1

Artificial Intelligence in the Behavioral Health Professions: An Overview

Artificial intelligence (AI) is now prominent in the behavioral health professions, being used in clinical, administrative, advocacy, policy, and educational contexts. Behavioral health practitioners are using AI to provide clinical services and interventions, conduct risk assessments, forecast clinical outcomes, and address potential and actual systemic bias in the delivery of services and employee hiring. Increasing numbers of behavioral health practitioners are using AI to document clinical services, provide training and supervision, and educate students.

These are impressive developments. That said, AI comes with noteworthy ethical challenges and legal risks. Key challenges include issues of informed consent and client autonomy; privacy and confidentiality; transparency; client misdiagnosis; client abandonment; client surveillance; plagiarism, dishonesty, fraud, and misrepresentation; and algorithmic bias and unfairness (Plante, 2023; Reamer, 2023a; Rubeis, 2022; Tambe & Rice, 2018; Terra et al., 2023).

This book has several goals. First, I provide a brief history of AI and explore its diverse uses in behavioral healthcare.* Second, I provide an overview of key ethical issues and challenges associated with behavioral health practitioners' use of AI. Third, I discuss ethics-informed "best practices" associated with behavioral health practitioners' use of AI, including guidelines

* My discussion of AI tools includes references to specific platforms and products. My references to these AI platforms and products are for illustrative purposes; they are not endorsements.

and protocols that practitioners and program administrators can use to ensure the ethical use of AI. I then discuss a wide range of risk management issues associated with practitioners' use of AI, including practical ways to protect clients and prevent malpractice litigation and licensing board complaints.* Throughout the discussion I include case examples to illustrate key concepts.**

HISTORY AND NATURE OF ARTIFICIAL INTELLIGENCE

The term *artificial intelligence* was introduced in 1955 by Stanford University professor John McCarthy (Petrosyan, 2024; Reamer, 2023a). Today, AI draws on computer science and datasets to simulate human intelligence and enable problem solving in a wide variety of settings, including behavioral health. AI includes what is known as *machine learning* (ML), which uses historical data to predict and shape new output. The term *generative AI* refers to the creation of images, videos, audio, text, and 3D models using learning patterns from existing data to generate new content and outputs. According to Copeland (n.d.), AI is the ability of digital computers or a computer-controlled robot to perform tasks commonly associated with intelligent beings.

AI can include what are known as expert systems, natural language processing, speech recognition, and machine vision (i.e., the ability of machines to see, analyze, and act). AI uses algorithms to generate responses to queries and provide guidance and suggestions. In healthcare professions, such as medicine and nursing, AI has been used to diagnose disease, facilitate patient treatment, automate redundant tasks, document clinical encounters, provide customer service, reduce dosage errors, provide robot-assisted services, analyze patient scans, and detect fraud (Rong et al., 2020).

More specifically related to behavioral health, what is known as *affective computing* and *emotion AI* apply this technology to practitioners' efforts to assist struggling individuals and those who support them (Diez, 2023; Luxton, 2016; Pascoe, 2023; Royer, 2021). Key AI options include chatbots,

---

* My discussion of risk management issues includes references to laws (for example, statutes, regulations, case law) and legal concepts related to the use of AI. I am not an attorney and cannot offer legal advice. My advice is limited to ethical and risk management issues. If you feel the need for legal consultation on these and related issues, please consult an attorney licensed in your jurisdiction who has expertise pertaining to the topics addressed in this book.

** In this book, I use pseudonyms in case examples to protect the identities of the clients.

social robots, machine translation, search engines, predictive analytic tools, speech-to-text tools, text recognition, speech generation, image recognition and generation, and research tools to assist people who struggle with mental health, substance use disorders, and other behavioral health challenges.

Chatbots

A *chatbot* is a computer program that simulates human conversation to solve users' queries. When an individual who is struggling with a behavioral health issue reaches out, the chatbot is there to welcome them and address their problems. Some chatbots connect users with a human agent. These conversational agents can hold discussions with users in the form of text, voice, or a combination of both. For example, a chatbot might provide a user who is experiencing depression or anxiety with assessment questions, self-help suggestions, resources, and referrals. Chatbots use *natural language processing*, which entails speech recognition and text analysis to simulate human conversations via computer programs and create and understand clinical documentation.

If security measures are lax, unauthorized access to chatbot data can lead to a data breach. Chatbot developers must explore encryption options to prevent unauthorized access to sensitive information about, among other things, users' personal lives, mental health symptoms, substance use, health challenges, financial and legal problems, and trauma history. The communication channel between the user and the chatbot in addition to the chatbot data themselves can also be a target for attacks. If the data transmission is inadequately encrypted, data could be intercepted by third parties, leading to the potential exposure of sensitive information.

Two unique risks with chatbots are referred to as (1) *model inversion attacks* in which an attacker reconstructs sensitive data shared during a chatbot exchange and (2) *adversarial attacks* in which slight changes to input data can cause the model to make incorrect decisions or reveal sensitive information. Another risk is an *injection attack* in which someone introduces malicious data that the chatbot mistakenly processes, thus allowing unauthorized access to sensitive data by third parties.

Chatbot developers must be cognizant of ethical issues related to unauthorized access, implementation of regular security audits, compliance with federal and state privacy statutes and regulations, and the possibility that users will insert malicious code or scripts.

Case 1.1: Terrance L. and the Chatbot

A 42-year-old man, Terrance L. recently separated from his wife after 20 years of marriage. He was struggling with intense feelings of loneliness and sadness. Suspecting he might be depressed, Terrance logged on to a behavioral health website featuring a chatbot that incorporates key features of cognitive–behavioral therapy (CBT). He engaged the chatbot, which led him through a series of questions about his unique circumstances, feelings, and symptoms. The chatbot provided Terrance with empathic responses and suggested some breathing exercises, self-help ideas, and online resources. The chatbot also encouraged Terrance to consult with a local psychiatrist to explore possible medication.

In his exchanges with the chatbot, Terrance disclosed many personal and intimate details about his life. The behavioral health organization that sponsors the chatbot has an ethical duty to alert Terrance to risks associated with these disclosures, obtain Terrance's informed consent to participate in the chatbot exchanges, and take constructive steps to ensure protection of Terrance's privacy.

Social Robots

A *social robot* is a robot capable of interacting with humans and other robots. Social robots can provide emotional support, companionship, and personal assistance services to older adults, people with disabilities, children with special needs, and other vulnerable populations. Social robots are developed using AI and are often equipped with sensors, cameras, microphones, and other technology so they can respond to touch, sounds, and visual cues much like humans would. Using AI, a robot can decipher facial expressions, engage in conversations, respond with a smile, read text and email messages, place video calls, tell stories and jokes, and track humans with their eyes to prove it is paying attention.

Social robots use AI to gather information from users, such as nursing home residents and vulnerable people who live independently. Thus, it is important for developers to address ways to protect the privacy of this information and prevent discrimination based on race, ethnicity, national origin, religion, sexual orientation, gender expression, disability, or other personal factors. Behavioral health professionals who use social robots must explore the extent to which clients understand that these devices are mimicking human

emotion and expression and cannot truly empathize with clients. For clients who struggle with cognitive impairment or other challenges, practitioners must be specifically concerned about the extent to which their own use of social robots is deceptive and misleading.

Case 1.2: Emma S. and Social Robots

Emma S. is the director of social services at a nursing home where many of the residents struggle with social isolation and some form of cognitive impairment. Emma recently attended a workshop session at a professional conference that provided an overview of the ways in which some nursing homes are using "social robots" and AI to serve residents. The life-size social robot is dressed in human-style clothing and, using AI software, interacts with nursing home residents.

Behavioral health organizations that use social robots featuring AI software are obligated to consider relevant ethical issues and implications. These include the possibility that clients, especially those who are cognitively impaired, will be misled about the fact that they are not communicating with a human being. Further, behavioral health organizations that provide services to clients using social robots must comply with prominent ethical standards concerning informed consent. For clients who are not competent to provide consent, behavioral health organizations must ensure that a surrogate decision maker recognized by law has the authority to provide proxy consent on the client's behalf.

Machine Learning and Translation

ML is a branch of AI and computer science that focuses on the use of data and algorithms to enable AI to imitate the way that humans learn. Over time, ML uses these data to gradually improve its accuracy. According to the University of California–Berkeley School of Information (UC Berkeley, 2020), ML involves using statistical learning and optimization methods that let computers analyze datasets and identify patterns. ML techniques use what is known as *data mining* to identify historical trends and inform future models. Since ML algorithms update autonomously, in theory, the accuracy improves with each run as the algorithm teaches itself from the data it analyzes. This is called *iteration*.

There are several ML models: supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning (UC Berkeley, 2020):

> **Supervised learning:** The dataset being used has been prelabeled and classified by users to allow the algorithm to see how accurate its performance is.
>
> **Unsupervised learning:** The raw dataset being used is unlabeled, and an algorithm identifies patterns and relationships within the data without help from users.
>
> **Semisupervised learning:** The dataset contains structured and unstructured data to help the algorithm make independent conclusions. The combination of the two data types in one training dataset allows ML algorithms to learn to label unlabeled data.
>
> **Reinforcement learning:** The dataset uses "rewards" and "punishment," which provide feedback to the algorithm to learn from its own experiences by trial and error.

The typical supervised ML algorithm consists of three components (UC Berkeley, 2020):

> **1. A decision process:** A recipe of calculations or other steps that takes in the data and "guesses" what kind of pattern the user's algorithm is looking to find.
>
> **2. An error function:** A method of measuring how good the software's guess was by comparing it with known examples.
>
> **3. An updating or optimization process:** A method in which the algorithm looks at the "miss" (inaccurate guess) and then updates how the decision process comes to the final decision to minimize "misses" in the future.

In addition, the concept of *deep learning* involves automatically learning from datasets without introducing human rules or knowledge. This requires massive amounts of raw data for processing; the more data that are received, the more the predictive model improves (UC Berkeley, 2020). Deep learning relies on what are known as *deep neural networks* that are trained on large datasets to identify and classify phenomena, recognize patterns and relationships, evaluate possible responses to queries, and make predictions and decisions.

In behavioral health, ML software is designed to enhance accuracy in diagnosing mental health conditions and predicting client outcomes. For example, users who seek information to help them cope with mood disorders or anxiety can submit queries to chatbots that engage in dialogue and generate diagnostic questions, conduct assessments, and suggest treatment options and resources.

Behavioral health professionals who rely on ML and translation must address potential ethical issues related to the privacy and surveillance of users' data, possible bias and discrimination resulting from AI's use of historical data, and threats to users' autonomy if results from ML make decisions for them.

Case 1.3: Melinda G. and ML

Melinda G. was in recovery after many years of struggling with her addiction to heroin. She was on probation following her arrest for shoplifting and possession of heroin with intent to sell. As a condition of probation, Melinda was required to participate in an outpatient substance use disorders treatment program.

Melinda missed three of her previous five counseling appointments; twice, she overslept, and once, her car would not start. The director of the treatment program notified Melinda that she was being terminated from the program due to her attendance problems. Melinda knew that this termination would violate her probation conditions, which might lead to her arrest and incarceration. She felt desperate for help and looked online for crisis intervention services. Melinda connected with an online crisis intervention service that uses AI. The AI software was developed using a "reinforcement learning" model, which posed a series of questions to Melinda and offered responses to help her manage her crisis. The algorithm used Melinda's responses as feedback and made adjustments accordingly throughout their online exchange.

Behavioral health organizations that use AI-based reinforcement learning to serve clients must consider key ethical issues, including the possibility that some clients may not understand fully that this software may surveil and monitor clients' behaviors, emotions, and activities. Further, AI-generated responses to clients may reflect algorithmic bias that discriminates based on demographic factors, such as race, ethnicity, gender, income, sexual orientation, and gender expression. This would violate behavioral health practitioners' ethical duty to not discriminate against clients.

Search Engines

Powerful *search engines* use natural language processing and ML to enhance the quality of results provided in response to users' queries. Moving beyond popular search engines, such as Bing and Google, sophisticated search engines (e.g., Andi, Brave, Perplexity, Phind, and You) use natural language processing and ML to generate text and image results. For example, users can post questions about behavioral health challenges they are experiencing and receive detailed information about diagnostic criteria, treatment options, and resources.

Behavioral health practitioners who use sophisticated search engines must address ethical issues related to possible reliance on inaccurate or false information, manipulation of search engine results and rankings, privacy risks associated with search engines' collection of users' data, use of search engine queries for surveillance purposes, and violation of copyright law.

Case 1.4: Tess R. and Search Engines

Tess R. is a transgender woman who recently transitioned. She was distressed that her parents were not supportive of her decision and found herself engaging in suicidal thoughts. She logged on to a search engine and posed the question, "What kind of help is there for people who are having suicidal thoughts?" Within seconds, the search engine provided Tess with guidance under several major headings: direct support and professional help, connecting with support services, action steps for communicating with someone who may be suicidal, psychotherapies (i.e., CBT and dialectical behavior therapy [DBT]), warning signs of suicide, resources and brochures, and links to crisis services.

Search engines' use of sophisticated natural language processing can clash with prominent ethical standards in the behavioral health professions (Reamer, 2023c). In theory, search engine results can mislead clients, provide them with inaccurate information, collect users' personal information without their full knowledge and consent, and be used against users during legal proceedings. Sponsors of search engines that rely on AI must comply with key ethical standards related to informed consent to ensure that users understand potential benefits and risks.

Predictive Analytic Tools

Predictive analytic tools use AI to extract insights from large volumes of data and forecast outcomes. Google Cloud BigQuery analyzes large datasets to find patterns that can help predict future behavior. Microsoft Azure Machine Learning helps organizations build, implement, and manage predictive models. Qlik Sense analyzes data to generate visualizations and insights to enhance decision making. SAP Predictive Analysis generates predictive models that can be used by organizations to make data-driven decisions. Sisense helps users explore data, create visualizations, and identify trends. TIBCO provides users with predictive analytics that can be used for decision making and to improve outcomes. MATLAB uses ML to analyze data and create simulations. Certilytics analyzes clinical episodes to help organizations identify opportunities for cost efficiencies, quality improvement, and enhanced client outcomes.

Behavioral health practitioners who rely on predictive analytic tools must explore ethical issues related to the possibility of inaccurate or misleading results, invasion of users' privacy, and systemic bias (based on these tools' use of historical data that may be linked to ethnic, racial, cultural, economic, political, sexual orientation, gender expression, and other forms of bias). One prominent example is the controversial use of AI predictive analytic tools in public child welfare agencies' decisions about foster care placements. According to Trail (2024):

> In the last decade, child welfare agencies have increasingly used big data to develop and implement predictive models to help them make decisions about the lives of children in foster care. With the goal of improving and making more consistent decisions, the models are attractive because the promise of better decisions should also lead to better outcomes for children. However, these models are far from perfect, and they have attracted criticism for their use of biased data, disregard of individual rights, and arbitrary weights. (para. 2)

The earliest versions of these predictive models were implemented in California and Illinois to predict potential child abuse, but the models produced so many false positive results that both states terminated the programs (Trail, 2024).