Minding the AI: Ethical Challenges and Practice for AI Mental Health Care Tools

8

Nicole Martinez-Martin

8.1 Introduction

The use of artificial intelligence (AI) for mental health applications raises questions regarding the potential impact on fiduciary obligations in the therapeutic relationship, oversight, bias, and data protection. Health technologies utilizing AI present particular challenges for regulation. AI technologies that address mental and behavioral health may be used in different domains, from healthcare to education and consumer uses, and in some domains, there are not regulations or practices that provide protections for users' health data. There are also ways that bias can enter into AI tools, such as during the data collection and preparation stages. It is therefore necessary to consider how to utilize AI for mental health applications so that the resulting tools do not reflect and reinforce existing social problems and inequality. At the same time, AI can present opportunities for addressing existing inequalities and discrimination in mental health care. There is the potential for misuse of data and health information gathered from individual users, and users may not be sufficiently aware of negative repercussions from sharing their data. Finally, AI tools will likely impact the fiduciary obligations generally expected in the therapeutic relationship and it will be necessary to carefully consider likely areas of concern in order to prepare processes for integrating these tools appropriately into mental health care. This article chapter will engage emerging recommendations for best practices in this area, along with areas for empirical ethics research.

N. Martinez-Martin (✉)
Stanford Center for Biomedical Ethics, Stanford, CA, USA
e-mail: nicolemz@stanford.edu

F. Jotterand, M. Ienca (eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*, Advances in Neuroethics,
https://doi.org/10.1007/978-3-030-74188-4_8

8.2 Artificial Intelligence

"Artificial intelligence" generally refers to the use of machines to perform tasks that resemble cognitive functions that we associate with human intelligence, such as learning or solving problems [1]. Artificial intelligence (AI) can take different forms, as software or hardware, including intelligent autonomous agents, distributed networks, or robotics [2]. Although the term "machine learning" (ML) is sometimes used interchangeably with AI, ML more specifically applies to approaches that train computers to "learn"—recognizing patterns in massive datasets, including complex data interactions, and in the algorithms that are used within AI applications [3]. Machine learning has been used for data mining, image recognition, natural language programming, statistical learning methods, and neural networks, among other applications [4]. The ability of ML to detect patterns and connections that the humans programming the model would not have necessarily known to look for can bring significant benefits to scientific research. For example, ML can be used in order to analyze large quantities of data, such as electronic health records, in order to detect patterns and associations that may be relevant to patient health and outcomes [5]. These patterns can, in turn, be used for the purpose of predictive analytics and decision-making models, in ways that outperform traditional clinical prediction models [6]. ML has also been used to examine social media and websites in order to determine patterns in health-related behaviors [7]. ML and neural networks have been applied to constructing expert systems and clinical decision support systems, which are systems meant to provide and supplement the type of knowledge and skills generally supplied by human experts [8]. By incorporating ML, clinical decision support systems can provide recommendations without needing preprogrammed knowledge. As will be discussed more below, while the benefits of using ML to analyze massive datasets are considerable, the reasons or reasoning underlying the output of ML can be difficult to scrutinize, sometimes even for those who set up the ML system [9]. This is one reason that ML can present a challenge for regulation and oversight.

Natural language processing (NLP) is a subfield of AI that has a number of applications in mental health care [10]. NLP uses computational techniques to examine and classify language. NLP can be used for analysis of social media or vocal data to identify patterns relevant to mental health and behavior [11]. For example, NLP can be used as part of scanning clinical text for identifying symptoms of severe mental illness [12, 13], or for providing and analyzing psychotherapy encounters [14]. NLP can also be used to construct chatbots or virtual humans who can interact with people through text or voice [15]. AI techniques, such as ML and NP, have also been used for virtual reality and augmented reality technologies in order to make the virtual environments more engaging and interactive for the participant [16]. In looking at the different types of AI, one can note features that contribute to the challenging aspects of ethical applications of AI for mental health. The use of massive data sets presents areas of tension with data protection and privacy. The difficulty in knowing the reasoning or potential for bias in algorithms generated by ML can make evaluation of these technologies more difficult. Furthermore, when it comes

to chatbots and VR in particular, AI-generated mental health technologies will have an impact on the therapeutic relationship in ways that go beyond traditional healthcare tools. Some of that impact could be positive, such as providing useful options for patients who may prefer sharing their feelings and information with a nonhuman. Potential negative repercussions include insufficient data protection or lack of clarity regarding liability for mistakes, which can also undermine trust overall in AI approaches to mental health care. The impact on the therapeutic relationship will need to be studied in order to better understand the benefits and burdens of AI mental health tools.

8.3 AI Mental Health Applications

AI is being integrated into a number of technologies for mental health care, from computing methods that can use massive data sets to assist in clinical decision-making, diagnosis, and treatment, to apps and wearables that can be used by patients and consumers for mental health assistance, and public health applications that assist in identifying behavioral health risks and solutions [17]. Some applications of AI involve boosting the capabilities of existing techniques and treatments, such as utilizing AI for deep brain stimulation approaches that respond dynamically to the needs of the patient [18]. It should be noted that the contexts for these different applications influence the types of ethical challenges encountered for that use. In the USA, for example, there are statutes that provide some protection for health information; however, these statutes and regulations generally apply to health information generated within the context of healthcare institutions and healthcare providers [19]. Even though some consumer AI applications can generate information about a person's mental health or behavior, that information may not have the same privacy protections that health information in the healthcare domain would be afforded [20]. Thus, as AI is being increasingly applied to mental health, it is important to note that many of these applications may be used in domains to which different privacy and user protection concerns are relevant, such as healthcare, consumer, or government institutions.

AI tools are being incorporated in the construction of expert systems [21, 22]. In clinical contexts, AI-informed expert systems may be used for such purposes as suggesting appropriate medications for a patient [23]. Predictive analytics are being increasingly utilized in healthcare environments, with AI often utilized for analyzing the data [24] These expert systems have traditionally been used in order to derive clinical rules or recommendations from the large amounts of data available in health systems, but, with the advent of more sophisticated AI, have become more focused on assisting with choices of differing probabilistic pathways [23]. Some have raised concerns that these decision-making tools will eventually replace the role of physicians, but the general goal is to make clinicians more effective with these tools. Providing sufficient training and support so that clinicians can utilize the information and findings provided by these AI-enhanced tools effectively remains a challenge [25]. In developing these decision-making tools, it is important

to take account for how they may be influenced and affected by the context in which they are placed. In other words, the treatment decisions that are recommended by an AI tool will, in turn, impact the clinical environment, thus becoming another factor that will need to be accounted for in the analyses performed by the AI tools [26]. It is therefore important to carefully consider how the expert system will be implemented, so that it can be appropriately aligned with its environment and stakeholders.

AI has also become useful for development of technologies that provide simulations for therapeutic purposes. Autonomous conversational agents can be used to engage with a person, respond to text or vocal queries, and even provide some aspects of therapeutic interactions [16]. Chatbots can be used to respond to basic text queries regarding mental health needs in order to inform or direct the user to resources or services, and also for more complex interactions meant to provide aspects of therapy [27]. For example, Woebot is a conversational agent meant to address people with depression that incorporates tools drawn from cognitive behavioral therapy and can assist in monitoring mood, find learning videos and resources, and walk the user through "self-directed" therapy [28]. There is also an increasing role for robotics technology, incorporating AI, for mental health purposes. Robotics can be useful in cases where there may not be a person who can fill the role, such as robotics that can serve a companionship and support role (e.g., assisting users with getting exercise) for a patient [29]. Robots may be particularly useful in cases where the user may have reasons to prefer not to interact with a human for the therapy service. For example, robots have shown promise in assisting people on the autism spectrum develop skills, such as play [30] or social interaction [31]. With both chatbots and robotic technology, one of the ethical challenges relates to the possibility of blurred boundaries in user interactions with the bot, where users may lose sight of the fact that they are sharing information with a technology that can collect and pass along that data. It will be key to ensure that users are adequately informed about how privacy and confidentiality apply to the interactions, and how the design of the technologies may be used to address these types of concerns (such as switches or signals to the user when information is being recorded) [32].

Virtual reality (VR) is a technology that allows a user to experience a computer-generated simulated environment and interact with virtual persons or beings in that environment [33]. VR has become a tool for addressing a variety of mental health concerns, from use in PTSD treatments to assisting with diagnosis [34]. VR can also be used as a way to provide a virtual therapy space for real-time therapeutic interactions [35]. Augmented reality (AR) refers to the combination of VR with the world around someone by placing computer-generated images into the live video. AR has been used to help train mental health clinicians, remind psychiatric patients to take medications, and assist children who have autism learn to recognize facial emotions [33].

Mobile mental health applications also have been incorporating features through the use of AI. "Digital phenotyping" is a term commonly used to refer to approaches in which smartphones and mobile sensors are used to gather personal data from users, which is then analyzed in order to assess the user's cognitive and mental state, as well as make predictions [36, 37]. The data collected could be physiological

functions, such as pulse, location information, tapping and keyboard interactions, or voice features [38]. Some approaches to digital phenotyping include analysis of social media posts and other internet use in order to assess behavioral health risks [39]. For clinical uses, the user would generally be asked to give informed consent and download an app onto their phone, which would passively collect the relevant personal data as the user goes about their usual daily activities. Beyond clinical usage, there are a range of institutions and organizations that may utilize digital phenotyping tools, such as educational institutions interested in assessing risk of suicide or of a student dropping out, the military assessing behavioral risks of recruits, insurance companies using such tools to set rates, employers, or consumer digital phenotyping for marketing purposes [40, 41].

As noted above, ethical concerns will differ depending upon the context of the application (e.g., different regulations and guidelines for data protection generally apply in healthcare contexts as opposed to consumer contexts). For uses that take place outside of healthcare, it is particularly important to examine the potential repercussions of inferences that can be drawn from an individual's personal data [42].

8.4 Ethical Challenges

Ethical challenges related to safety, effectiveness, or privacy are familiar areas of concern for new health technologies. Of course, AI tools in mental health care will raise varying ethical concerns according to their function. A conversational agent, for example, will likely raise concerns regarding how users interact with it therapeutically, that are different than concerns regarding how predictive analytics impact mental health care. Generally speaking, AI has some features that can pose difficulties for the traditional frameworks for addressing such ethical issues. The use of ML to generate algorithms, which puts the "reasoning" behind decisions into a proverbial "black box" can make it particularly challenging to examine and review the reasons behind the outputs generated by the algorithms. Thus agencies, such as the FDA, which is responsible for oversight of medical devices in the USA, have had to consider how to appropriately evaluate the accuracy and applications of AI technologies [43, 44]. A second issue, the use of these technologies in domains outside of healthcare, also impacts accountability and oversight. Information gathered in a healthcare setting would generally need to follow HIPAA privacy protections and involve informed consent procedures, which include protecting health and identifying information [45]. Digital phenotyping tools that could generate information and predictions about behavior and mental health, but are for consumer use, generally have fewer protections for user data or need for informed consent, often confined to notice about data practices on associated "terms and conditions" page. In some cases, the terms and conditions are misleading, not letting know the companies who may be receiving the data [46].

In the current big data environment, information that previously might be considered mundane or uninteresting, such as a grocery purchase or location at a particular

moment, can be combined with other information and be transformed into health information [47]. Yet the paradigm for protection of health information is still based on traditional frameworks in which healthcare institutions and physicians are envisioned as the main domain for healthcare information [47]. Moreover, the massive amounts of data and techniques used for ML are often characterized as providing more objective results, but need to be carefully scrutinized for ways that bias may enter into the findings [48, 49]. Finally, while some argue that AI tools should just be seen as the same kind of device as any previous methods of assessing health risks, there are indications that people may regard AI tools as more objective than the human clinician or even as a third party involved in the clinical interaction [50]. For that reason, AI tools will likely influence the therapeutic relationship [51]. Because many ethical obligations are rooted in the therapeutic relationship, it is important to empirically study how AI impacts the therapeutic relationship in order to address any repercussions for associated ethical duties.

8.5 Therapeutic Relationship

The therapeutic relationship or alliance refers to the relationship that develops between a patient and the mental health care provider in order to achieve the goals for the patient [52]. In mental health care, the therapeutic relationship can involve the patient providing sensitive and emotionally charged personal information. The mental health care provider has professional obligations to protect the patient from harm and provide a foundation for achieving desired treatment outcomes [53]. For this reason, ethical values such as trust and confidentiality are key to the therapeutic relationship [54]. When it comes to the use of AI technologies for mental health care, there are many questions that may impact the therapeutic relationship. How might continuous monitoring affect trust? How do clinicians manage the massive amounts of data in order to extract meaningful information and communicate it to patients? Is the technology experienced as a "third party" in the clinical relationship? How will clinicians evaluate and incorporate findings from AI tools into their professional judgment and how patients will respond in terms of perceived stigma or bias in the predictions? There will be a need for empirical research to investigate the impact on trust in the clinician or digital tool, as well as how physicians rely and communicate health information and how patients view the competence of physicians and devices. Elderly and people with severe mental illness may face particular challenges in understanding the risks and benefits of using AI technology, or have different views regarding prioritizing ethical values, such as privacy [55]. When it comes to AI technologies such as conversational agents or robots that are used to interact with patients, designers and healthcare obligations must consider how the devices will affect these ethical obligations associated with the therapeutic relationship. Conversely, when these applications are used outside of healthcare institutions, are there ethical obligations generally found in the therapeutic relationship that should be addressed—for example, if a website analyzes its users' behavior, are there any duties to warn or direct users to resources that should be instituted [56].

For clinical use of these tools, organizations such as the American Psychiatric Association have been proactive in trying to establish recommendations for appropriate integration of these tools into clinical practice [57].

8.6 Safety and Effectiveness

Oversight for safety and efficacy of health technologies utilizing AI is still evolving. Regulation of health devices based on machine learning presents challenge because the reasons for particular results or findings may not be accessible for evaluation. In the USA, medical devices that utilize AI are subject to regulation by the Federal Drug Administration (FDA). The FDA has made significant efforts in recent years to establish effective approaches to regulate digital health technologies, including those that incorporate AI. The FDA has announced a Digital Health Program and a Pre-certification Program for manufacturers, which involves a shift from a product-based approach to a more process-based approach and does not address the issue of evaluating specific machine learning devices [58]. Professional organizations for computer science and AI have also discussed the need for designing AI systems that include mechanisms for a clinician or other user to receive more explanation of the bases of the results or findings that they have received [59].

Going forward, one significant issue for clinical applications of AI will be embedding established clinical standards in the ways that the tool is designed and used. ML approaches require large datasets and population sizes in order to produce validated models for expert systems and predictive analytics, and so issues of data sharing are important to consider. As systems and tools based on AI are increasingly integrated into healthcare, professionals will need to consider what the appropriate applications of AI for mental health care are, as well as the scope and limitations of the systems. In particular, interdisciplinary collaboration is needed for assessing if, when, and how AI applications are implemented, and different end users (clinicians, healthcare administrators, or patients) should be included in the development process in order to support ethical design and use of these tools [60]. As AI-based systems and tools are placed into different contexts and used among different populations, professionals using the system will need information on how the tools may best be used among different populations. Systems may need safeguards in place in order to ensure that the technologies are being used in the manner and for the population in which they have been validated. Of course, for technologies such as mental health apps or digital phenotyping, that may be used outside of healthcare institutions or, particularly, by consumers, it can be more difficult to establish lines of accountability and oversight that can ensure appropriate understanding and scope of use of the tools. In those instances, regulations that protect user data and require more robust user consent can help to inform users and require consent for use of their data.

Accountability for AI systems also involves questions regarding which entities are responsible for monitoring how the systems are functioning and being used, as well as liabilities for problems. If an AI tool causes harm or is not working as

expected in a particular context, who is responsible for reporting and to whom they report? Furthermore, there will need to be consideration of how technologies such as expert systems or digital phenotyping may need to capability of monitoring risks of harm to patients or other users. In mental health contexts, where patients may disclose information that indicates a potential to harm self or others, how should tools monitor and assess such information and to whom will they need to report? These questions have come up in relation to conversational agents, in terms of whether these agents need to be programmed to provide resources or alert others if suicide risk is found [61]. Digital phenotyping is an area where, even if there is not direct disclosure from the patient about harming self or others, inferences could potentially be drawn from user data that leads to a prediction of harm [62]. Design of such tools need to incorporate consideration of monitoring and reporting potential harms, and institutions utilizing such systems need plans about how predictive tools and monitoring of potential for harm will be undertaken. Depending upon the jurisdiction, laws regarding duty-to-warn and other requirements will need to be taken into account.

8.7 Bias/Fairness

An important issue related to effectiveness and scope of use is methods for addressing the potential for bias in ML tools. The potential for bias can be viewed in terms of the potential for bias in the data used to construct the algorithms and bias in the algorithms themselves, as well as the potential for bias in how the algorithms may be used within a particular local context [63]. Because massive datasets are used in order to train ML systems to identify patterns in the data, the accuracy of the resulting algorithm depends on the quality of data in those training and validation sets [64]. Furthermore, if the dataset do not accurately reflect the population to the technology will be applied, then the bias in the data will be seen in the outcomes generated by the ML algorithm [65]. Thus ML systems could unfortunately both reflect and reinforce biases that are found in society. In terms of mental health applications of AI, social factors such as race, gender, and class can influence many aspects of mental health diagnosis outcomes. If the data used to generate an algorithm does not contain a representative sample, then the findings of the algorithm can be skewed. One of the reasons that it can be important to design "explainability" of the algorithm's reasoning into a tool is that algorithms may not have sufficient information (beyond the issue of a representative sample) to take into account why there may be certain associations between social factors and a particular mental health outcome. For example, an algorithm used for criminal justice sentencing may make an association between race and recidivism, but not have the data to take into account the impact of existing racial biases on recidivism [66]. These kinds of issues can not only limit the benefits that people from underrepresented racial and ethnic populations may receive from AI tools, but can exacerbate discrimination against particular groups. Efforts to increase the diversity of populations in datasets used for ML mental health research are critical. Professional organizations, such as the Institute

of Electrical and Electronics Engineers (IEEE), have been conducting efforts to formulate recommendations and methods for reducing bias in ML algorithms [67, 68]. One important aspect is to include input from stakeholders for stakeholders, in order to provide feedback from clinicians and mental health consumers that can inform efforts to reduce bias. In implementation of ML-informed tools and systems, some institutions have also taken an approach to create an impact assessment of the tool beforehand, so that a plan can be developed and implemented to inform relevant stakeholders of potential impacts of the tool and make efforts to minimize that impact [69]. There needs to be reflection on the ethical implications of potential AI applications in mental health throughout the stages of development. As early as the stage formulation of the question or goal of the AI application, reflection on the ethical issues may be needed, because algorithms designed to identify psychiatric genetic risk for purposes or decide on allocation of healthcare resources can potentially raise ethical challenges regarding discrimination. In domains such as insurance or employment, there is also potential for discrimination in the construction of algorithms. In the USA, because laws regarding discrimination often rely on finding discriminatory intent, it may be more difficult to address such algorithmic discrimination through the courts [70]. There may be a need to consider regulations that would make certain types of discrimination based on behavioral predictions unlawful.

8.8 Privacy/Trust

Privacy and data protection have been identified as particularly important issues when it comes to big data approaches and AI technologies. In the mental health context, an important issue is that for some AI technologies, such as digital phenotyping, data may be collected in ways that individuals may not ordinarily associate with health information or even as sensitive data (such as speed of typing or tapping patterns on digital devices). Data may be collected outside of contexts in which healthcare information is protected by existing standards, such as HIPAA. Next, the data may be highly granular, especially in combination. Some data may be de-identified, but individuals may not be aware that, in combination with other data, the risk of identifiability may increase. Currently, patients or mental health consumers may not be aware of the ways that their personal data may be shared or sold to different organizations and companies, or that those companies can generate additional behavioral or health inferences about individuals. The data protection policies of mental health applications can have repercussions for individuals in areas such as employment, insurance, litigation that people may not reasonably have expected. At the same time, privacy concerns need to be balanced against data sharing practices that advance scientific research. In Europe, the General Data Protection Regulation presents a model for stronger data protections, including stricter consent provisions for the collection of data [71]. California has enacted similar provisions in the California Consumer Privacy Act [72]. While these regulations are useful for protecting personal data, the inferences that can be drawn from the data people share

may still pose concern [73]. A reliance on consent as an approach to mitigate data protection concerns can also be problematic if not giving consent means that the user will be barred from using useful services. Stronger consent rules for use of personal data are necessary, particularly in contexts outside of healthcare, and provided at appropriate reading levels. At the same time, a focus on individual consent can overlook the need to include a broader range of people for broader discussion of how data may be ethically collected and the appropriate societal goals in doing so [74].

8.9 Surveillance

Technologies utilizing AI to monitor people's behavior may also have surveillance applications that are ethically challenging [75]. There are a number of institutions and companies that have an interest in monitoring individual behavior and conducting predictive analysis of mental states for a variety of reasons. Recently, the US government had proposed monitoring data from a range of wearables and apps to identify individuals for their potential to conduct mass shootings [76]. People diagnosed with mental illness were mentioned as a particular focus of such monitoring. While there was immediate pushback to this proposal from mental health and privacy advocates, the desire to use AI technologies to monitor people with mental illness for such purposes is not surprising. The use of facial recognition and genetic technologies for surveillance purposes in China has also received attention and criticism, as these technologies have been used to conduct behavioral surveillance, particularly targeting ethnic minority populations [77, 78]. There have been some laws passed on a local level to limit the use of facial recognition technology for surveillance [79], and there is a need to consider whether more regulation is needed. The use of technologies for behavioral and mental health surveillance can undermine the trust that people have in these technologies, use of their data, and the healthcare system. In the consumer domain, the massive collection and brokering of personal data is a part of what has been termed "surveillance capitalism." Beyond the issue of access to personal data, the inferences from these data can be used in attempts to influence and manipulate individuals for marketing and political purposes that raise ethical concerns on a societal level [80].

8.10 Conclusion

As efforts move forward to formulate guidelines and identify solutions to the ethical challenges presented by AI applications in mental health, there is a need for stakeholders with expertise in a range of disciplines, as well as patients and consumers, to come together and provide input. Transparency and informed consent have been commonly identified as goals, particularly in order to address some of the data protection and privacy challenges, in order to educate users and advise them of the potential repercussions of sharing their data. With AI technologies that are used for

identifying and addressing behavioral issues outside of healthcare, ensuring meaningful consent of individuals remains challenging and elusive. Even though transparency and informed consent are important components of ethical use of mental health applications of AI, there remains a need to consider regulation to protect the privacy and safety of consumers, guard against discrimination in relation to predictive technologies, and overall ensure broader discussion and action take place regarding realizing societal as well as individual benefits from behavioral and mental health applications of AI.

References

1. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719. https://doi.org/10.1038/s41551-018-0305-z.
2. Price N. Artificial intelligence in health care: applications and legal issues. The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School. https://petrieflom.law.harvard.edu/resources/article/artificial-intelligence-in-health-care-applications-and-legal-issues. Accessed 2 Mar 2019.
3. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015;16(6):321–32. https://doi.org/10.1038/nrg3920.
4. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry. ArXiv:1705.10553 [Stat]. 2017. http://arxiv.org/abs/1705.10553.
5. Rose S. Machine learning for prediction in electronic health data. JAMA Netw Open. 2018;1(4):e181404. https://doi.org/10.1001/jamanetworkopen.2018.1404.
6. Scalable and accurate deep learning with electronic health records. npj Digital Medicine. n.d. https://www.nature.com/articles/s41746-018-0029-1. Accessed 29 Aug 2019.
7. Hao B, Li L, Li A, Zhu T. Predicting mental health status on social media. In: Rau PLP, editor. Cross-cultural design. Cultural differences in everyday life. Berlin Heidelberg: Springer; 2013. p. 101–10.
8. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20(5):e262–73. https://doi.org/10.1016/S1470-2045(19)30149-4.
9. Mols B. In black box algorithms we trust (or do we?). https://cacm.acm.org/news/214618-in-black-box-algorithms-we-trust-or-do-we/fulltext. Accessed 31 Aug 2019.
10. Price WN. Regulating black-box medicine. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network; 2017. https://papers.ssrn.com/abstract=2938391.
11. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform. 2009;42(5):760–72. https://doi.org/10.1016/j.jbi.2009.08.007.
12. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, Roberts A, Dobson RJ, Stewart R. Natural language processing to extract symptoms of severe mental illness from clinical text: the clinical record interactive search comprehensive data extraction (CRIS-CODE) project. BMJ Open. 2017;7(1):e012012. https://doi.org/10.1136/bmjopen-2016-012012.
13. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid [Research article]. 2016. https://doi.org/10.1155/2016/8708434.
14. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. Trans Assoc Comput Linguist. 2016;4:463–76. https://doi.org/10.1162/tacl_a_00111.
15. Denecke K, May R, Deng Y. Towards emotion-sensitive conversational user interfaces in healthcare applications. Stud Health Technol Inform. 2019;264:1164–8. https://doi.org/10.3233/SHTI190409.

16. Miner A, Chow A, Adler S, Zaitsev I, Tero P, Darcy A, Paepcke A. Conversational agents and mental health: theory-informed assessment of language and affect. In: Proceedings of the fourth international conference on human agent interaction, 123–130. HAI '16. New York, NY: ACM; 2016. https://doi.org/10.1145/2974804.2974820.
17. Luxton DD. Chapter 1—An introduction to artificial intelligence in behavioral and mental health care. In: Luxton DD, editor. Artificial intelligence in behavioral and mental health care; 2016. p. 1–26. https://doi.org/10.1016/B978-0-12-420248-1.00001-5.
18. Patel UK, Anwar A, Saleem S, Malik P, Rasul B, Patel K, et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. J Neurol. 2019; https://doi.org/10.1007/s00415-019-09518-3.
19. Rothstein MA. Health privacy in the electronic age. J Leg Med. 2007;28(4):487–501. https://doi.org/10.1080/01947640701732148.
20. Martinez-Martin N. What are important ethical implications of using facial recognition technology in health care? AMA J Ethics. 2019;21(2):180–7. https://doi.org/10.1001/amajethics.2019.180.
21. Bennett CC, Doub TW. Chapter 2—Expert systems in mental health care: AI applications in decision-making and consultation. In: Luxton DD, editor. Artificial intelligence in behavioral and mental health care; 2016. p. 27–51. https://doi.org/10.1016/B978-0-12-420248-1.00002-7.
22. Masri RY, Jani HM. Employing artificial intelligence techniques in Mental Health Diagnostic Expert System. In: 2012 international conference on computer information science (ICCIS), vol. 1. 2012. p. 495–99. https://doi.org/10.1109/ICCISci.2012.6297296.
23. Singh VK, Shrivastava U, Bouayad L, Padmanabhan B, Ialynytchev A, Schultz SK. Machine learning for psychiatric patient triaging: an investigation of cascading classifiers. J Am Med Inform Assoc JAMIA. 2018;25(11):1481–7. https://doi.org/10.1093/jamia/ocy109.
24. Koh HC, Tan G. Data mining applications in healthcare. J Healthcare Inform Manag JHIM. 2005;19(2):64–72.
25. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med. 2018;15(11):e1002689. https://doi.org/10.1371/journal.pmed.1002689.
26. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med. 2018;378(11):981–3. https://doi.org/10.1056/NEJMp1714229.
27. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248–58. https://doi.org/10.1093/jamia/ocy072.
28. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Mental Health. 2017;4(2):e19.
29. Riek LD. Chapter 8—Robotics technology in mental health care. In: Luxton DD, editor. Artificial intelligence in behavioral and mental health care. San Diego: Academic Press; 2016. p. 185–203. https://doi.org/10.1016/B978-0-12-420248-1.00008-8.
30. Robins B, Dautenhahn K. Tactile interactions with a humanoid robot: novel play scenario implementations with children with autism. Int J Soc Robot. 2014;6(3):397–415. https://doi.org/10.1007/s12369-014-0228-0.
31. Vanderborght B, Simut R, Saldien J, Pop C, Rusu AS, Pintea S, Lefeber D, David DO. Using the social robot Probo as a social story telling agent for children with ASD. Interact Stud. 2012;13(3):348–72. https://doi.org/10.1075/is.13.3.02van.
32. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. JAMA. 2017; https://doi.org/10.1001/jama.2017.14151.
33. Lányi CS. Virtual reality in healthcare. In: Ichalkaranje N, Ichalkaranje A, Jain LC, editors. Intelligent paradigms for assistive and preventive healthcare; 2006. p. 87–116. https://doi.org/10.1007/11418337_3.
34. Virtual reality might be the next big thing for mental health. n.d. Scientific American Blog Network website: https://blogs.scientificamerican.com/observations/virtual-reality-might-be-the-next-big-thing-for-mental-health/. Accessed 20 Aug 2019.

35. Anderson PL, Price M, Edwards SM, Obasaju MA, Schmertz SK, Zimand E, Calamaras MR. Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. J Consult Clin Psychol. 2013;81(5):751–60. https://doi.org/10.1037/a0033559.
36. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA. 2017;318(13):1215–6. https://doi.org/10.1001/jama.2017.11295.
37. Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology. 2016;41(7):1691–6. https://doi.org/10.1038/npp.2016.7.
38. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela J-P. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. Npj Digit Med. 2018;1(1):15. https://doi.org/10.1038/s41746-018-0022-8.
39. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol. 2015;33(5):462–3. https://doi.org/10.1038/nbt.3223.
40. Kantrowitz L. When Facebook and Instagram think you're depressed. 2017. Vice website: https://www.vice.com/en_us/article/pg7d59/when-facebook-and-instagram-thinks-youre-depressed. Accessed 26 Oct 2017.
41. Dans E. The rise of real-time, context-based insurance. n.d. Forbes website: https://www.forbes.com/sites/enriquedans/2017/03/12/the-rise-of-real-time-context-based-insurance/. Accessed 29 Sept 2018.
42. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. Npj Digit Med. 2018;1(1):68. https://doi.org/10.1038/s41746-018-0075-8.
43. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med. 2014;371(4):372–9. https://doi.org/10.1056/NEJMhle1403384.
44. Center for Devices and Radiological Health. Digital Health [WebContent]. n.d. FDA.gov website: https://www.fda.gov/medicaldevices/digitalhealth/. Accessed 20 Feb 2018.
45. Glenn T, Monteith S. Privacy in the digital world: medical and health data outside of HIPAA protections. Curr Psychiatry Rep. 2014;16(11):494. https://doi.org/10.1007/s11920-014-0494-4.
46. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open. 2019;2(4):e192542. https://doi.org/10.1001/jamanetworkopen.2019.2542.
47. Bloss C, Nebeker C, Bietz M, Bae D, Bigby B, Devereaux M, et al. Reimagining human research protections for 21st century science. J Med Internet Res. 2016;18(12):e329. https://doi.org/10.2196/jmir.6634.
48. Danks D, London AJ. Algorithmic bias in autonomous systems. In: Proceedings of the 26th international joint conference on artificial intelligence. 2017. p. 4691–7. http://dl.acm.org/citation.cfm?id=3171837.3171944.
49. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics. 2016;22(2):303–41. https://doi.org/10.1007/s11948-015-9652-2.
50. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA. 2016;316(22):2353–4. https://doi.org/10.1001/jama.2016.17438.
51. Luxton DD. Artificial intelligence in psychological practice: current and future applications and implications. Prof Psychol Res Pract. 2014;45(5):332–9. https://doi.org/10.1037/a0034559.
52. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. Journal of Medical Internet Research. 2012;14(4). https://doi.org/10.2196/jmir.2084.
53. Torous J, Roberts LW. The ethical use of mobile health technology in clinical psychiatry. J Nerv Ment Dis. 2017;205(1):4–8. https://doi.org/10.1097/NMD.0000000000000596.
54. Rendina HJ, Mustanski B. Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the United States. J Med Internet Res. 2018;20(7):e233. https://doi.org/10.2196/jmir.9019.

55. Nebeker C, Lagare T, Takemoto M, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Transl Behav Med. 2016;6(4):577–86. https://doi.org/10.1007/s13142-016-0426-4.
56. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Mental Health. 2018;5(2). https://doi.org/10.2196/mental.9423.
57. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E-Health: Offic J Am Telemed Assoc. 2015;21(12):1038–41. https://doi.org/10.1089/tmj.2015.0002.
58. Center for Devices and Radiological Health. Digital health—digital health software precertification (Pre-Cert) program [WebContent]. n.d. https://www.fda.gov/MedicalDevices/DigitalHealth/UCM567265. Accessed 2 Aug 2018.
59. Koene A. Algorithmic bias: addressing growing concerns [leading edge]. IEEE Technol Soc Mag. 2017;36(2):31–2. https://doi.org/10.1109/MTS.2017.2697080.
60. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff. 2014;33(7):1139–47. https://doi.org/10.1377/hlthaff.2014.0048.
61. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med. 2016;176(5):619–25. https://doi.org/10.1001/jamainternmed.2016.0400.
62. Torous J, Onnela J-P, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry. 2017;7(3):e1053. https://doi.org/10.1038/tp.2017.25.
63. Glymour B, Herington J. Measuring the biases that matter: the ethical and casual foundations for measures of fairness in algorithms. In: Proceedings of the conference on fairness, accountability, and transparency. FAT* '19. Atlanta, GA: Association for Computing Machinery; 2019. p. 269–78. https://doi.org/10.1145/3287560.3287573.
64. Towards trustable machine learning. Nat Biomed Eng. 2018;2(10):709. https://doi.org/10.1038/s41551-018-0315-x.
65. Tunkelang D. Ten things everyone should know about machine learning. n.d. Forbes website: https://www.forbes.com/sites/quora/2017/09/06/ten-things-everyone-should-know-about-machine-learning/. Accessed 13 Jan 2018.
66. Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. Sci Adv. 2018;4(1):eaao5580. https://doi.org/10.1126/sciadv.aao5580.
67. Winfield A, Halverson M. Artificial intelligence and autonomous systems: why principles matter. n.d. IEEE Future Directions website: http://sites.ieee.org/futuredirections/tech-policy-ethics/september-2017/artificial-intelligence-and-autonomous-systems-why-principles-matter/. Accessed 28 Aug 2019.
68. Policy recommendations: control and responsible innovation of artificial intelligence. 2018. The Hastings Center website: https://www.thehastingscenter.org/news/policy-recommendations-control-responsible-innovation-artificial-intelligence/. Accessed 5 Dec 2018.
69. Institute AN. Algorithmic impact assessments: toward accountable automation in public agencies. 2018. Medium website: https://medium.com/@AINowInstitute/algorithmic-impact-assessments-toward-accountable-automation-in-public-agencies-bd9856e6fdde. Accessed 31 Aug 2019.
70. Kleinberg J, Ludwig J, Mullainathan S, Sunstein CR. Discrimination in the age of algorithms. Journal of Legal Analysis. 2018;10. https://doi.org/10.1093/jla/laz001.
71. EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.
72. California Consumer Privacy Act of 2018.
73. Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network; 2019. https://papers.ssrn.com/abstract=3248829.

74. Costanza-Chock S. Design justice: towards an intersectional feminist framework for design theory and practice. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network; 2018. https://papers.ssrn.com/abstract=3189696.
75. Martinez-Martin N, Char D. Surveillance and digital health. Am J Bioeth AJOB. 2018; 18(9):67–8. https://doi.org/10.1080/15265161.2018.1498954.
76. Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI (SSRN Scholarly Paper No. ID 3248829). 2019. Social Science Research Network website: https://papers.ssrn.com/abstract=3248829.
77. Feng E. How China is using facial recognition technology. NPR.Org. n.d. https://www.npr.org/2019/12/16/788597818/how-china-is-using-facial-recognition-technology. Accessed 11 Mar 2020.
78. China uses DNA to map faces, with help from the west. The New York Times. n.d. https://www.nytimes.com/2019/12/03/business/china-dna-uighurs-xinjiang.html. Accessed 11 Mar 2020.
79. Conger K, Fausset R, Kovaleski SF. San Francisco bans facial recognition technology. The New York Times. 2019, May 14. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.
80. Big other: surveillance capitalism and the prospects of an information civilization—Shoshana Zuboff, 2015. n.d. https://journals.sagepub.com/doi/10.1057/jit.2015.5. Accessed 11 Mar 2020.