple tasks for health-care providers. The twenty-first century Cures Act (CURES) excludes from FDA oversight apps "for maintaining or encouraging a healthy lifestyle and is unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition."<sup>10</sup> By contrast, according to an expert at Woebot, the FDA might be more motivated in the future to regulate technologies that make clear claims regarding a particular clinical indication.

Some legal commentators are troubled by the FDA's current approach, in light of the potential risks that patients may face when using psychotherapy apps that don't work in accordance with accepted clinical procedures. Specifically, on the subject of CBT, Theodore Lee has commented, "untreated or poorly treated conditions can lead to adverse outcomes for the patient or others. The lack of conformity with established CBT principles for most CBT-based apps suggests that patients may not be getting adequate treatment" (Lee, 2018). Despite these concerns, however, the FDA appears reluctant to regulate mobile mental health apps, and Congressional action like the CURES Act has reinforced this hands-off stance by removing many such apps from FDA purview.

### 3.3.4.3 The FTC

The Federal Trade Commission is charged to "protect consumers against unfair or deceptive acts or practices in commerce." The agency accomplishes this in a variety of ways, including legal enforcement and providing education. Although the FTC has not initiated legal action against any manufacturers of mental health apps, it recently filed a complaint against a "brain training" app called Luminosity that claimed to provide cognitive benefits. Specifically, the app promised to "protect against . . . age-related conditions such as mild cognitive impairment, dementia, and Alzheimer's disease; and will reduce cognitive impairment associated with . . . post-traumatic stress disorder, traumatic brain injury, attention deficit hyperactivity disorder, Turner syndrome, stroke, and other health conditions."<sup>11</sup> The FTC claimed that this language was deceptive. Under a settlement agreement, Luminosity paid \$2 million and was enjoined from future

<sup>10</sup> CURES Act, Section 3060 (page 1130).

<sup>11</sup> Complaint at 5, *F.T.C. v. Lumos Labs, Inc. d/b/a Luminosity*, No. 3:16-cv-00001 (N.D. Cal. January 4, 2016), ECF No. 1, [www.ftc.gov/system/files/documents/cases/160105lumoslabscompt.pdf](http://www.ftc.gov/system/files/documents/cases/160105lumoslabscompt.pdf) [<https://perma.cc/5ME3-A3Y9>].

deceptive conduct.<sup>12</sup> The suit serves as a cautionary tale to any company that cannot substantiate the benefits they claim to provide.

#### 3.3.4.4 The VA and DARPA

Interestingly, the US Department of Veterans Affairs (“VA”) has developed and delivered e-Mental Health tools to soldiers as well as the public. As Davor Mucic and Donald Milty explain in their book *e-Mental Health*,

Once again, the military may be in the lead for providing Internet-based care options, with foci of remote screening and assessment, post-deployment adjustment, suicide prevention and management, and delivery of training/education. Many personnel prefer to receive care at home – particularly those with anxiety, PTSD, and phobia – in addition to those who prefer to avoid the stigma of entering a mental health facility (Mucic, Milty, Parish, and Yellowlees, 2016).

The VA has developed and made available several apps for treating PTSD and other forms of anxiety. These include PTSD coach, an app that provides users with the ability to track and monitor PTSD symptoms; CBT-i, an app designed to deploy CBT to treat insomnia; ACT Go, which delivers acceptance and commitment therapy; and Moving Forward, an app designed to help reduce stress. All of these apps are provided to the public at no charge.

DARPA, the branch of the US military that focuses on developing new technologies, has also been involved in developing therapy chatbots. In 2014, the agency funded a study of “Ellie,” a chatbot therapist developed by researchers at the University of California’s Institute for Creative Technologies (Gratch, 2014).

#### 3.3.4.5 Intellectual Property

Intellectual property protection is also relevant to chatbot therapy apps. Patent protection is designed to encourage investments into developing new technologies by offering inventors an exclusivity window of twenty years. To receive patent protection, an invention must be new, non-obvious, and useful. Importantly, it also must qualify as patentable subject matter. For companies seeking to patent chatbot therapy software, this requirement could present some challenges. Although software has traditionally enjoyed robust patent protection since the late 1990s,

<sup>12</sup> Stipulated Final Judgment and Order for Permanent Injunction and Other Equitable Relief at 5–9, *F.T.C. v. Lumos Labs, Inc. d/b/a Lumosity*, No. 3:16-cv-00001 (N.D. Cal. January 8, 2016), ECF No. 10, [www.ftc.gov/system/files/documents/cases/160105lumoslabsstip.pdf](https://www.ftc.gov/system/files/documents/cases/160105lumoslabsstip.pdf) [https://perma.cc/7JZL-GK7Y]. These qualifications to the settlement are not atypical of FTC enforcement actions in digital health. See, e.g., Stipulated Final Judgment and Order for Permanent Injunction and Other Equitable Relief against Defendants Kristi Zuhlke Kimball and New Consumer Solutions LLC at 8–9, No. 1:15-cv-01614 (N.D. Ill. April 30, 2015), ECF No. 27, [www.ftc.gov/system/files/documents/cases/new\\_consumer\\_solutions\\_5-1-15.pdf](https://www.ftc.gov/system/files/documents/cases/new_consumer_solutions_5-1-15.pdf) [https://perma.cc/UJQ8-V3LV].

a Supreme Court ruling from 2014 appeared to narrow this form of protection to include only software that does not constitute an abstract idea and that includes an inventive concept. The Court's lack of specificity over the meaning of these terms has called into question the validity of many software patents since the ruling.

Despite these uncertainties about patenting algorithms, however, there still appear to be opportunities for patenting in the realm of chatbot therapy apps. Ginger.io, a company that provides a therapy chatbot service, has received eleven patents for, among other things, the modeling of psychological states (e.g., moods, conditions) based on what a user has expressed to their app, and for providing therapeutic interventions such as CBT to users based on those models. IBM, which created the widely known "Watson" AI system, has also received a number of patents in this space. The potential for acquiring patents may explain why academics who have sought to commercialize therapy chatbots have left their universities: Universities often require their employees to assign intellectual property they develop to their employers.

Trade secrecy, meanwhile, seeks to encourage investments in innovation by a different means: providing a cause of action that can be asserted against anyone who wrongfully misappropriates valuable and secret information. Trade secret protection is provided through state and federal statutes. Because valuable software systems and data can often be kept secret through tools like encryption and other security measures, trade secret protection has played an important role in many industries that rely upon data.

### 3.3.4.6 Public–Private Partnerships

Public–private partnerships are yet another place where law and policy have affected the mental health app industry. In 2018, state and county mental health officials in California worked closely with two Silicon Valley companies, Mindstrong and 7Cups, to evaluate the efficacy of mental health apps for patients of the public mental health system (Carey, 2019). Mindstrong's app excels in identifying patterns in user behavior that could signal the onset of certain types of psychological episodes. 7Cups, meanwhile, connects patients with trained "listeners" who can connect them to trained therapists.

As of this writing, press reports indicate that the effort has faced challenges related to recruitment and informed consent relating to the disclosure of user data. Commenting to the *New York Times*, Dr. John Torous, director of the division of digital psychiatry at the Beth Israel Deaconess Medical Center in Boston stated, "If we're excited about the potential of data, we should be equally worried about the risks, and those seem to be evolving faster than the scientific benefit . . . There may be guarantees the companies make about not sharing data, but if the company is sold to another company, that data goes with it," he said. "A lot of apps have that clause buried in 13 pages in mouse print" (Carey, 2019).

### 3.4 GOALS, OBJECTIVES, AND HISTORY

The Woebot story began in 2016 when Alison Darcy, a Stanford University clinical research psychologist, began exploring the potential of a therapy chatbot to help people with common mental health disorders. At the time, Dr. Darcy was working alongside Andrew Ng, an eminent expert in the field of artificial intelligence. Dr. Darcy became convinced that a well-designed chatbot could truly make a difference in the lives of people who suffer from anxiety and depression – so convinced, in fact, that she decided to leave academia to build a commercial product full-time.

The science behind Woebot is based largely on cognitive-behavioral therapy. As Darcy explained in a recent interview with *Wired*, “CBT asks people to recast their negative thoughts in a more objective light. Patients are encouraged to talk about their emotional responses to life events, and then stop to identify the psychological traps that cause their stress, anxiety, and depression” (Molteni, 2017). Woebot delivers CBT in the form of tightly scripted conversations, which are usually triggered when users describe problems they are experiencing. In response to a user who tells the system they are anxious, for instance, Woebot might tell the user about common distorted patterns of thinking in people who experience anxiety. The system then might walk the user through a common CBT exercise, such as writing down a recent anxious thought and searching for cognitive distortions in it.

Dr. Darcy and the team she assembled believed that Woebot couldn’t succeed if users didn’t enjoy using it and returning to it. The threshold challenge for her and her team thus became a question: “How do you make CBT in a way that people want to use every day?” Dr. Darcy and her team found answers to this question in an unlikely place: the world of gaming. “As a field, gaming has developed some of the most sophisticated engagement methodologies you can find,” she explained. Although gaming and therapy might seem like disconnected fields on the surface, Dr. Darcy and her team saw a meaningful connection: “A game places you in a suspended reality, and therapy is also a suspended reality. For example we typically want someone to ask themselves, “what if you didn’t have to think in a particular way”? In a sense, “you get to try out, and experience a different identity.” One technique that seems to reflect Woebot’s gaming DNA are quizzes: after teaching a user a new idea, the system often presents a short quiz designed to test the user’s understanding.

During the development process, which took about nine months (Molteni, 2017), Dr. Darcy and her colleagues asked a team of writers and therapists to help make Woebot’s conversational abilities sound as natural as possible [10]. According to the company, this work was successful. “People started calling it by name,” Dr. Darcy explained. She emphasized, however, that Woebot is not designed to fool users into thinking it is a human. Instead, it is very transparent – humorously so, at times – that its entire personality is nothing more than artifice. Interestingly, Dr. Darcy believes

that a human user's awareness that Woebot is not a human could lead to be more candid conversations.

A peer-reviewed study conducted by researchers at Stanford and Dr. Darcy concluded that Woebot was "a feasible, engaging, and effective way to deliver CBT." The study compared the outcomes for two groups of individuals recruited from a university setting who agreed to either engage with Woebot or receive information about mental health from the National Institute of Mental Health's website. The study reported as follows: "The study confirmed that after 2 weeks, those in the Woebot group experienced a significant reduction in depression" (Fitzpatrick, Darcy, and Vierhile, 2017).

### 3.5 ATTRIBUTES: THE CHARACTERISTICS OF INFORMATION SHARED

Woebot collects a wide variety of data from users. Nearly all of this information is unstructured text collected through conversational prompts presented to users in the Woebot app, which is available for iOS and Android operating systems (Molteni, 2017). Importantly, as Dr. Darcy explains, the system absorbs "only as much as you divulge to it." Woebot could, but does not necessarily, record a user's name, the problems the user wishes to address – for example, anxiety, depression, loneliness – and specific thoughts the user has had associated with those problems – that is, "I'm not popular enough" and "I'm worried I'll fail an upcoming test."

To appreciate the variety of data that can be collected from users, consider the following exchange. (This is a real exchange that I carried out with Woebot for this book chapter.)

- WOEBOT: Howdy Mike, so nice to hear from you. How are you feeling today?  
 USER: Anxious.  
 WOEBOT: Are you experiencing these feelings right now?  
 USER: Yes, now.  
 WOEBOT: I see. Let's get ready to identify & write your negative automatic thoughts in this moment . . .  
 USER: OK, I'm worried that I might develop a serious illness because I'm feeling a bit unwell.  
 WOEBOT: btw, I've remembered your most frequent distortions in thinking. Would you say that "All or Nothing" was in the thought?  
 USER: Yes, actually.  
 WOEBOT: Now it's time to rewrite your thought without any distortions. Give it a go!  
 USER: Feeling a bit unwell does not mean that I have, or will have, a serious illness.

This short dialog shows how Woebot uses its memory of a user's name and a cognitive distortion the user has experienced to deliver a CBT technique. The exchange also illustrates why a patient-driven modality like CBT is so well suited to a chatbot interface. As Dr. Darcy explained in an earlier interview with *Wired*, "A good CBT therapist should facilitate someone else's process, not become a part of it" (Molteni, 2017). According to Woebot, the service engaged in two million conversations per week in early 2019 (Nutt, 2017).

According to Dr. Darcy, the system retains records of all of the chats it has with users and stores them on a secure server. The company then de-identifies and aggregates these records to draw new insights from them. This underscores an important point: Woebot does not use machine learning techniques to generate its written responses to users. Instead, it uses ML to initiate partially scripted exchanges with users. Commenting for a recent magazine interview, Darcy explained, "Woebot uses machine learning to direct the right content or methods to that person in that moment. However, it is true that we do this in a highly controlled way" (Lee, 2018).

The company has already found interesting insights in its pool of data. "Some of the things we're learned have surprised us in the aggregate," Dr. Darcy explains. "There are things you just don't see in a traditional setting with one data point. We have been surprised, for instance, to see how many people experience loneliness. Some younger users report feeling lonely while they are in social settings. We've also learned from people reaching out in the middle of the night that they have postpartum depression." Dr. Darcy and her team have also been struck by the remarkable similarity in what users write. "It's not just that people say similar kinds of things – they say *the exact same thing*," she says. Woebot's data scientists apply insights like these to improving the topics that Woebot focuses on, and how the software interacts with users. In this way, the more that each user interacts with the system, the more helpful the system can become to all users.

### 3.6 GOVERNANCE

Woebot Labs is a privately held corporation founded by Dr. Alison Darcy in 2017. The company's headquarters are in San Francisco. According to reports provided by PrivCo, a service that provides research data on privacy companies, Woebot received \$8 million from venture capital investors in March, 2018. In exchange for this "Series A" investment, the investors received equity in the company. The company's website lists the CEO (Alison Darcy), the Chief Clinical Officer (Athena Robinson), the Chairman of the Board of Directors (Andrew Ng), and the Chief Technology Officer (Casey Sackett) (Knight, 2017). Decisions relating to Woebot and the user data that underlies it are entirely made within the hierarchical structure of the corporation. Corporate decisions may be constrained to some extent by the company's contractual relationship with its source of VC capital. Although the contract is not public, such

deals often grant the shareholders significant power over key corporate decisions, such as the sale of the company.

Beyond the formal corporate structure, professional standards, and codes of conduct constrain the behavior of Woebot's leadership. The CEO, for instance, is a clinical research psychologist and adjunct faculty member of Stanford School of Medicine. The company's CCO is a licensed clinical psychologist and an adjunct clinical associate professor in psychiatry and behavioral sciences at Stanford School of Medicine. Thus, the two most important decision-makers at the company are familiar with and must adhere in varying respects to codes of conduct and ethical standards promulgated by Stanford University. These require a psychologist to take reasonable precautions to protect patient confidentiality. Moreover, as a clinical psychologist, the CCO must adhere to similar codes and standards promulgated by the American Psychological Association and the state of California. However, whether Woebot is regarded as a therapeutic intervention and whether users of a therapy chatbot might be "patients" under a particular rule could be debated.

The company's privacy policy offers a view of how user data is treated [11]. Importantly, the service is available in two forms: through a stand-alone app and through Facebook's Messenger chat service. Woebot's privacy policy states, "Conversations with Woebot within Facebook Messenger are subject to the Facebook privacy policy. Facebook can see that you are talking to Woebot, and they can see the content of the conversations." Users seeking greater privacy than this will likely opt for Woebot's standalone app, which is available on iOS and Android, the two most widely used mobile operating systems. Without Facebook acting as an intermediary, Woebot alone can control the flow of user data.

To help explain this information flow, the company helpfully provides a chart on its website that divides user data into different categories and explains the company's purpose for collecting and processing it. The categories of user data include: "Personal information (including first name), password, referral code for participation in special programs, and email address," "Financial and billing information," and "Conversation data," which the company explains includes "Information, participation data, text, graphics, video, messages, responses to treatment and satisfaction surveys, or other materials generated through your interactions with Woebot." Under a section of the policy titled, "Where does Woebot obtain my personal data from," the policy states, "Most of the personal data we process is obtained from you when, through the application you: register for a Woebot account and exchange messages with Woebot."

The policy then defines what the company may do with this data. Under a section of the policy titled "Use of De-identified and Aggregated Information," the policy states that user data may be used to "create de-identified and aggregated information," for "analyses we create." Such analyses would presumably include the types of internal studies of user interactions with the system described by Darcy earlier in this chapter.

Woebot does not make user data directly available to any users. As explained earlier in this chapter, the company takes precautions against unauthorized data access by storing user data on a secure server. The company's privacy policy states that the company may share it (with some restrictions) with certain third parties, however. According to the policy, these parties include "service providers who provide technical and support services to help us provide an improve the product and its services." The policy also explains that the company may share personal data with third parties for a variety of legal reasons, including compliance with applicable law, law enforcement or national security requests, subpoenas and the like. The policy also explains that the company reserves the right to share user data with insurance companies, financial institutions, or in connection with a corporate reorganization, merger, or acquisition. Finally, it states that user data may be shared with third parties in connection with "Special Programs." The term "Special Program" is not defined in the policy, but the document states that relevant third parties under this provision may include "your employer, certification authorities, or other medical and academic partners who conduct the study."

On the whole, Woebot's governance of user data shows the unique potential that chatbots may have to aggregate health data and, in turn, address the mental health crisis. By allowing users to receive help through a smartphone app, the service removes the sense of embarrassment that some people may feel about visiting a therapist in person. This, along with the company's treatment of user data, affords potential users with a sense that their conversations will be kept private. To benefit from the service, users must converse with Woebot. With each such conversation, the designers of Woebot have more data with which they can improve the service for others. And, because Woebot can talk to many users at the same time, the service can connect with more people than the current population of mental health workers is able to.

### 3.7 CHALLENGES

Because therapy chatbots have been available to the public for just a short time, a clear picture of the challenges Woebot and its competitors might face hasn't yet developed. It is possible, however, to discuss challenges that the company and experts in the industry perceive.

Although many bodies of law and regulation are relevant to therapy chatbots, one of the most significant relates to user privacy. Privacy is critical to Woebot for several reasons. First, the product solicits or otherwise prompts users to share information they consider private related to mental health conditions. Interestingly, a 2018 public blog post written by Dr. Darcy suggests that decisions about user data will be guided by long-standing rules and norms from the worlds of medicine and science: "First and foremost, this company was established by psychologists, which means we are

subject to an ethical code of conduct. For us, this is so obvious that it seems self-evident.”

The economic potential for mental health data seems significant. As an expert on medical ethics for this book chapter explained, insurance companies have an enormous interest in obtaining this data. “The question that policymakers should focus on,” she said, “is whether they should have access to it. Should people with preexisting conditions revealed by their data still receive coverage? Can an insurer require you to sign up for one of these apps and use them?” These questions led the expert to opine that new laws limiting the data that insurers may receive and act upon should be enacted.

In addition to insurers are advertisers. In the 2010s, a common business model for online services is targeted advertising – the practice of directing advertisements to users based upon what they might like to buy. Online services usually accomplish this by collecting and processing data about their users and building up profiles that reflect users’ probable likes, dislikes, interests, and so forth. For a fee paid by retailers, they direct ads to users. This business model has been phenomenally lucrative to ad sellers, helpful to ad buyers, and it has helped fuel countless avenues of important research that stands to benefit the public. Privacy advocates have long argued, however, that this practice can be harmful when users do not have a clear view of what information has been collected about them, with whom this information has been shared, and the purposes for which it is has been shared (Wu, 2017). These perspectives gained support when, in 2018, press reports revealed that user data collected by Facebook had been shared with an organization called Cambridge Analytica for the purpose of political advertising. The news stories garnered enough of the public’s attention to blossom into a scandal that, over time, has brought privacy to the forefront of public policy discussions about technology.

Woebot’s creators are keenly aware of how these developments might harm existing users and discourage potential users. Since the product was launched in 2017, Woebot has been unequivocal in their pledge not to monetize user data for third-party advertising (Molteni, 2017). Interviewed for this book chapter, Dr. Darcy explained that advertising runs counter to the company’s core mission: “What we’re best at,” she explained,

is this ability to challenge distorted thinking. And one of the most commonplace distorted thoughts is “I’m not X enough.” Almost everybody has that core distortion, and it’s something the ad industry has fed off of. It is the belief that allows ads to be effective. The idea that we would sell our data for advertising goes against everything the company stands for.

The post goes on to explain that selling data to advertisers would “instantly undermine” the trust users have placed in Woebot. “Our investment has been in symptom change outcomes – not in data gathering. If it were the latter, we would have built our product to orient around gathering as much data as possible” [12].

Consistently, the company has been a champion of GDPR. As Dr. Darcy explained, “We supported GDPR, and last May we did a lot of work towards complying with it. [The GDPR] shifts technology companies toward providing data ownership at the individual level, and we go for that level of privacy because we believe in it.” Dr. Darcy added that, because leaders within Woebot come from human subjects research backgrounds, they are familiar with obtaining “seriously painstaking informed consent.” She added, “Since the company launched in 2017, all user data is limited to the minimum amount required for the service to function. Data is not linked to users in an identifiable way” [13].

Interestingly, Woebot isn’t the only mental health firm in favor of robust privacy. Another CEO interviewed for this article suggested that HIPAA should be extended to cover mental health data collected by corporations. “[I]f we can get a policy that says digital biomarkers are protected health information, then companies like Facebook and Google which could collect this stuff in the past suddenly couldn’t do that anymore. You’d need to treat the data in a manner consistent with HIPAA.”

Another problem that could undermine public confidence in therapy chatbots is lack of demonstrated efficacy. As discussed earlier, Woebot and other services like it are not required to obtain FDA clearance because of the FDA’s hands-off approach. (As mentioned, if the FDA decides to be more involved in the future, it will likely focus on technologies that make clear claims regarding particular clinical indications – so-called digital therapeutics.) For now, though, consumers have no guarantee that these apps will deliver the results they promise. Additionally, a mental health app that is poorly designed could theoretically harm someone struggling with a mental health disorder.

Regulation isn’t the only way to address this problem, but it might be the most effective. Peer-reviewed studies like the one cited earlier, for instance, could help consumers believe that an app works, but consumers are unlikely to search for and read scientific literature. *Psyberguide*, also discussed earlier in this chapter, helpfully provides a review system [8]. Ultimately, though, it cannot block a deficient app from being available to the public. Companies that administer app stores, such as Apple and Google, do have this kind of gatekeeping power. Experts interviewed for this article suggested that app stores are not the best solution, however. As the CEO of a Silicon Valley mental health startup commented, “You wouldn’t want Walgreens to put a stamp of approval on drugs, so why would you want Apple to be making decisions about the efficacy and safety of mental health interventions?”

Dr. Darcy suggests that more regulation could be helpful in this area: “I would love to see more regulation, honestly, because the fact is, there hasn’t been any. If health and wellness apps are quickly adopted but lack any support for efficacy, that threatens the public’s confidence.” Of course, a lengthy FDA review process would seem to run counter to the fluid, ongoing process of product improvement that therapy chatbots are based on. To balance the need for regulation with the realities of innovation, Dr. Darcy suggests a new approval process that focuses on companies

rather than products – a process that is already in play at the FDA. “I love the idea of shifting away from approving products to approving companies. Allowing them the freedom to innovate quickly. The government might say, ‘you have to show us how your company operates, how your QA systems are carried out, and, and so forth.’”

### 3.8 CONCLUSION

The emerging body of scientific literature on therapy chatbots suggests that these services could help large numbers of people reduce their levels of anxiety and depression. This possibility seems important for all of the reasons that these mental health disorders are so harmful: Anxiety can drain the enjoyment from life, interfere with personal relationships, as well as their productivity. Moreover, the problem only seems to be growing more widespread.

Therapy chatbots also may offer some unique advantages to traditional therapy. They can serve millions of people in a cost-effective way, for example. Because they can be accessed within one’s home, they don’t involve the stigma that some people associate with seeking mental health care. They can be accessed by people who cannot easily travel to a clinic. Finally, although these tools are not designed to replace traditional mental health treatment, they could be a cost-effective supplemental form of help for many. These technologies are still new, however, and some in the field of psychology believe their full potential remains an unknown. As one expert recently stated in a newspaper interview, “These things can work well on a superficial level with superficial conversations, but are they effective tools, do they change outcomes, and do they deliver more efficient care? It’s still early” (Nutt, 2017).

Ultimately, the success of therapy chatbots may not only depend on how well they can be engineered but how effectively they can manage pools of private user data. As discussed in the introduction to this chapter, chatbots suggest a promising new technique for aggregating and making use of large sets of mental health. By conditioning access to the benefits of the existing pool – for example, knowledge – on new interactions – for example, typed exchanges – they effectively establish an incentive that might not otherwise exist. By delivering aggregated knowledge through an interactive system that is tightly scripted – rather than, say, a database – chatbots can effectively obscure private user data. This privacy is reinforced by other measures the company has taken, such as encryption.

As effective as this approach seems, the Knowledge Commons Framework shows why privacy remains an important area of work for the company. In light of the stigma that some people feel regarding mental health disorders, there is a need for users to feel that their data will not be disclosed. The degree of privacy the law ensures, meanwhile, is relatively low. HIPAA doesn’t appear to cover chatbots billed as wellness apps, and the FDA has not sought to regulate data security in these types of apps tightly. Meanwhile, high-profile hacks and disclosures at companies like

Facebook have made the public aware that companies have an enormous financial impetus to use private data to make money. As a result, the app's creators and other companies like it must rely on public messaging. This has taken a variety of forms, including pledges not to use user data for advertising, corporate advocacy in favor of more robust privacy laws, and frank statements to the press that professional codes of conduct limit the behavior of corporate leaders. Ultimately, changes to the law combined with ancillary efforts like these may lead to a future where therapy chatbots can help people who suffer from anxiety and reveal new insights about the disorder itself.

#### LINKS

- [1] <https://adaa.org/about-adaa/press-room/facts-statistics>
- [2] [www.apa.org/monitor/2016/06/numbers](http://www.apa.org/monitor/2016/06/numbers)
- [3] [www.nhs.uk/Conditions/stress-anxiety-depression/Pages/benefits-of-talking-therapy.aspx](http://www.nhs.uk/Conditions/stress-anxiety-depression/Pages/benefits-of-talking-therapy.aspx)
- [4] <https://adaa.org/finding-help/transcranial-magnetic-stimulation>
- [5] <https://adaa.org/finding-help/treatment/complementary-alternative-treatment>
- [6] <https://nccih.nih.gov/health/tips/stress>
- [7] [www.ableto.com/news/ableto-acquires-joyable-2019/](http://www.ableto.com/news/ableto-acquires-joyable-2019/)
- [8] <https://psyberguide.org>
- [9] <https://adaa.org/social-media-obsession>
- [10] [www.youtube.com/watch?v=AE966uR0ges](http://www.youtube.com/watch?v=AE966uR0ges)
- [11] <https://woebot.io/privacy>
- [12] <https://woebot.io/blog/why-well-never-sell-your-data-to-advertisers/>
- [13] <https://woebot.io/blog/what-is-gdpr/>
- [14] [www.x2ai.com/outcomes](http://www.x2ai.com/outcomes)

#### REFERENCES

- Andrilla, C. Holly A., Davis G. Patterson, Lisa A. Garbersen, Cynthia Coulthard, and Eric H. Larson. "Geographic Variation in the Supply of Selected Behavioral Health Providers." *American Journal of Preventive Medicine* 54, no. 6 (2018): S199–S207.
- Carey, Benedict. "California tests a digital 'fire alarm' for mental distress." *The New York Times*, June 17, 2019. [www.nytimes.com/2019/06/17/health/mindstrong-mental-health-app.html](http://www.nytimes.com/2019/06/17/health/mindstrong-mental-health-app.html)
- Chin, Monica. "This AI therapist is helping me get over my ex." *Mashable*, February 22, 2018. <https://mashable.com/2018/02/22/worlds-first-ai-therapist/>
- Condliffe, Jamie. "How machine learning may help tackle depression." *MIT Technology Review*, April 5, 2017. [www.technologyreview.com/s/604075/how-machine-learning-may-help-tackle-depression/](http://www.technologyreview.com/s/604075/how-machine-learning-may-help-tackle-depression/)
- De Salvo, Karen B. and Jocelyn Samuels. "Examining oversight of the privacy & security of health data collected by entities not regulated by HIPAA." *Health IT Buzz*, July 19, 2016. [www.healthit.gov/buzz-blog/privacy-and-security-of-ehrs/examining-oversight-privacy-security-health-data-collected-entities-not-regulated-hipaa](http://www.healthit.gov/buzz-blog/privacy-and-security-of-ehrs/examining-oversight-privacy-security-health-data-collected-entities-not-regulated-hipaa).
- Department of Health and Human Services. Examining oversight of the privacy & security of health data collected by entities not regulated by HIPAA." 2016. [www.healthit.gov/sites/default/files/non-covered\\_entities\\_report\\_june\\_17\\_2016.pdf](http://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf)

- Dobbs, David. "The smartphone psychiatrist." *The Atlantic*, July/August 2017. [www.theatlantic.com/magazine/archive/2017/07/the-smartphone-psychiatrist/528726/](http://www.theatlantic.com/magazine/archive/2017/07/the-smartphone-psychiatrist/528726/)
- Dreyfuss, Emily. "LGBTQ+ youth prefer to seek mental health help digitally." *Wired*, June 11, 2019. [www.wired.com/story/lgbtq-mental-health-digital-outreach/](http://www.wired.com/story/lgbtq-mental-health-digital-outreach/)
- DuPont, Robert L., Dorothy P. Rice, Leonard S. Miller, Sarah S. Shiraki, Clayton R. Rowland, and Henrick J. Harwood. "Economic Costs of Anxiety Disorders." *Anxiety* 2, no. 4 (1996): 167–172.
- Fitzpatrick, Kathleen Kara, Alison Darcy, and Molly Vierhile. "Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial." *JMIR mental health* 4, no. 2 (2017): e19.
- Frischmann, Brett M., Michael J. Madison, and Katherine J. Strandburg, eds. *Governing Knowledge Commons*. New York: Oxford University Press, 2014.
- Fulmer, Russell, Angela Joerin, Breanna Gentile, Lysanne Lakerink, and Michiel Rauws. "Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial." *JMIR mental health* 5, no. 4 (2018): e64.
- Garg, Parie and Sam Glick. "AI's potential to diagnose and treat mental illness." *Harvard Business Review*, October 22, 2018. <https://hbr.org/2018/10/ais-potential-to-diagnose-and-treat-mental-illness>
- Ghinassi, Cheryl Winning. *Anxiety*. Biographies of Disease Series. Santa Barbara, CA: Greenwood Publishing Group, 2010.
- Gionet, Kylie. "Meet Tess: the mental health chatbot that thinks like a therapist." *The Guardian*, April 25, 2018. [www.theguardian.com/society/2018/apr/25/meet-tess-the-mental-health-chatbot-that-thinks-like-a-therapist](http://www.theguardian.com/society/2018/apr/25/meet-tess-the-mental-health-chatbot-that-thinks-like-a-therapist)
- Gratch, Jonathan. "The computer will see you now." *The Economist*, August 20, 2014. [www.economist.com/science-and-technology/2014/08/20/the-computer-will-see-you-now](http://www.economist.com/science-and-technology/2014/08/20/the-computer-will-see-you-now)
- Green, Eric P., Nicholas Pearson, Sathyanath Rajasekharan, Michiel Rauws, Angela Joerin, Edith Kwobah, Christine Musyimi, Chaya Bhat, Rachel M. Jones, and Yihuan Lai. "Expanding Access to Depression Treatment in Kenya Through Automated Psychological Support: Protocol for a Single-Case Experimental Design Pilot Study." *JMIR research protocols* 8, no. 4 (2019): e11800.
- Health Resources and Services Administration/National Center for Health Workforce Analysis; Substance Abuse and Mental Health Services Administration/Office of Policy, Planning, and Innovation. 2015. National Projections of Supply and Demand for Behavioral Health Practitioners: 2013–2025. Rockville, Maryland.
- Hilty, Donald M. and Davor Mucic. "Technology, Health, and Contemporary Practice: How Does E-Mental Health Fit and What Does It Offer?" In *e-Mental Health*, Davor Mucic and Donald M. Hilty, eds., Cham: Springer, 2016, 3–27.
- Hernandez, Daniela. "Meet the chatbots providing mental health care." *The Wall Street Journal*, August 9, 2018. [www.wsj.com/articles/meet-the-chatbots-providing-mental-healthcare-1533828373](http://www.wsj.com/articles/meet-the-chatbots-providing-mental-healthcare-1533828373)
- Hoge, Elizabeth, David Bickham, and Joanne Cantor. "Digital Media, Anxiety, and Depression in Children." *Pediatrics* 140, no. Supplement 2 (2017): S76–S80.
- Horwitz, Allan V. *Anxiety: A Short History*. Baltimore, MD: JHU Press, 2013.
- Joerin, Angela, Michiel Rauws, and Mary Lou Ackerman. "Psychological Artificial Intelligence Service, Tess: Delivering On-demand Support to Patients and Their Caregivers: Technical Report." *Cureus* 11, no. 1 (2019): e3972.

- Kleisiaris, Christos F., Chrisanthos Sfakianakis, and Ioanna V. Papatthanasiou. "Health care practices in ancient Greece: The Hippocratic ideal." *Journal of Medical Ethics and History of Medicine* 7 (2014): 6.
- Knight, Will. "Andrew Ng has a chatbot that can help with depression." *Technology Review*, October 18, 2017. [www.technologyreview.com/s/609142/andrew-ng-has-a-chatbot-that-can-help-with-depression/](http://www.technologyreview.com/s/609142/andrew-ng-has-a-chatbot-that-can-help-with-depression/)
- Lee, Debra. "Alison Darcy." *New Learning Times*, July 9, 2018. <https://newlearningtimes.com/cms/article/5417/alison-darcy>
- Lee, Theodore T. "Recommendations for Regulating Software-Based Medical Treatments: Learning from Therapies for Psychiatric Conditions." *Food & Drug LJ* 73 (2018): 66.
- Mathotaarachchi, Sulantha, Tharick A. Pascoal, Monica Shin, Andrea L. Benedet, Min Su Kang, Thomas Beaudry, Vladimir S. Fonov, Serge Gauthier, Pedro Rosa-Neto, and Alzheimer's Disease Neuroimaging Initiative. "Identifying Incipient Dementia Individuals Using Machine Learning and Amyloid Imaging." *Neurobiology of aging* 59 (2017): 80–90.
- McCaney, Kevin. "AI programs can help in early detection of mental health issues." *Government CIO Media and Research*, January 24, 2019. [www.governmentciomedia.com/ai-programs-can-help-early-detection-mental-health-issues](http://www.governmentciomedia.com/ai-programs-can-help-early-detection-mental-health-issues)
- Molteni, Megan. "The chatbot therapist will see you now." *Wired*, June 7, 2017. [www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist/](http://www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist/)
- Morris, Robert R., Kareem Kouddous, Rohan Kshirsagar, and Stephen M. Schueller. "Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions." *Journal of Medical Internet Research* 20, no. 6 (2018): e10148.
- Mucic, Davor, Donald M. Hilty, Michelle B. Parish, and Peter M. Yellowlees. "Web- and Internet-Based Services: Education, Support, Self-Care, and Formal Treatment Approaches." In *e-Mental Health*, Davor Mucic and Donald M. Hilty, eds., Cham: Springer, 2016, 173–192.
- Nutt, Amy Ellis. "The Woebot will see you now – the rise of chatbot therapy." *Washington Post*, December 3, 2017. [www.washingtonpost.com/news/to-your-health/wp/2017/12/03/the-woebot-will-see-you-now-the-rise-of-chatbot-therapy/](http://www.washingtonpost.com/news/to-your-health/wp/2017/12/03/the-woebot-will-see-you-now-the-rise-of-chatbot-therapy/)
- Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press, 1990.
- Pantic, Igor. "Online Social Networking and Mental Health." *Cyberpsychology, Behavior, and Social Networking* 17, no. 10 (2014): 652–657.
- Rachman, Stanley J. *Anxiety*. London: Psychology Press, 2020.
- Rieland, Randy. "Can artificial intelligence detect depression in a person's voice?" *Smithsonian Magazine*, November 5, 2018. [www.smithsonianmag.com/innovation/can-artificial-intelligence-detect-depression-in-persons-voice-180970702/](http://www.smithsonianmag.com/innovation/can-artificial-intelligence-detect-depression-in-persons-voice-180970702/)
- Rosso, Cami. "Is the robot psychologist the next big AI app?" *Psychology Today*, October 17, 2018. [www.psychologytoday.com/intl/blog/the-future-brain/201810/is-the-robot-psychologist-the-next-big-ai-app](http://www.psychologytoday.com/intl/blog/the-future-brain/201810/is-the-robot-psychologist-the-next-big-ai-app)
- Shoemaker, Erica Z. and Donald M. Hilty. "e-Mental Health Improves Access to Care, Facilitates Early Intervention, and Provides Evidence-Based Treatments at a Distance." In *e-Mental Health*, Davor Mucic and Donald M. Hilty, eds., Cham: Springer, 2016, 43–57.
- Stephens, Taylor N., Angela Joerin, Michiel Rauws, and Lloyd N. Werk. "Feasibility of Pediatric Obesity and Prediabetes Treatment Support through Tess, the AI Behavioral Coaching Chatbot." *Translational Behavioral Medicine* 9, no. 3 (2019): 440–447.

- U.S. Department of Health and Human Services. "The U.S. health workforce chartbook: part IV: behavioral and allied health." September 2018. <https://bhw.hrsa.gov/sites/default/files/bhw/health-workforce-analysis/research/hrsa-us-health-workforce-chartbook-part-4-behavioral-and-allied-health.pdf>
- Wallach, Eric. "An interview with Jo Aggarwal, co-inventor of Wysa." *The Politic*, March 28, 2018. <http://thepolitic.org/an-interview-with-jo-aggarwal-co-inventor-of-wysa/>
- Wu, Tim. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Vintage, 2017.