Behavior Analysis in Practice
https://doi.org/10.1007/s40617-023-00868-z

DISCUSSION AND REVIEW PAPER

Starting the Conversation Around the Ethical Use of Artificial Intelligence in Applied Behavior Analysis

Adrienne M. Jennings[1] · **David J. Cox**[2,3]

Accepted: 2 October 2023

Abstract

Artificial intelligence (AI) is increasingly a part of our everyday lives. Though much AI work in healthcare has been outside of applied behavior analysis (ABA), researchers within ABA have begun to demonstrate many different ways that AI might improve the delivery of ABA services. Though AI offers many exciting advances, absent from the behavior analytic literature thus far is conversation around ethical considerations when developing, building, and deploying AI technologies. Further, though AI is already in the process of coming to ABA, it is unknown the extent to which behavior analytic practitioners are familiar (and comfortable) with the use of AI in ABA. The purpose of this article is twofold. First, to describe how existing ethical publications (e.g., BACB Code of Ethics) do and do not speak to the unique ethical concerns with deploying AI in everyday, ABA service delivery settings. Second, to raise questions for consideration that might inform future ethical guidelines when developing and using AI in ABA service delivery. In total, we hope this article sparks proactive dialog around the ethical use of AI in ABA before the field is required to have a reactionary conversation.

Keywords Ethics · Artificial intelligence · Applied behavior analysis

Artificial intelligence (AI) is a field within computer science that often aims to mimic human intelligence via computational processes and technological systems. AI has many different subdomains such as computer vision, which is aimed at mimicking humans' ability to differentially respond to stimuli within their visual field; robotics, which is aimed at mimicking movement through an environment without running into things and getting injured; speech recognition, which is aimed at understanding and responding to vocal-verbal behavior; and natural language processing / analysis, which is aimed at responding to and emitting textual stimuli.

Researchers have used many different approaches to teach computer systems to implement tasks that mimic human behavior (see Kautz, 2022, for an article-length review of the history of AI). One method that dominated AI during the 20th century is referred to as symbolic AI. Here, intelligence is assumed to result from the manipulation of abstract representations (e.g., symbols) where AI systems work by implementing a series of logic-like steps of reasoning using language-like representations of problems (Garnelo & Shanahan, 2019; Newell & Simon, 1976). Early examples of symbolic learning involved "expert-based systems" (a.k.a. "good old-fashioned AI"—GOFAI; Haugeland, 1985) wherein the rules for reasoning and logic for completing an intelligent task were defined and implemented in code by collaborations between subject matter experts and computer engineers. Though expert-based systems continue to exist even in applied behavior analysis (ABA; e.g., RethinkFirst's first-generation Medical Necessity Algorithm), symbolic AI has long since moved beyond GOFAI (Garcez & Lamb, 2020).

An alternative approach to teaching computer systems to implement tasks that mimic human behavior is sometimes referred to as a connectionist approach (Garnelo & Shanahan, 2019). In this approach intelligence is assumed to occur by learning associations from data (Goel, 2021). Connectionist approaches, arguably, dominate the current AI landscape in research and industry and involve supervised machine learning (e.g., Müller & Guido, 2016), unsupervised learning (e.g., Everitt et al., 2011; Patel, 2019),

✉ Adrienne M. Jennings
ajenning@daemen.edu

[1] Department of Behavioral Science, Daemen University, 4380 Main Street, Amherst, NY 14226, USA

[2] Institute for Applied Behavioral Science, Endicott College, Beverly, MA, USA

[3] RethinkFirst, 49 W 27th St, 8th floor, New York, NY 10001, USA

Published online: 03 November 2023

and reinforcement learning (e.g., Sutton & Barto, 1998). Connectionist approaches involve techniques ranging from classical machine learning techniques (e.g., Müller & Guido, 2016) to neural networks and deep learning techniques (e.g., Goodfellow et al., 2016). A full review of these techniques and how they work is well outside the scope of this article and readers are recommended to review the book-length citations for each of the areas above or the tutorial specific to behavior analysis by Turgeon and Lanovaz (2020). However, the primary distinction of connectionist approaches is that the computer systems learn associations from data with little or no prior knowledge.

Although humans can analyze and find patterns in data, the speed and efficiency whereby computers can analyze and process data often exceeds what humans are capable of. Further, recently developed AI technologies suggest these systems identify knowledge about the natural world without making the same assumptions and taking the same approaches that humans do to generate knowledge (e.g., Evans & Gao, 2016; Marchant, 2020; Sadler & Regan, 2019). This speed and efficiency for analyzing and making sense of large, complex datasets creates significant promises and possibilities for the use of AI in the delivery of ABA services (see Cox & Jennings, 2023, for a recent discussion). However, as this article will discuss, AI is another useful technological tool like the microwave oven, cellular phones, or magnetic resonance imaging (MRI) machines. And, like other tools, AI can be used in ways that align with, or run counter to, people's ethical preferences.

Some of the most well-known uses of AI include robotics, autonomous vehicles, digital assistants, and the now (in) famous ChatGPT (OpenAI, 2023). For example, robotic vacuums map the layout of the house to determine the best route and avoid obstacles (e.g., iRobot's Roomba). Robots assist surgeons to perform coronary artery bypass surgeries and appendectomies (e.g., Food & Drug Administration, 2022). Autonomous vehicles use AI to navigate roads based on data from sensors, cameras, and stored maps (e.g., Matheson, 2019; Viter, 2019). Digital assistants such as Siri, Alexa, or Cortana use AI to provide information, send messages, set reminders, and so on (Oracle, n.d.). Educators use AI to assign lessons and check for plagiarism (e.g., Institute for Ethical AI in Education, 2020). And, ChatGPT is available for conversation about any topic users would like to discuss (though the accuracy of ChatGPT's responses is not guaranteed).

Although the notion of AI is likely somewhat familiar to readers, practicing behavior analysts may question why this topic is relevant to them. Our rationale is that practicing behavior analysts who learn to use AI may be in a better position to help their clients more effectively or efficiently.[1] For example, an experiment by Cantin-Garside et al. (2020) showed how AI can automatically detect and monitor self-injurious behavior via motion sensors. Other researchers have used AI to detect stereotypy (Dufour et al., 2020; Fasching et al., 2013; Plotz et al., 2012), to assist with diagnosing autism spectrum disorder (ASD; Erden et al., 2021; Song et al., 2019), and to identify assessment questions that best predict ASD (e.g., Bone et al., 2015; Kosmicki et al., 2015). And, still others have used AI to identify patient profiles for more accurate hour recommendations (e.g., Cox et al., 2023). Other researchers have sought to improve data analysis via AI. For example, researchers have used AI to analyze single-case graphs with AI performing similarly or better than humans (e.g., Lanovaz et al., 2020; Lanovaz & Hranchuk, 2021; Taylor & Lanovaz, 2022). AI is even being used in administrative processes such as with CentralReach's "smart scheduling" system that seeks to maximize authorized hours by accounting for client availability and approved hours, therapists' location and drive times, therapists' credentials, as well as real-time events such as cancellations (CentralReach, 2020).

Despite these exciting innovations, AI is like many technologies in that it is developing faster than its corresponding oversight (Sacasas, 2018; Schneier, 2019). Although AI was introduced in the 1950s, serious conversations around the ethical use of AI have largely emerged in the last few years (e.g., Ashok et al., 2022; Floridi & Cowls, 2019).[2] Such a lag between technology development and ethical oversight leaves open the possibility for its unethical use. For example, Cambridge Analytica harvested personal data from millions of Facebook users and used AI to generate targeted political ads (Zialcita, 2019). A lack of oversight allowed Facebook to be lax in their data security and Cambridge Analytica to use data without consent; the combination led to an estimated 200+ elections around the globe being affected.

As a community, we have an opportunity to discuss proactively how AI should be used as a tool by practicing behavior analysts and what we expect—ethically—from technology companies when they build and publish new AI technologies. To help instigate this conversation, we chose to focus our discussion around topics at the intersection of two domains. First, topics that currently dominate the ethical use of AI in healthcare. Second, topics common to practicing behavior analysts' discussions of ethical clinical practice and in research.

The topic of AI ethics is broad and deep and a full treatise is well beyond the scope of this article (e.g., Dubber et al., 2020). Nevertheless, the literature around the ethical use of AI

[1] For a article-length review of AI use cases relevant to ABA, see Cox and Jennings (2023).

[2] Even this article is a case in point. It was originally drafted in the summer and submitted in the fall of 2022. In the months following its submission and throughout its peer-review process, ChatGPT 3.5 and 4.0 were both released, which significantly changed the general awareness and knowledge of what AI is, its widespread utility, and its regular use. We can only speculate what AI capabilities will exist when this article finally makes it to print, let alone in the few years postpublication.

in healthcare currently places heavy emphasis on topics such as security and privacy of patient data (e.g., Murdoch, 2021); algorithm/model transparency and how to avoid or respond to mistakes made by AI (e.g., Mörch et al., 2020); what counts as reasonable algorithm/model explainability (e.g., Amann et al., 2020; Loh et al., 2022; Martinho et al., 2021); and how to navigate and define equity issues (e.g., Berdahl et al., 2023; Lamont & Favor, 2017). In the sections below we highlight how each of these areas from the healthcare ethical AI literature intersects with the clinical ethics literature in behavior analysis.

There are several topics within the clinical ethics literature that are often referenced in behavior analytic discussions around clinical ethics. These include the ethical principles espoused in the Belmont Report (Office of the Secretary, 1979) and the principlist approach most notably described by Beauchamp and Childress's *The Principles of Biomedical Ethics* (1979). We chose to lead from this perspective for two reasons. First, these ethical principles form the foundation for many current codes of ethics and guidelines for responsible conduct for many healthcare professions, including behavior analysts (e.g., Behavior Analyst Certification Board [BACB], 2020; Byrd & Winkelstein, 2014). This makes the language and principles espoused likely to be familiar to practicing behavior analysts. Second, the Belmont Report has served as one of the foundational guides for current conversations around AI ethics (e.g., IBM, 2021). These ethical principles allow behavior analysts to leverage, extend, and adapt existing work in AI ethics to the unique research and practice use cases of AI in ABA.

By definition, ethics involves statements about "right" and "wrong" for a social group. Thus, guidelines and rules for the ethical use of AI in ABA should include input and expertise from many behavior analysts and the people who seek them out for their services. Below, we take a first pass at framing the ethical use of AI in ABA around the ethical principles from the Belmont Report (Office of the Secretary, 1979) and Beauchamp and Childress (1979). In so doing, many unanswered questions arise that the social group referred to as "practicing behavior analysts" will likely have to answer. Relatedly, we also realize this is one perspective and one approach to discussing this important topic. Our ability to manage the use of AI as a field of collegial professionals will likely improve to the extent that many voices are included. We are excited to begin this discussion and we hope that many readers of diverse opinions and experiences join us in this important conversation.

Notes on Scope

There are two important comments around the scope of what follows that are important to make explicit before moving further. First, there are rather robust and substantially large literature bases on relevant ethical considerations such as technology ethics, data ethics, and AI ethics, as well as clinical ethics, medical ethics, and public health ethics. We chose to focus this article on some of the common major topics across data and AI ethics as relevant to their intersection with common topics in clinical–ethical decision making by practicing behavior analysts. Thus, the list of questions and considerations below is necessarily incomplete and provides much opportunity for future ethical work. But, as noted by Gasser and Schmitt (2020), norms of ethics from the professions implementing AI often provide initial guidance and governance around the use of an AI system and, thus, are used for guidance around initial conversations on the use of AI in ABA.

The second important note is the type and level of technology development and deployment we refer to with AI systems in this article. Research and clinical ethics in ABA often involve decisions of similar scope given researchers' use of within-subject designs that practicing behavior analysts can directly replicate (Cox et al., 2022; Normand & Donohue, 2023). That is, researchers research at the level of the individual client and practicing behavior analysts can implement that research at the level of the individual client.[3] In contrast, technology is developed and deployed at the level of groups of people. Though individual considerations and preferences are likely kept in mind, it is often impractical or impossible to build technological platforms that are perfectly customizable to the individual needs and preferences of every individual user. Thus, the AI systems referenced in this article are not the one-off, small *N*, proof-of-concept research demonstrations for how AI can be used in ABA and that are often published in academic journals. Rather, the AI systems referenced in this article are those that have been (or will be) pushed into production and widely accessible by practicing behavior analysts as built by technology companies for scale of impact[4] (e.g., ChatGPT). As with the emphasis on behavior analysts in practice above, this highlights the size of this topic and the many areas for fruitful future work.

[3] This does not mean that the contingencies that surround research and practice are identical. The differential contingencies often lead to differential claims about what is the "right" thing to do in a specific context (e.g., the necessity of conducting a reversal following an intervention that changed behavior in socially significant ways).

[4] It is possible for individual researchers to build and deploy AI models built from a small sample size that are widely accessible by practicing behavior analysts. However, a critical concern here is how well the participant characteristics within the small sample sizes apply to the client characteristics for whom a practicing behavior analyst might want to use the technology. Unlike research in ABA that develops methodologies that can be tailored based on the individual client characteristics and clinical expertise of the supervising BCBA, most technologies cannot be tailored in the same manner. The section on model bias and generalizability makes more explicit this distinction (see below). Further, data security and privacy are a critical concern when PHI is entered into platforms leveraging AI models online. Users should always verify that those who deploy models in online applications have adhered to the necessary legal requirements around data security and privacy.

Beneficence and Nonmaleficence

The first ethical principles we will discuss are *nonmaleficence* and *beneficence*. *Nonmaleficence* has its roots in the Hippocratic Oath and might be the oldest rule to which healthcare practitioners have historically been held accountable: "First, do no harm" (Edelstein, 1943). The logically related opposite end of the utilitarian perspective to *nonmaleficence* is *beneficence*: maximize benefit. Together, *beneficence* and *nonmaleficence* often work in tandem to guide researcher and practitioner behavior to make utilitarian justified decisions that "lead to the greatest benefit for the greatest number of people" (Driver, 2014).

Beneficence and *nonmaleficence* create numerous relevant considerations for the ethical use of AI in ABA. Broadly, *beneficence* and *nonmaleficence* would suggest that AI should be developed in a way that maximizes benefit for as many individuals as possible while, simultaneously, avoiding foreseeable harms when developing, deploying, and embedding AI into ABA service delivery. This differs slightly from clinical decisions where clinicians make decisions that maximize benefit and minimize harm for each client; though it might be somewhat similar to decisions clinicians make where they need to consider maximizing benefit and minimizing harm for all clients on their caseload (e.g., how to allocate supervision time each week). Technologies, in contrast to individual clinical decisions and significantly scaled up from caseload decisions, are designed and built to apply to groups of people and thus the application of ethical principles has a different frame of reference.

To make these principles more tractable at the scope of developing and deploying AI technologies at scale, we next review common topics in AI ethics where *beneficence* and *nonmaleficence* are applicable to the current or future use. Table 1 summarizes some of the questions around *beneficence* and *nonmaleficence* that arise with the development and deployment of AI systems in ABA.

Data Security and Privacy

At its core, AI technologies are built by developing mathematical and computational models that relate one set of data (e.g., independent variables, environmental characteristics, inputs) to another set of data (e.g., dependent variables, patterns of behavior, outputs; see Table 2). For our purposes, the inputs might be client characteristics, programmed or unprogrammed respondent and operant contingencies surrounding environment–behavior relations, and the behavioral or technological systems that implement those contingencies. The outputs might be the socially significant patterns of behavior targeted with the ABA program that we are trying to describe, predict, and control. The computational and

Table 1 List of Questions Around the Ethical Use of AI Specific to the Ethical Principles of Nonmaleficence and Beneficence from the Text

Principle	Topic	Question(s)
Beneficence & Nonmaleficence	Data Security and Privacy	How is the AI system builder adhering to legal requirements related to HIPAA and HITECH? How certain does the success of a new AI system have to be to ethically allow practitioners and researchers to collect sensitive data outside the scope of current client programs? What are the benefit–risk tradeoffs that justify collecting sensitive data outside the scope of current programs? Is it ethical to collect and use PHI/PII from clients that do not directly benefit?
	Cost to Build and Maintain	Is it okay to potentially hinder current client progress by allocating resources to build out AI systems that will potentially benefit future clients at an unknown time?
	Transparency and Mistakes	How can we describe the "intent" of the AI system sufficiently to warrant consent to participate as "informed?" How can the AI system developers demonstrate they have identified all probable benefits and harms from building and deploying the AI system? How probable of an event makes it worth sharing in the consent process? What counts as sufficient protection from harms that may arise when the AI system is used to inform treatment? How will we know an error has been made? What is our corrective protocol when an error is detected? How will the occurrence of an error be communicated to clients? Who is held responsible when an AI system makes a mistake?
	Model Bias & Generalizability	What is the best way to communicate the scope of clients and clinical contexts for whom the AI system was trained? How should client choice with data sharing affect their ability to access the AI system?

Table 2 List of Terms and Definitions Relevant to Computer Science and AI as Used in This Article

Term	Definition
Artificial Intelligence	". . . software that is developed with [specific] techniques and approaches and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." (European Commission, 2021)
Model	Program that analyzes datasets to find patterns and make predictions, decisions, or perform specific tasks.
Algorithm	A set of rules or instructions AI systems are programmed to follow.
Prediction	The output of an AI model determined by the inputs.
Training Data	Information or exemplars that are provided to teach the model how to respond or formulate outputs.
Validation Data	The portion of the dataset that is used to tune architectural parameters of a model to prevent overfitting or underfitting during training.
Testing Data	The dataset used to evaluate the performance of an AI model after it has been trained.
Cross Validation	A statistical method used to estimate the overall variability and performance of models.

mathematical system that maps these inputs to the outputs is called what we are referencing as a model (see Cox & Vladescu, 2023, for a more in-depth discussion on what counts as a model in behavior analysis).

Generally, the data used to build the models that form the foundation of AI come from actual clients. Though not a perfect truism, the ability for AI models to predict behavior often improves based on at least two characteristics of the underlying data. First, model performance often improves with more detailed and personalized information from each client. Second, model performance often improves with more data and greater variability in the data inputs. Stated differently, model performance often will improve with greater data variability across client characteristics, information about programmed and unprogrammed contingencies, and the *who, what,* and *when* around the implementation of those contingencies. In total, the more personal health information (PHI) and personally identifiable information (PII) included in the dataset used to build AI systems, the more potential the models likely have to be better.

Accessing and storing large volumes of PHI and PII data with many clients creates many opportunities where sensitive health or education privacy can be compromised. Federal regulation through the Health Insurance Portability and Accountability Act of 1996 (HIPAA; Pub. L. No. 104-191) requires healthcare providers adhere to legal obligations around protecting PHI. And, the Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH; Pub. L. No. 111-5) expanded legal obligations to protect PHI stored in electronic health records (Congressional Research Service, 2009). Outside of legal requirements, related ethical requirements are included in the ethics code for board certified behavior analysts (BCBAs); in particular, Codes 2.03 and 2.05 which pertain to protecting confidential information and documentation around data protection and retention (BACB, 2020). In short, BCBAs are responsible for protecting confidentiality, as well as "storing, transporting, retaining, and destroying" electronic documentation. ABA practitioners are not unique in this realm as most professions make similar requirements (e.g., the code of ethics for the American Public Health Association likewise stipulates "collect only data elements . . . [that are] necessary. . ." [American Public Health Association, n.d.]).

Thus, some clinical–ethical ethical guidance exists around *what* data behavior analysts collect—only the data necessary to support or improve client progress. But many questions remain when we start to consider collecting large volumes of data to develop AI tools targeted at populations of people. Here questions the field has to answer surround, "necessary for what" exactly, and what counts as "support" or "improvement" in client progress? Restricting necessary to only the success of current and individual clinical/educational programs and systems will likely limit future advancements. But, including necessary information for building AI systems with future *potential* benefits adds uncertainty and may involve collecting data that turns out to be useless. For example, data recorded for a skill acquisition program minimally consists of the date and percent correct, and perhaps the therapist's initials and prompt level. However, for building an AI system that could recommend troubleshooting strategies or future programs, it may be beneficial to include other information related to that client's profile (e.g., diagnosis, age, assessment scores), as well as information on the skills and abilities of the therapist (e.g., age, education, on-the-job training, on-the-job performance with varying clients and similar programs). This also raises the question *for whom* are the data collected? Clients whose PHI/PII are contributed to training data sets for AI systems may not directly benefit as much as future clients given how long robust AI systems can take to build. Questions the field will have to answer include, how certain does the success of a new AI system have to be to allow for collecting sensitive data outside the scope of current programs? What are the benefit-risk tradeoffs that justify collecting sensitive data outside the scope of current programs? And, is it ethical

to collect and use PHI/PII from clients who do not directly benefit?

A related ethical question is *where* the data are stored once collected. The days of pen and paper data collection are (we hope for the sake of efficiency) over. But, the use of electronic data collection platforms takes the data "out of the hands" of BCBAs. Further, it is unlikely that ABA organizations have employees who are proficient at the design and implementation of AI. The result is that ABA organizations are increasingly likely to pay for technological systems that allow the therapists and BCBAs to collect data electronically, store the data securely in the cloud, access the data whenever needed for clinical decisions, and make decisions about when and how to use AI. ABA organizations need to do their research before signing onto these systems to ensure the technology platform that provides data storage and AI tools has protection and storage policies that align with state and federal laws, and ethical codes. Although data security is not guaranteed by one specific law, multiple federal and state laws offer protection (Klosowski, 2021). Given the details of these laws can vary from state to state, readers should familiarize themselves with the requirements in the geographical locations within which they provide services.

Cost to Build and Maintain

The cost to build and maintain AI systems is another ethical consideration under the principles *beneficence* and *nonmaleficence*. Creating and deploying AI systems that are technologically robust and applicable at the population level requires a lot of data from a diverse set of clients, as well as trained personnel to build, monitor, and improve the AI system (e.g., computer scientists, data scientists, data engineers, domain experts). Further, AI systems require hardware and software to store, maintain, and operate the AI system. All of this involves money, time, and other resources that companies could allocate to other clinical and organizational systems that may improve client outcomes. This leaves the ethical question of whether it is okay to potentially hinder current client progress by allocating resources to build out a system that will potentially benefit future client progress at an unknown future point in time. Similar to above, a risk-benefit analysis could be used to determine how each of the different variables at play might lead the field to determine under what circumstances pursuing an AI system is ethically justified. Some of the variables to consider when determining whether resources are ethically allocated to build AI systems include: the amount and duration of current hindrance to client progress; delay, probability, and amount of future potential benefit to clients; and the delay, probability, and amount of benefit to clients if those same resources were allocated to other available projects.

Transparency and Mistakes

A third potential area for harm mitigation involves transparency when building and deploying AI systems. Transparency of AI systems minimally involves detailed information about how a system works, what data are used, who the system has been optimized for, and the limitations of that system (e.g., Mörch et al., 2020). As described by the American Medical Association, transparency of AI means patients should be informed of "the intent behind the development of an AI system" (American Medical Association, 2022). Clients have the right to be informed how their data are used to develop current and future AI systems and how the AI system is used to inform treatment. Here, ethical questions the field will need to answer include, how can we describe the 'intent' of the AI system sufficiently to warrant consent to participate as being truly informed? How can the AI system developers demonstrate they have identified all material benefits and harms from building and deploying the AI system? How probable of an event makes it worth sharing in the consent process? And, what counts as sufficient protection from harms that may arise when the AI system is used to inform treatment? Though the *reasonable person standard* exists to guide what information should be included in informed consent processes (Odwazny & Berkman, 2017), AI systems are not yet widely accepted and understood technologies like *x*-ray or electrocardiogram machines. Given the novelty and potential misunderstanding around what AI systems are and how they work, it is unclear what a "reasonable person" might want to know about how AI systems influence what healthcare professionals claim as the benefits and harms of all available procedures.

In line with transparency,[5] it is also important to understand when and how the AI systems make errors in their predictions and recommendations, and how large those errors are. Though algorithms are often framed as being black boxes and the most advanced AI systems may exceed humans' abilities to fully comprehend them, a lot can still be known and made transparent about AI systems (Diakopoulos, 2020). Parroting an example offered by Diakopoulos (2020), the specifics of your favorite restaurant's recipes may only be known to the chef. Nevertheless, inspection of the kitchen can still identify issues with ingredients,

[5] It's helpful here to denote the difference between explanation and transparency in AI systems. As defined by Diakopoulos (2020), "Explanation entails a system articulating how it made a particular decision and is typically causal (e.g., input influence or sensitivity-based) or involves case-based comparisons, whereas transparency disclosure involves descriptions of system behavior and design intent but leaves any final causal explanation of system behavior to the evaluation of information disclosures by interested stakeholders" (p. 204).

kitchen cleanliness, and food handling that allow for the assessment and improvement of food safety—despite its lack of complete information about the food being served. Likewise, developers of AI systems can produce information that makes transparent its design and implementation, the processes and outputs, and how data are handled and used. Here, practical transparency means AI system developers produce enough information to promote their effective governance, the accountability of the system, and what accountability can be placed on the users of the system.

Before AI systems are incorporated into ABA services, ABA organizations and practitioners should plan for how they might catch and respond to errors based on the transparency-related information made available by the AI system developers. Important questions here that are already being debated in the AI ethics in healthcare literature include: How will we know an error has been made? What is our corrective protocol when an error is detected? Who is held responsible when an AI system makes a mistake? And, how will the occurrence of an error be communicated to clients? As noted above, it takes a village to build, deploy, maintain, and monitor AI systems. Everyone involved should be informed of their role and the systems for accountability if any harms occur.

Model Bias and Generalizability

A fourth important topic related to the *beneficence* and *nonmaleficence* of AI systems involves model bias and generalizability based on the data used to develop the AI system. Many researchers and practitioners are likely familiar with the notion that the accuracy and validity of our claims decreases when talking about environment-behavior relations outside of the situations for which we have collected data. For example, if you only have data on functional behavior–environment relations in a school setting, you are likely to be much less accurate speaking about functional behavior–environment relations in the home. A more succinct guideline is, "Don't speak beyond your data." The same is true when building quantitative and computational models. Delay discounting researchers typically speak about choice within the range of delays measured, matching researchers typically speak about choice within the reinforcement ratios measured, and AI systems are designed to perform as well as possible based on the range of inputs that were used to create the underlying models. Generalizing predictions to data outside the range of data used during training is difficult and may lead to error greater than what was observed during model training, validation, and testing (see Table 2).

Model bias can be defined in several different, though interrelated, ways. One definition of bias could be claimed to come from psychology where bias refers to systematic errors that deviate from logic or rational behavior (e.g., Dunbar et al., 2014; Tversky & Kahneman, 1974). Translated behavior analytically, bias can be defined as systematic deviations in responding from what is logically expected based on the observed reinforcement schedules in effect[6] (e.g., Baum, 1974). An example meeting these definitions of bias in AI is historical racial discrimination in facial recognition technology (e.g., Najibi, 2020; National Institute of Standards & Technology [NIST], 2019). Described around the definitions above, the facial recognition technologies tested by NIST (2019) consistently demonstrated discrepancies in classification accuracy for different skin tones and sexes. This is not rational because skin tone and sex should not affect the ability for technology to recognize faces (psychological definition of bias). It also seems unlikely the technology creators purposively "reinforced" and taught the system to make mistakes specific to skin tone or sex. Thus, the technology was likely behaving in ways that deviated from the feedback schedules programmed to train the model (behavioral definition of bias).

Both of the above definitions of bias are unique, though not necessarily distinct, from the definition of bias present in the classic bias–variance tradeoff when developing AI models (e.g., Kohavi & Wolpert, 1996; Neal, 2019). Here, the errors made by a model can be separated into three components: bias, variance, and noise. As an equation using words: Error for observation x = Bias^2 + Variance + Noise. As an equation using mathematical symbols:

$$\mathrm{Error}(x) = \mathrm{E}\left(\left[\hat{f}(x)\right] - f(x)\right)^2 + \mathrm{E}\left[\hat{f}(x) - \mathrm{E}[\hat{f}(x)]^2\right] + \varepsilon. \quad (1)$$

Here, $\mathbf{E}[\hat{f}(x)]$ refers to the average prediction after the model has been trained over several independent data sets,[7] $f(x)$ refers to the actual value of the data point attempting to be predicted, and ε refers to remaining noise.[8] In this

[6] We are not implying here that matching is equivalent to maximizing as research has shown this relation to not always hold (e.g., Mazur, 1981). Rather, we are highlighting that a precise and quantitative definition of bias can and has been offered from an operant perspective and that could be aligned theoretically with definitions from other areas of psychology (*a la* Skinner, 1945). What is the "best" definition of bias will obviously depend on the context and the function of the speaker's behavior.

[7] Readers interested in more detailed discussions around various methods for test–train–validation splits of datasets and cross-validation are referred to Müller and Guido (2016). Definitions are offered in Table 2.

[8] For a fun, interactive, in-depth, and intuitive explanation of the bias–variance tradeoff in machine learning, readers are referred to Wilber and Werness (2021).

definition, bias is defined as how off a model's predictions are from the actual values (i.e., $\mathbf{E}[\widehat{f}(x)] - f(x)$) and variance is defined as how much predictions vary for any given data point (i.e., $\mathrm{E}[(\widehat{f}(x) - E[\widehat{f}(x)])^2]$). Note that all of these calculations are made using *only* the data available to the AI system developers when they develop an AI model.

We can generalize the various definitions of bias above back to the current article. All AI models are necessarily limited in how well they generalize beyond the training data. The more similar that unseen data are to the data used to train the model, the more likely the model is to perform well on the new, unseen data. The further that unseen data are from the data used to train the model, the less likely the model is to perform well on the new, unseen data. Here, similarity might be defined as the unique combination of variables that make up a new observation. Or, similarity might be defined relative to the range of values present for each variable in the dataset used for training, testing, and validating (e.g., historical racial discrimination in facial recognition technology). We can describe the bias of a model precisely in various ways (e.g., bias term in the matching law; bias term in Equation 1). But, regardless of how we describe bias quantitatively, the model built on training data might behave in ways that the users and developers do not logically or rationally want it to behave.

When bias is observed in an AI system, one commonly attempted solution is to improve the datasets used to train and test the models to better represent the variability present in the data input by users of the product (see Roach, 2018, for how Microsoft attempted this for their facial recognition technology). Thus, to reduce potential bias from the outset, AI system developers try to obtain enough training data representative of a large sample, to produce a model that could apply, or generalize to a variety of clients. To illustrate, consider an AI system that was created to predict success rates with a specific behavior reduction protocol using data from clients ages 3–5 years old, with an ASD diagnosis, Autism Treatment Evaluation Checklist (ATEC; Rimland & Edelson, 1999) scores between 50–70, in upper-class socioeconomic status, all receiving early intensive behavioral intervention. Using the AI system with older individuals, with other diagnoses, with higher or lower ATEC scores, in a different socioeconomic status, and/or receiving different forms of behavior analytic services, may provide predictions that are less accurate than indicated when the model was trained. This does not mean that AI systems should never be applied to novel cases. Rather, it suggests that AI system users should be aware of what data went into training the AI system, how those data compare to the data the user will input, and users should exercise correlated caution based on how well the two align.

To maximize the benefit and minimize the potential harm of AI systems in ABA to as many clients as possible requires data be used to build AI systems from as diverse a group of individuals as possible. Diversity might include behavioral repertoires, cultural backgrounds of clients and staff, education and training of the staff, intervention design and procedures, and anything else that may play a role in the effectiveness of an ABA program. For each AI system use case, it will take time for models to be trained with enough data that represents a wide variety of diversity. Initial models may have biases due to limitations in the data or how the models were developed. Bias in the AI system may arise not because of any malicious intent of the people building the AI tool, but because the data collected are not representative of all the people to whom users of the AI system apply the technology. AI system developers should acknowledge and discuss the limitations for who and what situations the AI system has been trained. In turn, practitioners should understand the limitations of any AI system and talk with clients about the scope of an AI system before it is incorporated into treatment, in line with the ethics code for BCBAs to be truthful (Code 1.01).

Autonomy

A second ethical principle we will discuss is *autonomy*. *Autonomy* in healthcare generally refers to the right we each have to make independent decisions about our body and the healthcare we receive. In behavioral terms, each individual has a right to contact healthcare choices wherein the related contingencies do not lead to a choice different from what would be predicted by their healthcare preferences (i.e., free from undue coercion and constraint). There are at least two historical conversations around *autonomy* that are relevant to building, deploying, maintaining, and monitoring AI systems. First, we each have the right to know how our health-related data is being used and to control with whom that data is shared. Note here the potential conflict between *beneficence* and *nonmaleficence* and *autonomy*. As described in the previous section, maximizing benefit and minimizing harm suggests we would want to include as much data from as many different interventions and people as possible. Further, each client will likely benefit more if the AI system is trained using their data. Knowing this, researchers and practitioners may unknowingly place undue influence and constraint on their clients to share their data which might violate the principle of autonomy. A second conversation around *autonomy* is our right to decide how an AI system influences intervention recommendations and decisions. To do this, however, creates several requirements for AI system researchers. Table 3 highlights some of the questions around *autonomy* that arise with the development and deployment of AI systems in ABA.

Table 3 List of Questions around the Ethical Use of AI Specific to the Ethical Principles of Autonomy from the Text

Principle	Topic	Question(s)
Autonomy	Explainability	What does an explainable AI system look like? What counts as providing sufficient information about how the AI system works, and the probable benefits and harms of the AI system such that consent to use a client's data is "informed?" What level of detail is needed to make AI systems reasonably explainable, how does a reasonable explanation differ for different people, and at what point does it become critical to talk about? How will consent for data use be obtained and by whom (e.g., ABA organization, tech company)? What happens if consent is revoked? How often does information need to be communicated to clients?
	Data Ownership	Who owns the data collected by technicians and behavior analysts during ABA sessions? How might partial ownership be divided out? How does data ownership influence how data can be shared? Should data sharing be opt in or opt out? Should clients have a greater role in conversations about the data collection systems used by their providers?

Explainability

Reasonable explainability[9] might refer to the description of an AI system in a way that can be understood by others (e.g., Martinho et al., 2021). In the healthcare literature, researchers have described the detrimental effects due to lack of explainability (e.g., Amann et al., 2020) and have offered suggestions for how explainability can be improved (e.g., Loh et al., 2022). Reasonable explainability is buttressed by clinical ethics codes around requirements for informed consent. For example, the code of ethics for the American Psychological Association states services are described "using language that is reasonably understandable" (American Psychological Association, 2017). The code of ethics of the American Occupational Therapy Association has the requirement to "fully disclose the benefits, risks, and potential outcomes of any intervention" (American Occupational Therapy Association, 2020). And, the ethics code for BCBAs necessitates that interventions and assessments are described prior to implementation (Code 2.08, 2.16). Thus, from an ethical guideline standpoint, healthcare professionals would likely agree that AI systems should be explainable so consent is properly informed.

But, what does "an explainable AI system" look like? That is, to what degree of explanation for how the AI system works and is used meets our ethical obligation? Many ABA practitioners are likely familiar with explaining how preference assessments work, the importance and impact of conducting functional analyses, and deriving socially practical reinforcement schedules through conversations with parents and caregivers. But, how many practitioners have explained how their electronic data collection platform works? How many of us have been provided with an explanation for how the output of an echocardiogram, *x*-ray, or MRI machine influenced your doctor's treatment decisions? How many practitioners explain how published research, decision aides, and peer support/review networks in their company inform clinical decisions and recommendations? Should we be more direct in explaining how we make intervention decisions and recommendations? Or do only "novel" technologies and tools that are less widely used need to be explained? What level of detail is needed to make AI systems reasonably explainable, how does a reasonable explanation differ for different people, and at what point does it move from seemingly silly to talk about (e.g., the data collection platform someone uses) to being critical to talk about (e.g., the AI robot that performs surgery; the AI system making clinical recommendations around ABA treatment goals)?

Further, as the sophistication of AI systems increases through machine learning, explainability likely becomes more difficult. In some circumstances we do not understand the exact mechanisms for how something works and it may be a "black box." In these situations, AI systems can leverage explainability tactics used with other "black boxes" in healthcare and education. For example, we do not fully understand how some regularly prescribed medications work such as antidepressants and mood stabilizers (Institute for Quality & Efficiency in Health Care, 2020; McCoy et al., 2022). These medications continue to be used based on inferences for the underlying mechanisms and data from successful use cases. And, we suspect prescribing professionals

[9] A parallel might be drawn here to the "reasonable person standard" for informed consent. More complicated AI systems (e.g., large language models, neural networks) may not even be fully understood by the computer scientists who build and maintain them. Nevertheless, as noted by Martinho et al. (2021), many healthcare practitioners want to have a reasonable understanding of how the AI system works even if it requires more learning and education on their part. We suspect clients or patients whose care is affected by AI systems will likely want a "reasonable explanation" though the amount of detail may differ from those of healthcare practitioners.

have figured out how to explain how these medications work sufficiently such that patients feel fully informed to consent.

The takeaway seems to be that clients have a right to an informed decision. We may not question how an echocardiogram, *x*-ray, or MRI scan informs our doctor's decisions given that these are established technologies with well-known precision around the information they provide to physicians. But, the conversation is different with emerging, experimental, and still-to-be-proven technologies such as AI. Practitioners who use AI systems should be prepared to answer clients' questions about how an AI system influences their treatment decisions and recommendations. It is important to note that practitioners' ability to answer those questions creates demands for reasonable explainability on part of the researchers building the AI systems. What will those demands be, exactly? And, at the end of the day, who is responsible for making the AI system reasonably explainable and for explaining the AI system to the client in a manner that meets ethical obligations for informed consent? As of now, we suggest that answering these questions requires regular, ongoing collaboration and communication across all relevant parties (e.g., computer scientists, data scientists, data engineers, domain experts, behavior analysts) when new AI technologies are being developed. Once reasonable explainability has been established, how will consent for data use be obtained? What happens if consent for data use is revoked? And, because systems are always evolving and adapting, how often does information need to be communicated to clients?

Data Ownership

A related, and critical topic, at the intersection of *autonomy* and AI ethics is data ownership. Who owns the data collected by RBTs and BCBAs in ABA sessions? Does that data belong to the patient because it contains their PHI? Does it belong to the ABA organization who engaged in the hard work to devise the data collection system and collect the data? Does it belong to the electronic data collection platform who engaged in the hard work of building the software and hardware that allows the data to be collected, stored, accessed, and maintained? Or do all have partial ownership in some capacity? And, how do legal claims here (McGuire et al., 2019) interact with ethical claims?

Data ownership questions directly influence what happens with the data once collected. Can ABA organizations or data collection platforms use the data however they want (as applicable under the law) once the data are collected? If so, what are the implications of this? Should clients have to opt-in to or opt-out of sharing their data beyond data use specific to their ABA services? This becomes a nontrivial choice as past researchers have found that the default option we present is what will likely be selected by most clients (e.g., Davidai et al., 2012; Dholakia, 2021; Johnson et al., 2002). Further, few clients likely have input on the decisions that ABA organizations make around their data collection systems and processes. Should clients be more involved in these decisions? If so, what exactly does that look like and who is responsible for informing clients so they can make an informed decision? Once a decision is made, how long does that consent last? And, what happens when a client changes their choice?

Justice

A final ethical principle we will discuss is *justice*. Traditionally in healthcare, *justice* refers to treating people equitably. With AI systems in ABA services, *justice* might refer to ensuring that clients have equitable access to the benefits that result from incorporating AI, as well as fair distribution of the costs associated with developing the AI system. Table 4 highlights some of the questions around *justice* that arise with the development and deployment of AI systems in ABA.

Access to Benefits

In an ideal world everyone would have equal access to the benefits of healthcare technology that improves their quality of life. In the real world, however, this is extremely difficult to do. Developing technology often costs a lot of time, money, and other resources that need to be recouped when sold to customers. When many resources are needed to build technology, such as AI systems, the resulting cost of the system can make it inaccessible for ABA organizations or clients who cannot afford it. This can lead to only more affluent individuals and ABA organizations accessing the benefits of AI systems contributing further to healthcare disparities within United States or across countries (e.g., Summers-Gabr, 2020; Weinstein et al., 2017). In terms of ethics, few would likely argue this is "good" or "right." But, practical solutions that make access to healthcare technologies equitable are few and far between. How will we attempt equitable distribution of AI systems in ABA as these systems become more common?

"Data sharing" is a topic from previous sections that is also relevant to the principle of *justice*. How should a client's decision to share their data affect their ability to access the resulting tools? As described in the *beneficence* and *nonmaleficence* section, if the client's specific data (or someone like them) is not used to create the AI system, the resulting models may not easily generalize to their specific case. This means that AI tools without a client's data may not benefit them or might even cause harm. But, from the *autonomy* section, we also cannot force people to share their data for building AI systems. How do we strike the balance between *beneficence/nonmaleficence* and *autonomy* to meet ethical claims around equitable access to the benefits of AI systems? And, perhaps

Table 4 List of Questions around the Ethical Use of AI Specific to the Ethical Principles of Justice from the Text

Principle	Topic	Question(s)
Justice	Shared Access to Benefits	What barriers exist to the equitable distribution of the benefits of AI systems? How should client choice with data sharing affect their ability to access the AI system? How can the obtained benefits to clients be observed and measured? What systems are needed to ensure the equitable distribution of benefits? What aspects to the current market economy distribution of technology might prevent equitable distribution of benefits from AI systems?
	Shared Burden of Cost	What barriers exist to the equitable distribution of the costs to build, deploy, and maintain AI systems? How can we estimate the total cost to build and maintain an AI system? What are the different ways potential benefactors can share that cost? How is the cost monitored and distributed to clients as they move in and out of the clinical system where AI is used? What aspects to the current market economy distribution of technology might prevent equitable distribution of costs of AI systems? How might behavior analysts generate creative solutions to incentivize healthcare providers to participate in building AI systems?
	Defining Equitable	What definition of equitable is most appropriate or most preferred by ABA practitioners and researchers as AI systems are developed? Under what conditions might different definitions of equitable be most appropriate? Once a definition is determined as best, who is responsible for maintaining the contingencies around meeting that definition?

most important, how can we observe and measure "equitable" in this context (Stewart & Napoles-Springer, 2003)?

Shared Burden of Cost

The flipside to equitable distribution of benefits is the equitable distribution of costs. As noted above, developing, deploying, monitoring, maintaining, and improving robust AI systems that behavior analysts in practice are likely to contact requires a significant amount of resources.[10] It would seemingly be unfair for a subset of people to bear the cost of these AI systems only to have others come in and reap the majority of the benefits. In analogy, it would seemingly be unfair if a single person were to spend several years building out a luscious garden in their backyard only to have their neighbors come and take all the food before the owner got any. In an equitable system, all neighbors would either contribute their time and abilities equally to build and develop the garden or to compensate the gardener for their work via payment or bartering. The same holds for the development, deployment, monitoring, maintenance, and improvement of AI systems for ABA.

Creating an equitable cost distribution among all potential benefactors of AI systems is no easy feat and raises many challenging questions. For example, how can the total cost to build and maintain an AI system be estimated? What are the different ways the potential benefactors can share that cost (e.g., data sharing, monetary compensation, beta testers)? How is that cost monitored and distributed as clients move into and out of the clinical system wherein the AI system is used? We do not pretend to have any answers here. The current system within much of the technology economy is for organizations with the skill set and data access to build an AI system that is then made available at a cost to potential consumers. That is, startups or existing technology companies bear the cost and risk to develop the product and then recoup that cost by selling the product in a market. But, as noted above, relying solely on this type of a system will likely allow only more affluent individuals to access the benefits of the AI system. What creative solutions might we come up with to ensure equitable distribution of costs so that access is more just? As a historical parallel, part of the role of HITECH was to incentivize the adoption of electronic health records given the benefits to patient care. How might behavior analysts generate creative solutions to incentivize healthcare providers to collaborate and participate in building AI systems?

[10] We appreciate that researchers can spin up neural networks for little-to-no cost and gather data from a few participants to train proof-of-concept models. However, it is seems unethical to deploy one-off AI systems built using small-sample sizes for broad access by practicing behavior analysts. This is because of the high probability of bias and low generalizability of such AI systems given the lack of representative population level patient characteristics used to train the AI system. Given this article is targeting AI systems being used by behavior analysts in practice, here we are referencing robust AI systems that meet the legal, ethical, and clinical specifications to be used broadly in all or most ABA settings where the populations are likely to differ from those characterizing the few participants in published research studies.

Defining Equitable[11]

To this point in this section, we have used the term equitable without really defining what that looks like. A final set of questions the field will likely need to come to consensus on is what definition(s) of equality we choose to use as we create ethical guidelines for the equitable use of AI in ABA. Though far from a complete list, there are at least six different ways we can talk about defining equality (e.g., Lamont & Favor, 2017). These definitions include: (1) to each person an equal share (i.e., strict egalitarianism); (2) to each person according to need (i.e., welfare-based principle); (3) to each person according to effort (i.e., one desert-based principle as in deserving); (4) to each person according to contribution (i.e., a second desert-based principle); (5) to each person according to merit (i.e., a third desert-based principle); and (6) to each person according to free-market exchanges (i.e., libertarian principle). How might the field of ABA practitioners rank these different principles? Further, it seems likely that different definitions will be most appropriate under different conditions. How might we identify the best use cases for each definition? How will equitable distribution of AI systems based on that definition be observed and monitored? And, who is responsible for maintaining those contingencies? Once we can come to an agreement on these questions (and others that arise), we can return to the AI ethics in healthcare literature to provide guidance for methods that adequately address equity issues when developing, building, and deploying AI technologies (e.g., Berdahl et al., 2023).

Limitations and Future Directions

AI ethics provides a starting point for the fair and respectful use of this new technology, but at least two additional challenges remain. The first challenge is that AI ethics currently "have no teeth." That is, there is no overarching system in place to maintain contingencies around adhering to, or violating, AI ethics. From a respondent and operant perspective, we know that simply creating guidelines is not sufficient to control behavior (e.g., Cleek & Leonard, 1998; Rességuier & Rodrigues, 2020; Shung, 2019; Somers, 2001). As a field, we will need to answer the many questions noted above and many additional questions to choose how we want AI to be ethically incorporated into ABA. Once answered, we will also have to identify how we will create contingencies to support adherence to those guidelines and reinforce behavior that moves forward this area of ethics and ABA.

The second challenge is that ethical guidelines are often too vague or too broad to readily apply to everyday situations (e.g., Brodhead et al., 2022; Kelly et al., 2020; Rainie et al., 2021). The United States and the European Union are attempting to directly address this challenge by developing and adopting detailed regulations around developing, deploying, monitoring, and maintaining AI systems (e.g., Broadbent & Arrieta-Kenna, 2021; Engler, 2022; Tähtien, 2022). However, it is important these regulations are written in a way that allows for flexible interpretation given the rapid advances of AI (Harris, 2021). Nonetheless, federal regulations and laws typically involve a standard of behavior that is considered a minimum standard compared to the aspirational or ideal behavior described with ethical guidelines. Thus, practitioners and researchers will likely need to extend such regulations to practically and ethically guide the everyday behavior of people building AI systems, as well as to guide the use of those systems in ABA. To safely and ethically maximize what AI has to offer requires planning and forethought.

Lastly, we focused on one specific area where we adapted common topics discussed in AI ethics as relevant to common principles that inform clinical–ethical decision making by behavior analysts in practice. As noted in the introduction, the fields of technology ethics, data ethics, and AI ethics are broad and deep. Further, we did not discuss in detail ethical topics specific to experimental and applied research on the development of AI systems in ABA. There are, no doubt, a host of likely questions and concerns specific to researchers working in this realm that also will likely need to be raised and addressed as a community.

Summary

AI is increasingly being used to improve the delivery of healthcare services, including in ABA or in use cases with direct relevance to ABA (Cox & Jennings, 2023). As with any other technology, AI could be used by humans in a manner deemed ethical or unethical. However, many important questions need to be answered before we can start to make claims about the ethical and unethical use of AI by practicing behavior analysts to deliver ABA services. In this article, we offered many starting questions that the field will likely need to answer around the ethical use of AI in ABA. While we work on answering these questions and developing guidelines, we should focus on emphasizing, and perhaps strengthening, our critical thinking skills as we make clinical decisions.

[11] For the reader looking to dive deeper into this area, these are sometimes referred to in the bioethics literature as "material principles of justice."

The considerations and questions described above are not an exhaustive list. The aim was to spark conversation so we can proactively make decisions about what it means to ethically use AI in ABA to make service delivery more efficient and to augment[12] the decision making of practicing behavior analysts (e.g., IBM Technology, 2021). Given that ethics are rules meant to guide groups of people, we believe answering these questions is best accomplished collaboratively as a community of professionals. We hope that in starting this conversation others will weigh in with suggestions for the next steps, whether that involves creating guidelines or establishing committees. Fortunately, we do not have to reinvent the wheel. The literature on AI ethics in healthcare has already begun to discuss many important issues that align with the principles of the Belmont Report (Office of the Secretary, 1979), the principlist approach to ethical decision making offered by Beauchamp and Childress (1979) and are applicable to ABA service delivery. We have the opportunity to start considering the many ethical questions that arise before AI is thoroughly embedded in ABA service delivery. Those who want to contribute to this dialogue can join the conversation here: https://endicott.qualtrics.com/jfe/form/SV_3CPDsGp37CxSblk.

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of Interest The authors have no conflicts of interest to disclose.

References

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics Decision Making, 20*, 310. https://doi.org/10.1186/s12911-020-01332-6

American Medical Association. (2022). Advancing health care AI through ethics, evidence, and equity. https://www.ama-assn.org/practice-management/digital/advancing-health-care-ai-through-ethics-evidence-and-equity

American Occupational Therapy Association. (2020). AOTA 2020 occupational therapy code of ethics. https://scota.net/resources/Documents/AOTA%202020%20Code%20of%20Ethics.pdf

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). http://www.apa.org/ethics/code/index.html

American Public Health Association. (n.d.). *Public health code of ethics*. https://www.apha.org/-/media/files/pdf/membergroups/ethics/code_of_ethics.ashx

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management, 62*, 102433. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*(1), 231–242. https://doi.org/10.1901/jeab.1974.22-231

Beauchamp, T. L., & Childress, J. F. (1979). *Principles of biomedical ethics*. Oxford University Press.

Behavior Analyst Certification Board. (2020). *Ethics code for behavior analysts*. https://bacb.com/wp-content/ethics-code-for-behavior-analysts/

Berdahl, C. T., Baker, L., Mann, S., Osoba, O., & Girosi, F. (2023). Strategies to improve the impact of artificial intelligence on health equity: Scoping review. *JMIR AI, 2*, e42936. https://doi.org/10.2196/42936

Bone, D., Goodwin, M. S., Black, M. P., Lee, C. C., Audhkhasi, K., & Narayanan, S. (2015). Applying machine learning to facilitate autism diagnostics: Pitfalls and promises. *Journal of Autism and Developmental Disorders, 45*(5), 1121–1136. https://doi.org/10.1007/s10803-014-2268-6

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition, 118*(1), 2–16. https://doi.org/10.1016/j.cognition.2010.10.004

Broadbent, M., & Arrieta-Kenna, S. (2021). AI regulation: Europe's latest proposal is a wake-up call for the United States. Center for Strategic and International Studies. https://www.csis.org/analysis/ai-regulation-europes-latest-proposal-wake-call-united-states

Brodhead, M. T., Cox, D. J., & Quigley, S. P. (2022). *Practical ethics for the effective treatment of autism spectrum disorder* (2nd ed.). Academic.

Byrd, G. D., & Winkelstein, P. (2014). A comparative analysis of moral principles and behavioral norms in eight ethical codes relevant to health sciences librarianship, medical informatics, and the health professions. *Journal of the Medical Library Association, 102*(4), 247–256. https://doi.org/10.3163/1536-5050.102.4.006

Cantin-Garside, K. D., Kong, Z., White, S. W., Antezana, L., Kim, S., & Nussbaum, M. A. (2020). Detecting and classifying self-injurious behavior in autism spectrum disorder using machine learning techniques. *Journal of Autism & Developmental Disorders, 50*(11), 4039–4052. https://doi.org/10.1007/s10803-020-04463-x

CentralReach. (2020). CentralReach acquires AI-based scheduling algorithm to automate scheduling operations for autism and ABA care delivery. https://centralreach.com/centralreach-acquires-ai-based-scheduling-algorithm-to-automate-scheduling-operations-for-autism-aba-care-delivery/

Cleek, M. A., & Leonard, S. L. (1998). Can corporate codes of ethics influence behavior? *Journal of Business Ethics, 17*, 619–630. https://doi.org/10.1023/A:1017969921581

Congressional Research Service. (2009). Summary: P. L. 111-5–The Health Information for Economic and Clinical Health Act. https://crsreports.congress.gov/product/pdf/R/R40161/9

Cox, D. J., & Jennings, A. M. (2023). The promises and possibilities of artificial intelligence in the delivery of behavior analytic services. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-023-00864-3

Cox, D. J., & Vladescu, J. C. (2023). *Statistics for applied behavior analysis practitioners and researchers*. Academic.

[12] Note, an emphasis should be placed on the word "augment," because we would argue (and believe readers would agree) that practicing behavior analysts should remain responsible for final treatment decisions and recommendations.

Cox, D. J., Syed, N., Brodhead, M. T., & Quigley, S. P. (2022). *Research ethics in behavior analysis: From laboratory to clinic and classroom*. Academic.

Cox, D. J., D'Ambrosio, D., Pagliaro, J., & RethinkFirst Data Team. (2023). An artificial intelligence driven system to predict ASD outcomes in ABA. OSF Preprints. https://osf.io/3t9zc/

Davidai, S., Gilovich, T., & Ross, L. D. (2012). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences, 109*(18), 15201–15205. https://doi.org/10.1073/pnas.1211695109

Dholakia, U. (2021). The ethical quandary of default opt-ins. *Psychology Today*. https://www.psychologytoday.com/us/blog/the-science-behind-behavior/202104/the-ethical-quandary-default-opt-ins

Diakopoulos, N. (2020). Transparency. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 197–213). Oxford University Press.

Driver, J. (2014). *The history of utilitarianism*. Stanford University Press.

Dubber, M. D., Pasquale, F., & Das, S. (2020). *The Oxford handbook of ethics of AI*. Oxford University Press.

Dufour, M. M., Lanovaz, M. J., & Cardinal, P. (2020). Artificial intelligence for the measurement of vocal stereotypy. *Journal of the Experimental Analysis of Behavior, 114*(3), 368–380. https://doi.org/10.1002/jeab.636

Dunbar, N. E., Miller, C. H., Adame, B. J., Elizondo, J., Wilson, S. N., Lane, B. L., Kaufmann, A. A., Bessarabova, E., Jensen, M. L., Straub, S. K., Lee, Y.-H., Burgoon, J. K., Valacich, J. J., Jenkins, J., & Zhang, J. (2014). Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game. *Computers in Human Behavior, 37*, 307–318. https://doi.org/10.1016/j.chb.2014.04.053

Edelstein, L. (1943). *The Hippocratic oath: Text, translation and interpretation*. The Johns Hopkins Press.

Engler, A. (2022). The EU and U.S. are starting to align on AI regulation. Brookings Institution. https://www.brookings.edu/blog/techtank/2022/02/01/the-eu-and-u-s-are-starting-to-align-on-ai-regulation/

Erden, Y. J., Hummerstone, H., & Rainey, S. (2021). Automating autism assessment: What AI can bring to the diagnostic process. *Journal of Evaluation in Clinical Practice, 27*(3), 485–490. https://doi.org/10.1111/jep.13527

Evans, R., & Gao, J. (2016). DeepMind AI reduces Google data centre cooling bill by 40%. *DeepMind*. https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

European Commission. (2021). Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Fasching, J., Walczak, N., Toczyski, W. D., Cullen, K., Sapiro, G., Morellas, V., & Papanikolopoulos, N. (2013). Assisted labeling of motor stereotypies in video. [Poster presentation]. American Academy of Child and Adolescent Psychiatry, 60th Annual Meeting, Orlando, FL, United States.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Food & Drug Administration. (2022). Computer-assisted surgical systems. https://www.fda.gov/medical-devices/surgery-devices/computer-assisted-surgical-systems

Garcez, A. A., & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd wave. *arXiv*. https://doi.org/10.48550/arXiv.2012.05876

Garnelo, M., & Shanahan, M. (2019). Reconciling deep learning with symbolic artificial intelligence: Representing objects and relations. *Current Opinion in Behavioral Sciences, 29*, 17–23. https://doi.org/10.1016/j.cobeha.2018.12.010

Gasser, U., & Schmitt, C. (2020). The role of professional norms in the governance of artificial intelligence. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of ai* (pp. 141–160). Oxford University Press.

Goel, A. K. (2021). Looking back, looking ahead: Symbolic versus connectionist AI. *AI Magazine, 42*, 83–85. https://doi.org/10.1609/aaai.12026

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Harris, J. (2021). AI advances, but can the law keep up? Towards Data Science. https://towardsdatascience.com/ai-advances-but-cat-the-law-keep-up-7d9669ce9a3d

Haugeland, J. (1985). *Artificial intelligence: The very idea*. MIT Press.

Health Information Technology for Economic and Clinical Health Act. 2009 Pub. L. No. 111-5, § 13001, 123 Stat.227.

Health Insurance Portability and Accountability Act. 1996 Pub. L. No. 104-191, § 264, 110 Stat.1936.

IBM. (2021). AI ethics. https://www.ibm.com/cloud/learn/ai-ethics

IBM Technology. (2021). What is AI ethics? [Video]. YouTube. https://www.youtube.com/watch?v=aGwYtUzMQUk

Institute for Ethical AI in Education. (2020). Interim report: Towards a shared vision of ethical AI in education. https://tinyurl.com/4c5xuye2

Institute for Quality & Efficiency in Health Care. (2020). *Depression: How effective are antidepressants?* InformedHealth.org - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK361016/

Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, framing and privacy: Why opting in-opting out. *Marketing Letters, 13*(1), 5–15. https://doi.org/10.1023/A:1015044207315

Kautz, H. (2022). The third AI summer: AAAI Robert S. Engelmore Memorial Lecture. *AI Magazine, 43*(1), 105–125. https://doi.org/10.1002/aaai.12036

Kelly, E. M., Greeny, K., Rosenberg, N., & Schwartz, I. (2020). When rules are not enough: Developing principles to guide ethical conduct. *Behavior Analysis in Practice, 14*(2), 491–498. https://doi.org/10.1007/s40617-020-00515-x

Klosowski, T. (2021). The state of consumer data privacy laws in the U.S. (and why it matters). *The New York Times*. https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/

Kohavi, R., & Wolpert, D. H. (1996). Bias plus variance decomposition for zero-one loss functions. ICML, 96. http://robotics.stanford.edu/~ronnyk/biasVar.pdf

Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry, 5*, e514. https://doi.org/10.1038/tp.2015.7

Lamont, J., & Favor, C. (2017). Distributive justice. *The Stanford encyclopedia of philosophy*. Stanford University Press. https://plato.stanford.edu/entries/justice-distributive/

Lanovaz, M. J., & Hranchuk, K. (2021). Machine learning to analyze single-case graphs: A comparison to visual inspection. *Journal of Applied Behavior Analysis, 54*(4), 1541–1542. https://doi.org/10.1002/jaba.863

Lanovaz, M. J., Giannakakos, A. R., & Destras, O. (2020). Machine learning to analyze single-case data: A proof of concept. *Perspectives on Behavior Science, 43*(1), 21–38. https://doi.org/10.1007/s40614-020-00244-0

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods & Programs in Biomedicine, 226*, 107161. https://doi.org/10.1016/j.cmpb.2022.107161

Marchant, J. (2020). Powerful antibiotics discovered using AI: Machine learning spots molecules that work even against "untreatable" strains of bacteria. *Nature*. https://doi.org/10.1038/d41586-020-00018-3

Martinho, A., Kroesen, M., & Chorus, C. (2021). A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence. *Artificial Intelligence in Medicine, 121*, 102190. https://doi.org/10.1016/j.artmed.2021.102190

Matheson, R. (2019). Bringing human-like reasoning to driverless car navigation. *MIT News*. https://news.mit.edu/2019/human-reasoning-ai-driverless-car-navigation-0523

Mazur, J. E. (1981). Optimization theory fails to predict performance of pigeons in a two- response situation. *Science, 214*(4522), 823–825. http://www.jstor.org/stable/1686991

McCoy, L. G., Brenna, C. T., Chen, S. S., Vold, K., & Das, S. (2022). Believing in black boxes: Machine learning for healthcare does not need explainability to be evidence-based. *Journal of Clinical Epidemiology, 142*, 252–257. https://doi.org/10.1016/j.jclinepi.2021.11.001

McGuire, A. L., Roberts, J., Aas, S., & Evans, B. J. (2019). Who owns the data in a medical information commons? *Journal of Law, Medicine & Ethics, 47*(1), 62–69. https://doi.org/10.1177/1073110519840485

Mörch, C. M., Gupta, A., & Mishara, B. L. (2020). Canada protocol: An ethical checklist for the use of artificial intelligence in suicide prevention and mental health. *Artificial Intelligence in Medicine, 108*, 101934. https://doi.org/10.1016/j.artmed.2020.101934

Müller, A., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.

Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics, 22*(1), 1–5. https://doi.org/10.1186/s12910-021-00687-3

Najibi, A. (2020). Racial discrimination in face recognition technology. *Harvard University Blog on Science Policy, Special Edition: Science Policy & Social Justice*.https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology

National Institute of Standards & Technology (NIST). (2019). Face recognition vendor test (FRVT) Part 3: Demographic effects. https://doi.org/10.6028/NIST.IR.8280

Neal, B. (2019). On the bias-variance tradeoff: Textbooks need an update. *arXiv*. https://doi.org/10.48550/arXiv.1912.08286

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19*(3), 113–126. https://doi.org/10.1145/360018.360022

Normand, M. P., & Donohue, H. E. (2023). Research ethics for behavior analysts in practice. *Behavior Analysis in Practice, 16*(1), 13–22. https://doi.org/10.1007/s40617-022-00698-5

Odwazny, L. M., & Berkman, B. E. (2017). The "reasonable person" standard for research informed consent. *American Journal of Bioethics, 17*(7), 49–51. https://doi.org/10.1080/15265161.2017.1328540

Office of the Secretary. (1979). National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research–*The Belmont Report: Ethical principles and guidelines for the protection of human research*. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

OpenAI. (2023). ChatGPT [Large language model]. https://chat.openai.com/chat

Oracle. (n.d.). What is a digital assistant? https://www.oracle.com/chatbots/what-is-a-digital-assistant/

Patel, A. A. (2019). *Hands-on unsupervised learning using Python: How to build applied machine learning solutions from unlabeled data*. O'Reilly Media.

Plotz, T., Hammerla, N. Y., Rozga, A., Reavis, A., Call, N., & Abowd, G. D. (2012). Automatic assessment of problem behavior in individuals with developmental disabilities. *UbiComp '12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 391–400). https://doi.org/10.1145/2370216.2370276

Rainie, L., Anderson, J., & Vogels, E. A. (2021). Worries about developments in AI. Pew Research Center. https://www.pewresearch.org/internet/2021/06/16/1-worries-about-developments-in-ai/

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society, 7*(2), 1–5. https://doi.org/10.1177/2053951720942541

Rimland, B., & Edelson, S. M. (1999). Autism treatment evaluation checklist (ATEC). *APA PsycTests*. https://doi.org/10.1037/t03995-000

Roach, J. (2018). Microsoft improves facial recognition technology to perform well across all skin tones, genders. *The AI Blog*. https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/#:~:text=Microsoft%20announced%20Tuesday%20that%20it,recognize%20gender%20across%20skin%20tones.&text=With%20the%20new%20improvements%2C%20Microsoft,by%20up%20to%2020%20times.

Sacasas, L. M. (2018). Does technology evolve more quickly than ethical and legal norms? The Frailest Thing. https://tinyurl.com/2p8f7aky

Sadler, M., & Regan, N. (2019). *Game hanger: AlphaZero's groundbreaking chess strategies and the promise of AI*. New in Chess.

Schneier, B. (2019). We must bridge the gap between technology and policy making. Our future depends on it. World Economic Forum. https://www.weforum.org/agenda/2019/11/we-must-bridge-the-gap-between-technology-and-policy-our-future-depends-on-it/

Shung, K. P. (2019). Artificial intelligence and ethics: Part I. *Medium*. https://medium.com/@koolanalytics/artificial-intelligence-ethics-part-1-523eb06e04eb

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review, 52*(5), 270–277. https://doi.org/10.1037/h0062535

Somers, M. J. (2001). Ethical codes of conduct and organizational context: A study of the relationship between codes of conduct, employee behavior and organizational values. *Journal of Business Ethics, 30*, 185–195. https://doi.org/10.1023/A:1006457810654

Song, D. Y., Kim, S. Y., Bong, G., Kim, J. M., & Yoo, H. J. (2019). The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: A literature review. *Journal of the Korean Academy of Child & Adolescent Psychiatry, 30*(4), 145–152. https://doi.org/10.5765/jkacap.190027

Stewart, A. L., & Napoles-Springer, A. M. (2003). Advancing health disparities research: Can we afford to ignore the measurement issues? *Medical Care, 41*(11), 1207–1220. https://www.jstor.org/stable/3768410

Summers-Gabr, N. M. (2020). Rural-urban mental health disparities in the United States during COVID-19. *Psychological Trauma: Theory, Research, Practice, & Policy, 12*(S1), S222–S224. https://doi.org/10.1037/tra0000871

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Bradford.

Tähtien, S. (2022). What is the EU's artificial intelligence act and what will it change? *Towards Data Science*. https://towardsdatascience.com/what-is-the-eus-artificial-intelligence-act-and-what-will-it-change-b1f6812f5dd5

Taylor, T., & Lanovaz, M. J. (2022). Agreement between visual inspection and objective analysis methods: A replication and extension. *Journal of Applied Behavior Analysis, 55*(3), 986–996. https://doi.org/10.1002/jaba.921

Turgeon, S., & Lanovaz, M. J. (2020). Tutorial: Applying machine learning in behavioral research. *Perspectives on Behavior Science, 43*(4), 697–723. https://doi.org/10.1007/s40614-020-00270-y

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. http://links.jstor.org/sici?sici=0036-8075%2819740927%293%3A185%3A4157%3C1124%3AJUUHAB%3E2.0.CO%3B2-M

Viter, I. (2019). The future of autonomous driving with artificial intelligence. *Medium*. https://medium.com/swlh/the-future-of-autonomous-driving-with-artificial-intelligence-4aa2a85e8072

Weinstein, J. N., Geller, A., Negussie, Y., & Baciu, A. (2017). *Communities in action: Pathways to health equity*. National Academies Press.

Wilber, J., & Werness, B. (2021). The bias variance tradeoff. MLU-EXPLAIN. https://mlu-explain.github.io/bias-variance/

Zialcita, P. (2019). Facebook pays $643,000 fine for role in Cambridge Analytica scandal. NPR. https://www.npr.org/2019/10/30/774749376/facebook-pays-643-000-fine-for-role-in-cambridge-analytica-scandal