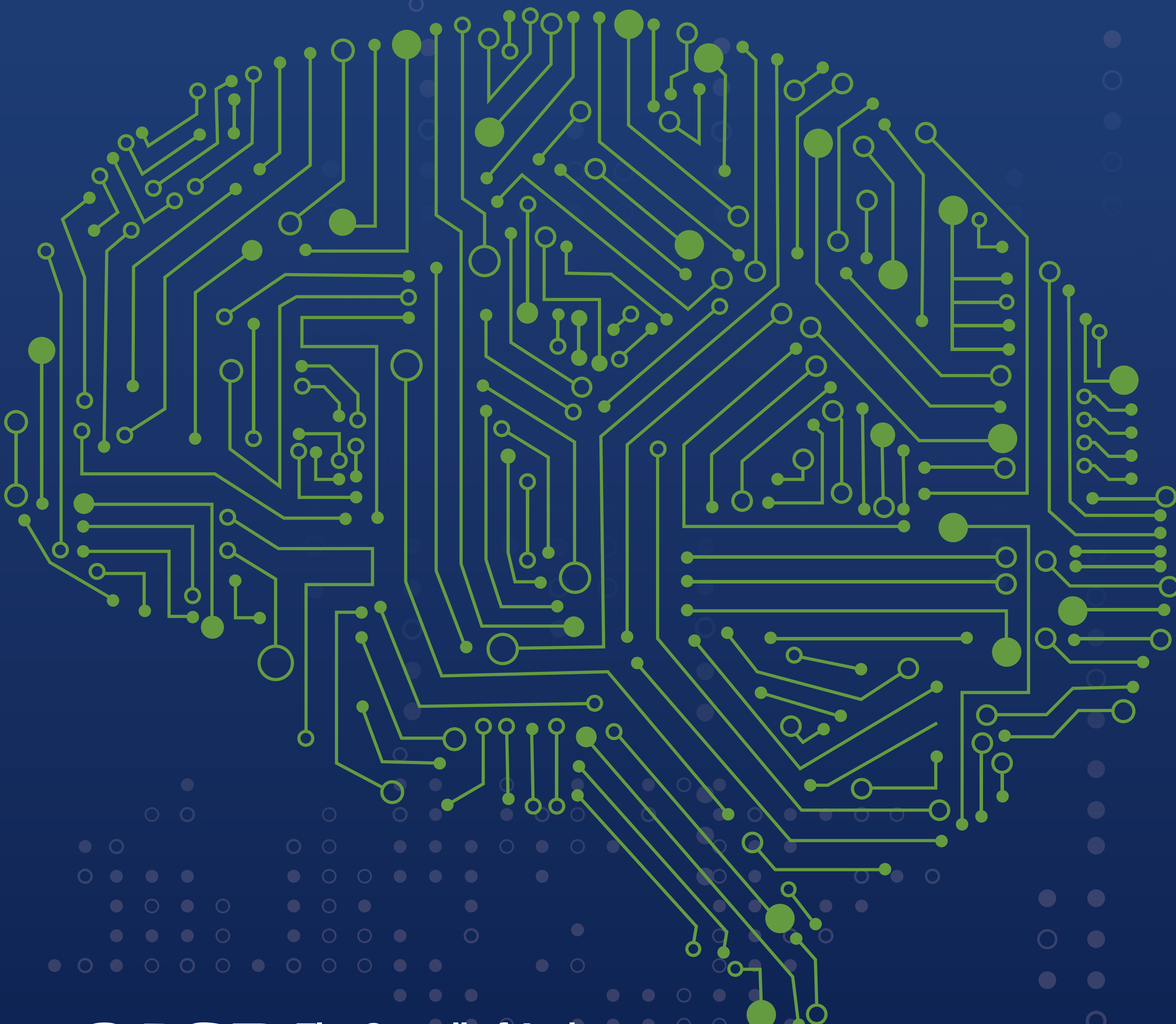


CASP The Council of Autism Service Providers

PRACTICE PARAMETERS FOR ARTIFICIAL INTELLIGENCE USE IN APPLIED BEHAVIOR ANALYSIS

TABLE OF CONTENTS

ACKNOWLEDGMENTS

THE COUNCIL FOR AUTISM SERVICE PROVIDERS (CASP) expresses its sincere appreciation to prior and current members of the AI Community Work Group for their invaluable insights in developing this resource.

Project Leaders

Summer Gainey, *Above and Beyond Therapy*

Missy Olive, *Florida Association for Behavior Analysis*

Rebecca Womack, *RAW Consulting Solutions*

Project Editors

David Cox, *Mosaic Pediatric Therapy*

Summer Gainey, *Above and Beyond Therapy*

Adam Hahs, *Caravel Autism Health*

Missy Olive, *Florida Association for Behavior Analysis*

Rebecca Womack, *RAW Consulting Solutions*

Content Contributors

T.J. Couey, *ABA Advocates*

David Cox, *Mosaic Pediatric Therapy*

Allison Frazier, *Elevate Learning*

Summer Gainey, *Above and Beyond Therapy*

Adam Hahs, *Caravel Autism Health*

Paul Heering, *Bierman Autism Centers*

Missy Olive, *Florida Association for Behavior Analysis*

Joseph Smith, *EMG Consulting*

Randall Swanson, *ABA Pathways*

Alexandra Tomei, *BlueSprig Pediatrics*

Mike Wasmer, *Council of Autism Service Providers*

Rebecca Womack, *RAW Consulting Solutions*

PREFATORY NOTES

THESE GUIDELINES ADDRESS ARTIFICIAL INTELLIGENCE (AI) and other advancements that use computer systems to mimic human reasoning and thinking. Throughout the document, we use the term "AI" to encompass Artificial Intelligence as well as various applications related to its use.

The designation of Board Certified Assistant Behavior Analyst® (BCaBA®), Board Certified Behavior Analyst® (BCBA®), Board Certified Behavior Analyst — Doctorate®(BCBA-D®) are given to those who have received the proper education, training, and supervision and obtained the required experience necessary to pass an exam from the Behavior Analyst Certification Board® (BACB®). In these guidelines, they are collectively referred to as "behavior analysts."

When these guidelines refer to "paraprofessionals," we're describing technicians and behavior technicians who undergo training, education, and supervision to be able to deliver one-to-one services under the supervision of a behavior analyst. A Registered Behavior Technician®(RBT®) is a person who has completed the required education, training, and competency assessment and passed an exam from the BACB.

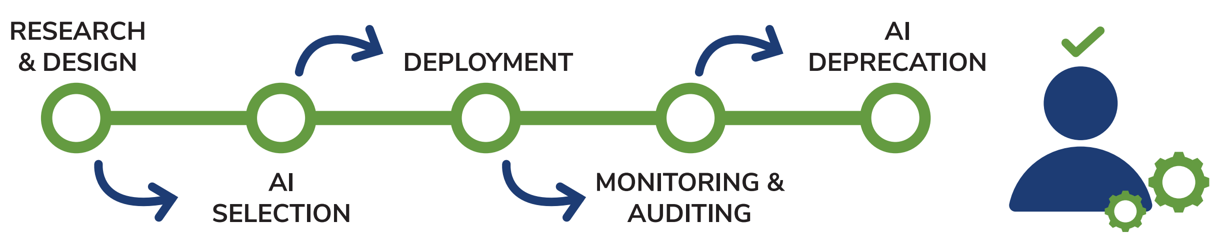
Behavior analysts, paraprofessionals, and RBTs may be collectively referred to as "provider(s)" when applicable.

Companies that create, design, develop, and/or sell AI technologies to an ABA organization are referred to as "vendors." Employees or contract workers tasked with creating, designing, or developing AI specific to the needs and circumstances of a particular ABA organization (and not for resale to the public) will be referred to as "developers." Some organizations may use both for implementing AI.

The terms "practice guidelines" and "generally accepted standards of care (GASC)" are used interchangeably throughout this document.

When we say "these guidelines," we're referring to the content of this document.

Finally, when we address "you," we're referring to both you the reader and your organization.



PREAMBLE

USING ARTIFICIAL INTELLIGENCE (AI) in applied behavior analysis (ABA) presents opportunities and challenges that need careful consideration. As AI continues to evolve, its application in behavior analysis must align with evidence-based practices, uphold the integrity of rigorous clinical decision making, and prioritize ethical service delivery to ensure beneficial outcomes for individuals receiving care. Given the increasing prevalence of AI in both clinical and administrative settings, organizations providing ABA services must adopt structured governance practices that maintain compliance with federal and state regulations, fulfill obligations to funding sources, and ensure responsible oversight.

Purpose

This document establishes clear guidelines for provider organizations to navigate AI integration while preserving the principles of behavior analysis and regulatory accountability. It outlines best practices for AI system selection, deployment, and ongoing monitoring to safeguard clinical efficacy and mitigate risks associated with automation. Additionally, it details auditing mechanisms, error reporting protocols, and criteria for the responsible deprecation of outdated AI systems—ensuring that AI-driven processes uphold ethical standards and maintain operational integrity. By adhering to these guidelines, organizations can foster transparency, mitigate potential risks, and optimize AI's role in enhancing both clinical and administrative excellence within ABA services.

Scope

These guidelines apply to AI use within ABA organizations, including but not limited to clinical decision making, administrative operations, data management, and service documentation. We emphasize safeguards to uphold clinical integrity, protect client welfare, and ensure adherence to evidence-based practices.

This framework outlines governance strategies to mitigate automation-related risks, ensuring AI systems operate transparently under organizational oversight. It covers topics such as practitioner responsibilities, AI data verification, legal and ethical compliance, and the limitations of AI in ABA.

By defining these parameters, provider organizations can responsibly integrate AI while maintaining ethical and regulatory standards, as well as fostering innovation without compromising service quality or efficacy.

Legal Disclaimer

This guide is for informational purposes only and does not constitute legal, regulatory, or professional advice. While efforts have been made to ensure accuracy, the authors, publishers, and contributors make no warranties regarding its completeness or applicability to any specific legal or business environment. CASP does not endorse any AI tool, software, or application; any references are illustrative only.

AI use in professional services is subject to evolving laws, regulations, and ethical considerations that vary by jurisdiction. Readers should consult qualified legal, compliance, and technical professionals before implementing AI solutions.

These guidelines do not establish any advisory relationship. The authors, publishers, and contributors disclaim liability for any damages resulting from reliance on its content. By using these guidelines, you assume full responsibility for any actions based on its information.

For specific legal or regulatory guidance, consult a licensed professional.

ANNOTATED GLOSSARY OF TERMS

THE PURPOSE OF THIS SECTION IS TO ESTABLISH CLEAR, COMPREHENSIVE, ADAPTABLE DEFINITIONS and frameworks for the use of AI and its subsets within the field of ABA. This section provides structured definitions that reflect AI's existing and emerging functionalities in ABA while also acknowledging the rapid evolution of technology.

Artificial Intelligence (AI) A field of computer science that simulates intelligence (often human intelligence) in machines. AI systems can analyze data, learn from interactions, and make decisions, enabling automation and predictive insights across various domains (Artificial Intelligence Definitions, 2025).

Example: Using technology to assist in the diagnosis of autism by tracking eye contact.

Algorithm A step-by-step procedure or a set of rules followed by a computer to solve a problem or accomplish a task. In the context of AI, algorithms define the processes used to analyze data, recognize patterns, and train a machine learning model.

Example: The set of rules that Netflix uses to provide recommendations based on viewing history.

Bias in AI Systematic errors in AI system outputs that arise when models reflect the limitations or imbalances in their training data, leading to unfair or skewed decisions. Bias is nearly impossible to eliminate entirely from any model. However, inappropriate AI system use can be mitigated through transparency around training data and model performance for subsets of the training data.

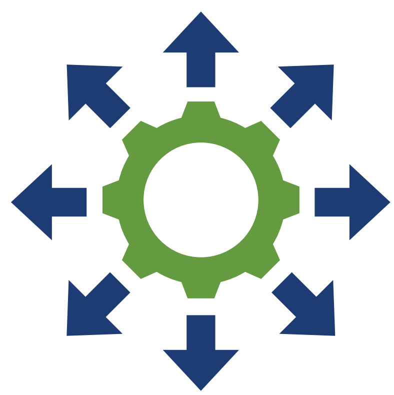
Example: A tool that is trained from characteristics from limited demographics may favor those demographics when screening resumes, leading to biased hiring decisions.

Chatbot A software application that enables conversation that is intended to mimic human interactions.

Example: In customer service, chatbots are used to answer commonly asked questions or direct customers to the appropriate resource.

Clinical Decision Support Systems (CDS) Clinical decision support (CDS) systems provide clinicians, staff, patients, and other individuals with knowledge and person-specific information that is intelligently filtered and presented at appropriate times to enhance health and healthcare (Berner, 2009).

Example: An effective CDS system may suggest alternative interventions or treatment plan changes if it detects that a patient's tantrums are not decreasing despite the interventions being implemented



Cognitive Offloading The use of external tools or systems—such as AI—to reduce mental effort by delegating tasks such as remembering, calculating, or decision making.

Example: A BCBA uses a smartphone to set a reminder for an upcoming appointment instead of trying to remember it, enabling them to focus on other activities.

Computer Vision A field that enables computers to interpret data derived from videos or photographs, mimicking the function of human vision.

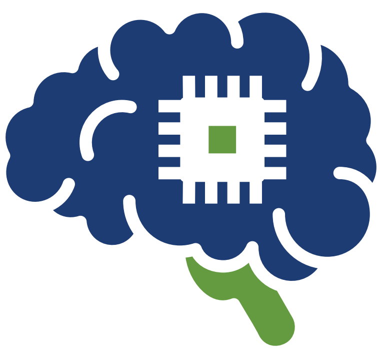
Example: In social media, computer vision is used to identify and automatically tag people in photos.

Confabulations An AI generation that is incorrect because it reconstructs data by filling in the gaps based on similar information (*Why Confabulation, Not Hallucination, Defines AI Errors*, n.d.).

Example: A program meant to assist in "filling in the blanks" of a goal you're writing but does so with the wrong target skill as a recommendation.

Data Model A blueprint or visual representation that defines how data elements are organized, standardized, and managed based on how they relate to one another and to the properties of real-world entities (DAMA International, 2017; What Is Data Modeling?, 2024). Data modeling is the strategy used to store and organize data in a way that data users understand.

Example: Within an address book, phone numbers are related to contact names and addresses and are organized alphabetically. Following this data model makes it easy for address book users to access and use the information.



Deep Learning A subset of machine learning that uses artificial neural networks with multiple layers to model complex patterns in data. These deep architectures enable the automatic extraction of features from raw data, facilitating tasks such as image and speech recognition. The term "deep" refers to the numerous layers through which data are processed, allowing for the learning of hierarchical representations (LeCun, Bengio, & Hinton, 2015). Deep learning can be contrasted with classical machine learning.

Example: In facial recognition, deep learning algorithms are used to recognize individuals in images and videos.

Generally Accepted Standards of Care (GASC) Clinical practices for serving patients with a specified condition recognized by healthcare providers in the relevant clinical specialty (CASP, 2025).

Example: CASP's *Applied Behavior Analysis Practice Guidelines for the Treatment of Autism Spectrum Disorder* are the generally accepted standards of care for providing ABA services to individuals with autism

Generative AI A subfield of artificial intelligence focused on generating new data samples that resemble a given training dataset.

Example: ChatGPT, Claude, and Gemini.

Hallucinations Instances where a generative AI model—particularly large language models (LLMs) or vision models—produces outputs that are factually incorrect, nonsensical, or fabricated (Sun et al., 2024).

Example: Providing a list of scholarly references where the authors, titles, or journals do not exist.

Labeled Data Raw data that have been assigned specific tags or labels, providing context and meaning. Labels allow machine learning models to learn how other features in the dataset combine to result in the assigned tag or label (IBM, 2025).

Example: Labeling: "car" in an image, "positive sentiment" in a text review, "dog barking" in an audio recording.

Large Language Model A generative AI model designed to process, interpret, and probabilistically generate human-like text based on received input and leveraging vast amounts of training data to learn statistical relations in language patterns.

Example: OpenAI GPT-4, Claude, Perplexity.

Machine Learning (ML) A subfield of AI where feedback is iteratively provided based on a computer system's output such that the system improves without being explicitly taught how to do so. Large amounts of data are generally necessary for these improvements to occur (Artificial Intelligence Definitions, 2025).

Example: Product suggestions based on browsing and purchasing history. Products are suggested, the system gets feedback on whether you buy the product, and the system iteratively learns to make better product suggestions over time.

Medical Necessity Medical necessity refers to healthcare services or supplies needed to prevent, diagnose, or treat an illness, injury, condition, disease or its symptoms and that meet generally accepted standards of care (Center for Medicare and Medicaid Services).

Example: [ABA] services must be medically necessary to ameliorate symptoms of a diagnosed disorder, build adaptive behaviors, and/or reduce maladaptive behaviors to enhance the patient's health, safety, and overall functioning and/or to prevent deterioration or regression (ABA Coding Coalition, 2022).

Model Alignment How well an AI system's output and underlying algorithmic processes reflect the values, goals, and intended use cases defined by its human stakeholders. This is especially important in sensitive domains such as healthcare and education. It's standard practice for model developers to have a model alignment phase *after* model training and *before* the system is made available for public consumption (e.g., Hendrycks et al., 2021; Shen et al., 2023).

Example: A medical diagnosis AI provides treatment recommendations consistent with clinical guidelines and patient safety priorities.

Natural Language Processing (NLP) A subfield of AI that enables computer programs to analyze, interpret, and generate human languages in a valuable way, utilizing computational and statistical linguistic methods. (Chowdhary, 2020)

Example: A clinician takes in trial-based data, and an algorithm processes the data—converting it into narrative session notes, which will then be vetted by the clinician for accuracy.

Performance Drift The gradual change in an AI system's accuracy or reliability over time due to changes in data patterns, environments, or user behavior that differ from the conditions under which the system was originally designed.

Example: A spam email filter becoming less accurate over time as new types of spam emails emerge that differ from the patterns it was trained to detect.

Prompt Engineering The practice of crafting precise, effective input instructions to guide an AI system's output toward a desired format, style, solution, or content. Prompt engineering is often an iterative process rather than a one-off behavior.

Example: A clinician writes a prompt like, "Summarize this session note in two sentences, highlighting behavioral improvements and any barriers encountered" to ensure the AI system produces concise and relevant summaries for documentation.

Software as a Service (SaaS) Software integration via web browser—versus a download to your computer/device—whereby the host maintains updates, security, and maintenance.

Example: Dropbox, Zoom, Gmail, and Microsoft Teams.

Structured Data Set A dataset organized in a predefined format, typically stored in tables with clearly defined rows and columns. Structured datasets follow a consistent schema outlined in a data model, making them easily searchable and analyzable using databases and statistical methods.

Example: A medical records database where each patient's information (e.g., name, age, diagnosis, treatment history) is stored in a structured table format (e.g., patients as the rows, information as the columns)—allowing for efficient querying and analysis.

Supervised Machine Learning (SML) SML is a type of machine learning where the model is trained on labeled data. The algorithm learns from a dataset containing input-output pairs (features and corresponding labels). The goal is to predict the output for unseen inputs (Mitchell, 1997)

Example: Using a model to predict the likelihood of a specific behavior occurring based on previous data and environmental factors, such as predicting when a child might engage in challenging behavior based on their history and current context.

Unstructured Data Set A dataset that lacks a predefined format or organizational structure, often consisting of raw text, audio, images, or video files. Unlike structured datasets, unstructured data doesn't fit neatly into rows and columns, requiring specialized techniques for processing and analysis.

Example: A collection of ABA session videos or conversational transcripts, where the videos or conversations are stored as .wav files or free-text responses without a consistent structure, making them more challenging to analyze without computer vision or natural language processing techniques.

Unsupervised Machine Learning (UML) UML is a type of machine learning where the model works with unlabeled data. The algorithm identifies patterns, structures, or groupings within the data without explicit instructions on what to look for (Hastie, Tibshirani, & Friedman, 2009)

Example: A behavior analyst collects session data on many patients, including the frequency and duration of behavior, environmental contexts, and intervention strategies used. UML is then used to group patients based on similar behavioral patterns and response to interventions to help individualize and refine intervention plans.

IN ORDER TO EFFECTIVELY SELECT, DEPLOY, AND IMPLEMENT AI, IT'S IMPERATIVE TO UNDERSTAND THE MANY VARIABLES external to an ABA organization that impact the developer, vendor, and end user. This section includes basic information leaders should know about AI prior to selecting and deploying it. These guidelines aren't exhaustive. They're also not intended to supplant comprehensive guidance that organizations may need for their unique circumstances.

Federal and State Laws

In order to recognize the ethical risks arising from AI use, it's important to understand the current legal landscape and trends that will impact the AI industry for the foreseeable future. At the writing of these guidelines, both state and federal laws regarding AI use in healthcare are continually evolving. Historically, state laws have differed regarding which entities can be held legally liable as "healthcare providers," with some omitting AI vendors (Mello & Guha, 2024). Additionally, some states have upheld licensing agreements that designate healthcare providers as being solely responsible for accurate clinical interpretation of the AI output (Mello & Guha, 2024). While state laws struggle to keep pace with technological developments, companies like Bryan Cave Leighton Paisner (BLCP), a global law firm, have developed tracking systems to update every state's residents on laws and regulations affecting AI development and deployment.

AI use in professional and healthcare services is subject to evolving laws, regulations, and ethical considerations that vary across states and jurisdictions. Leaders of ABA organizations are strongly encouraged to consult with qualified legal, compliance, and technical professionals before implementing AI solutions. They should also remain current on evolving laws and regulations impacting AI development and deployment.

Obligations to Funding Sources

At the writing of these guidelines, most healthcare funding sources don't address AI for service delivery in their ABA policies. It's unclear why. Regardless, ABA organizations shouldn't assume that payers will permit or reimburse every use of AI by practitioners. Some ABA policies describe clinical activities with broad language, while others are very specific. For example,

__At the writing of these guidelines, most healthcare funding sources don't address AI for service delivery in their ABA policies. It's unclear why. Regardless, ABA organizations shouldn't assume that payers will permit or reimburse every use of AI by practitioners.__

Clinicians should ensure any AI-powered work product aligns with the principles of behavior analysis and reflects individualized treatment decisions. AI should serve as a tool to support clinical work, not supplant it.

one policy may say: "… [treatment plans] should be completed" (Blue Cross and Blue Shield of Oklahoma, 2023). Another may say: "Data from treatment targets are most often collected by the behavior technician and analyzed by the behavior analyst on a regular basis" and specify the typical activities associated with each provider type (Cigna, 2025).

It's important to review each payer's policy to determine its requirements (if any) for AI use. If it's not specified in the policy—and you intend to use AI for activities that are billable directly (e.g., treatment plan writing) or indirectly (i.e., bundled into the rate of the service; e.g., writing session notes)—then it's best to provide the payer a transparent description of your intended use. Organizations shouldn't assume that payers will reimburse for AI use in ABA services just because they reimburse for it in other healthcare sectors (e.g., radiology).

Ethics

All providers holding a BACB credential are responsible for understanding and adhering to the BACB's Ethics Code for Behavior Analysts (2020) (the "Code") and the RBT Ethics Code 2.0 (2021). These codes govern every aspect of clinical practice relevant to certification, including the appropriate and ethical use of AI by clinicians (BACB, July 2024). While AI doesn't receive its own section in the code, you need to consider how AI use applies to every relevant area, including but not limited to ethical principles, professional obligations, service delivery, supervision, and research.

For example, obtaining informed consent to assess and treat is an essential component of ABA service delivery. If an organization chooses to use AI, they must have a thorough understanding of how client data entered into AI systems are stored, analyzed, and utilized. In this scenario, informed consent may provide the client with visibility into how AI is used, its risks and benefits, and the opportunity to decline its use in services at any time (BACB, 2020). It's important for organizations to carefully consider all the different features of AI, the ways this may impact the client, and how the ethics code applies prior to deployment.

The practitioner (i.e., BCBA, RBT) is responsible for clinical documentation. This includes full clinical responsibility and requires the practitioner to prevent harm; document the most effective, individualized treatment recommendations; and include evidence-based practices (BACB, 2020). Clinicians should ensure any AI-powered work product aligns with the principles of behavior analysis and reflects individualized treatment decisions. AI should serve as a tool to support clinical work, not supplant it. Clinicians should continue to rely on their training and expertise in order to follow the profession of ABA's GASC.

With any AI use, behavior analysts should ensure the BCBAs and RBTs they supervise receive proper education, training, and oversight. This may include, but is not limited to, establishing clear policies around the integration of AI in clinical work (e.g., session note documentation, recording data specific to a client's behavior) and delineating acceptable and unacceptable uses of technology. Behavior analysts should also regularly revisit the topic of appropriate, ethical AI use with their RBTs and BCBAs— ensuring they have the most current information about systems that are presently being used or will be in the future.

As with all forms of professional writing, maintaining integrity is essential. Writers are responsible for avoiding plagiarism and ensuring appropriate credit is given to original sources (Publication Manual of the American Psychological Association, 7th Ed., 2019). When using AI-generated content, carefully review and verify originality to avoid plagiarism. This extends to verifying that documentation reflects the specific encounter and does not reuse text from other sources or any proprietary information without permission from the intellectual property holder or owner. By incorporating their own clinical judgment, observations, and patient-specific details, clinicians can maintain the integrity and authenticity of their medical documentation. Further, transparency in the use of AI is an important aspect of professional integrity. Clinicians are expected to clearly disclose the use of AI in any work product where it has contributed to the content.

In addition to adhering to ethics codes, clinicians should remain mindful of applicable laws and regulations to ensure the safe, compliant use of AI technologies in healthcare. These legal requirements operate at both federal and state levels. For instance, the Health Insurance Portability and Accountability Act (HIPAA, 1996) mandates strict compliance with standards for privacy, security, and the handling of patient data. Therefore, when assessing AI vendors, clinicians must ensure that the platform's usage complies with all applicable laws and regulations designed to safeguard patient protections (WHO, 2021). Furthermore, given that laws governing AI are continuously evolving (Secinaro et al., 2021), clinicians should actively monitor emerging legislation and regulatory updates to maintain compliance.

Other Governance Guidelines

AI governance guidelines have rapidly emerged in response to the expanding influence of AI, including in healthcare and behavioral services. Foundational frameworks include the OECD Principles on AI (OECD, 2019), the EU AI Act (European Commission, 2021), and the U.S. Blueprint for an AI Bill of Rights (White House OSTP, 2022). These frameworks emphasize the principles of transparency, accountability, fairness, and societal values, with greater emphasis being placed in high-risk domains involving vulnerable populations (e.g., health care, ABA). AI ethics and governance areas are rapidly evolving from *which* and *how* ethical principles are relevant to enforceable regulatory mechanisms, such as risk classification systems, documentation requirements, auditing, and bias mitigation (Floridi et al., 2018; Mittelstadt, 2022). AI's usage in clinical decision support tools, automated assessment systems, and administrative technologies is growing rapidly. Therefore, it's important to stay informed about developments in AI governance. Practitioners should be especially mindful of guidelines related to explainability, data privacy, and accountability—ensuring any AI-enhanced tools used in practice uphold ethical standards and the rights of clients (BACB, 2020; Cabitza et al., 2017).

THE USE OF AI IN HEALTHCARE HAS GROWN OVER TIME, in part due to the increasing availability of large data sets. Innovations in big-data analytical methods have been a direct result of this trend (Congressional Research Service [CRS], 2024).

CRS (2024) categorizes the use of AI in healthcare into the following areas: diagnosis and treatment, patient engagement and adherence with treatment plans, and administrative applications. While AI has the potential to improve care delivery, organizations and individual users must also recognize (and plan for) new challenges as well as the potential to exacerbate existing challenges. The following section suggests factors to consider when selecting an AI tool and monitoring its use.

AI System Selection

AI selection begins with evaluating your current or predicted needs. Sokolow (2024) stated that "AI is the ability of a machine or software to process and generate information like a human." To that end, you need to consider the people and processes impacted by a potential AI system before evaluating the technology. For example, many AI solutions focus on improved efficiency, which has implications for staffing levels and workflow.

As part of the decision-making process, organizations should align their priorities with the type of AI technology being considered. For instance, you might consider an AI solution to reduce barriers to onboarding patients if that's currently a pain point. Alternatively, you may choose a documentation-auditing tool if you determine that your clinical documentation fails to support medical necessity.

You should also consider the integration pattern of AI into the clinical workflow. In a typical clinical workflow, raw data are analyzed by the provider; then interpretations are provided directly to the patient. One integration pattern could be implementing AI to assist reviewing raw data. The AI's report would then be reviewed by the provider. Finally, the provider would go over the final report with the patient (Pascoe et al., 2024). Additional models exist that omit human review, with AI generated-reports delivered directly to the patient.

When leveraging AI, you should address the transparency of its development and use as well as the interpretability of its processes and outputs.

AI technology comes with its own set of features. But that doesn't necessarily mean you should adopt them all. It's important to review all features, determine whether and how you'll implement them, and whether you're in compliance with applicable laws and regulations.

Transparency

Transparency in AI is a broad topic both in scope and application. The "black box problem" refers to how transparently AI arrives at an output and how that output impacts the lives of people receiving healthcare services (Linardatos et al., 2021). To achieve transparency, AI must be *explainable* (i.e., how the output was derived is knowable and communicated) and *interpretable* (i.e., the ability to comprehend something) (Kiseleva et al., 2022). Achieving AI interpretability depends on the extent to which it is explainable. When leveraging AI, you should address the transparency of its development and use as well as the interpretability of its processes and outputs (Kiseleva et al., 2022). For the purposes of these guidelines, AI transparency will be discussed across these three areas: accountability, safety and quality, and the ability to make informed consent (Kiseleva et al., 2022).

Accountability means "having the obligation to answer questions regarding decisions and/or actions" (Brinkerhoff, 2004). Organizations should hold the developers of AI accountable for explaining how the AI system works, as well as providing the information necessary to assess and justify its performance (Kiseleva, 2020a; Rieder and Hofmann, 2020). For example, if you use AI to help identify how many treatment hours are clinically indicated for a client, you should first understand how the AI system arrives at the output and what information informs the decision-making process.

Safety and quality go hand in hand with transparency, especially as it relates to the accuracy of AI outcomes (Kiseleva et al., 2022). Outcomes derive from the quality and relevance of your inputs as well as training and data validation processes (Kiseleva et al., 2022). Careful, ongoing oversight of the AI systems enables continuous improvement (e.g., testing, auditing, and debugging) and increased safety (Carvalho et al., 2019). This also provides a pathway for tracing errors to their source (Carvalho et al., 2019).

AI transparency is also critical for a client to provide informed consent, which protects welfare and individual rights (BACB, 2020). True informed consent is only possible if your organization understands AI's working mechanisms and its use in service delivery. You must grasp AI decision-making processes, when and how AI tools are appropriate, and how to verify results (Kiseleva et al., 2022). An organization should be prepared to disclose the extent to which AI is being used, its risks and benefits, and its adverse effects (BACB, 2020; Kiseleva et al., 2022). When possible, the organization should identify and offer non-AI alternatives to clients and families as an opt-out option (Kiseleva et al., 2022).

Social Significance

Baer et al. (1968, 1987) describe socially important behavioral changes as those leading to outcomes that are beneficial to the individual receiving the service. When integrating AI into the workflow of delivering ABA, consider not only the practitioner's benefits, but also how it will affect services and treatment outcomes for the client and their family.

One AI integration is predictive analysis. Alowais et al. (2023) state that healthcare can be optimized by using AI to improve predictive models' accuracy. Before adopting predictive analysis, consider the extent to which it enhances clinical processes and affects each client's services, treatment outcomes, and reason for starting care. Social significance is multifaceted. It optimizes clinical judgement and gives the client and family more efficient treatment—potentially decreasing time in care. You should continually monitor and evaluate the impact of AI, including how it supports or impedes clinical work as well as the treatment provided to the client and their family.

AI technologies are being explored for their potential to support data collection, pattern recognition, and decision-making processes (Walz, 2024). These tools may contribute to streamlining aspects of clinical workflows and providing additional insights to guide intervention planning.

In some cases, AI applications could assist in managing administrative tasks, organizing treatment data, or standardizing certain procedures (e.g., client onboarding). This may help organizations maintain consistency across service delivery. However, continuous assessment is encouraged to determine whether AI is supporting the identified areas of need (and whether it's positively impacting the client) (Walz, 2024).

As mentioned previously, the introduction of AI into ABA services raises important considerations related to data privacy, informed consent, and implementation fidelity. The potential effects on client outcomes and professional roles are areas that require further study. Ongoing evaluation and collaboration across disciplines will likely be necessary to understand the broader implications of these technologies in practice.

AI Deployment

We recommend you carefully consider four critical areas of AI integration: communication, timing, training, and implementation. These areas don't encompass everything you should consider, but they're a strong core that can guide you toward responsible AI integration.

Communication: The goal of a well-articulated communication strategy is to promote transparency, collaboration, and clarity among your employees. Ultimately, ensuring stakeholders understand the *what*, *why*, and *how* of AI integration helps support expectation management and engagement while increasing the likelihood of successful incorporation into their workflow. There must be clarity around the vision, goals, and objectives of using AI as well as how it aligns with the organization's strategic goals for accomplishing their mission (Martins, 2023).

Prior to communicating at the organizational level, the leaders should have a clear understanding of all ethical or legal considerations and the need for continuous learning (Martins, 2023). For example, if an organization plans to use AI to develop treatment goals, they should be prepared to obtain stakeholder consent, conform with applicable laws, and establish a cadence for ongoing training.

The communication may also include information about the benefits of using AI for the organization and its employees. Doing so helps illuminate which positive attributes and improvements will be experienced. It also allows employees to anticipate acquiring additional and improved skills, less burdensome work, and/or greater job satisfaction (Martins, 2023).

Be sure your communications regarding AI are consistent. This reduces confusion while maintaining alignment with your goals and mission. There should also be frequent communication to support engagement as well as ways to keep employees updated about modifications, important milestones, and special achievements (Martins, 2023). Finally, it's important that employees have a means of raising concerns, questions, or feedback to help maintain momentum for successful integration (Martins, 2023).

Timing: Organizational readiness, resources, and capacity will determine when AI integration is appropriate. Even an excellent program can face significant barriers *if* you don't deploy it out thoughtfully and have a plan for change management. You need to evaluate current technological infrastructure, culture, and workflows to properly assess organizational readiness.

Organizations might consider a phased approach, wherein you begin with low-risk use cases and gradually move to more complex tasks. It's also important to be mindful of change fatigue and rollout pacing. Implementing too many initiatives at once or moving too quickly through integration can impact adoption and fidelity of the AI tool, program, etc. Finally, organizations should consider aligning AI integration with naturally occurring annual cycles and planning cadence to avoid disruption of important day-to-day practices.

Training and Implementation

Onboarding: Once an AI system has been selected and approved for use, establishing an effective onboarding process is important. This includes developing comprehensive, and role-specific training materials tailored to the AI system's application within the organization. Initial training sessions should be hands on, ensuring clinicians feel comfortable using the technology. End users should also learn to critically evaluate AI system outputs before integrating them into routine tasks. Additionally, understanding how to report potentially inappropriate/inaccurate (i.e., hallucinations/confabulations) recommendations once the AI system is embedded in workflows is essential to the refinement of the tool. To enhance training effectiveness, real world use case scenarios should be incorporated, demonstrating how varied AI systems can streamline daily tasks. Additionally, ongoing technical support should be readily available to address challenges and ensure seamless implementation.

Regular Training: Ongoing staff training enhances AI-related skills and fosters collaboration in overcoming challenges. Encourage teams to share success stories, such as how they've effectively utilized the tool for various tasks or in innovative ways. Create dedicated spaces for discussing barriers and developing problem-solving strategies. Training should focus on building baseline proficiency and on mastery, ensuring staff feel confident integrating the AI system into their workflows. Maintaining ethical AI use aligned with the

The true value of AI systems lies in the ability to support patient care while concurrently reducing clinician burden.

organization's mission, vision, and values should remain an ongoing, structured dialogue. While AI systems are designed to support staff, they also bring risks. Employees must stay vigilant to prevent errors that could impact individual clients and staff members as well as the organization.

Implementation: AI system interfaces should be guided by user-centric design principles and processes. Mechanisms for continuous user feedback should be integrated throughout the discovery process and usability testing. This enables you to identify key user needs and ensure they're being met as intended.

Successfully integrating AI tools into your organization's workflow requires you to understand and meet user needs. Optimizing user experience (UX) and prioritizing seamless workflow integration should be front and center during implementation. Pascoe et al. (2024, p.673) state that "clear navigation, concise feedback, and minimal cognitive load are essential for clinicians to effectively use AI-driven functionalities within their workflow." Therefore, applying user-centered design principles helps ensure AI tools are tailored to clinicians' needs, preferences, and perspectives—ultimately improving usability and user satisfaction.

The true value of AI systems lies in the ability to support patient care while concurrently reducing clinician burden by delivering actionable insights or simplifying access to patient data (Pascoe et al., 2024). Understanding clinical workflows and user preferences is key to developing interfaces that support the following:

Contextual relevance: Presenting AI outputs alongside related data (e.g., trends and/or note summaries) enables informed decision making.

Human oversight: Features like color-coded, AI-generated notes for quick review allow clinicians to monitor outputs, reduce bias, and stay actively engaged.

Additionally, seamless integration into clinical workflows is essential for effective adoption of AI tools in ABA organizations. AI solutions should be embedded into clinicians' daily routines and aligned with their existing use patterns. This requires mapping/aligning AI functionalities to specific clinical tasks and potentially ensuring interoperability with other IT systems. By integrating AI tools that naturally fit into workflows, ABA organizations can enhance efficiency, minimize cognitive load, and further enhance overall performance (Howe et al., 2018).

Finally, integrating a continuous feedback loop between clinicians and AI system developers is essential to refine tool performance, enhance usability, and ensure the technology evolves in alignment with real-world clinical needs. This iterative process enables the system to adapt based on user input—leading to more accurate, relevant, trusted outcomes over time.

Monitoring and Auditing

Monitoring

Monitoring refers to ongoing oversight of AI systems to ensure they continue to operate as intended (e.g., Cheong, 2024; Sayles, 2024). This includes continuously evaluating system performance, safety, and relevance to the tasks for which it is used or marketed. Given the dynamic nature of AI models, organizations must implement procedures to detect unintended consequences, performance drift, or misalignment of AI system outputs with ethical standards and treatment objectives (e.g., Bayram et al., 2022; Merkow et al., 2024). Routine monitoring should assess how well the AI system supports behavior analysts' clinical judgment, upholds client rights, and functions across diverse populations (Jennings & Cox, 2024). Protocols should also specify who is responsible for monitoring and communicating concerns (e.g., vendors, AI system developers, ABA organization), the frequency with which monitoring must occur, and thresholds for acceptable variance in outputs and deviations from expected outputs (e.g., Cheong, 2024).

Effective monitoring might be automated or involve humans (i.e., human-in-the-loop; Drori & Te-eni, 2024). Organizations might integrate real-time dashboards to track AI system outputs and compare them with expected patterns, while behavior analysts and supervisors regularly review AI-informed recommendations to confirm clinical appropriateness (e.g., Drori & Te-eni, 2024; Jennings & Cox, 2024). Discrepancies should trigger a review of the AI system's logic, data inputs, and configuration settings. All monitoring efforts should be documented and included in organizational quality assurance processes to support transparency, facilitate error reporting, and enable continuous improvement (e.g., Cheong, 2024).

Auditing

For the purposes of these guidelines, auditing refers to the formal, systematic examination of AI systems used in ABA (e.g., Mökander, 2023). The purpose of auditing is to assess compliance with clinical, ethical, and legal standards (Mökander & Floridi, 2024). Unlike monitoring—which occurs continuously—auditing is periodic and retrospective and often involves a more in-depth analysis. Audits may include evaluation of the system's algorithmic decisions, data-handling practices, performance across demographic groups, and model alignment with medical necessity and ethical obligations (e.g., Funda, 2025). You should establish an audit schedule for all your AI systems, define key performance and fairness metrics and benchmarks, and document findings to ensure transparency and accountability (Funda, 2025; Mökander, 2024; Mökander & Floridi, 2024).

AI system auditing should be conducted by individuals or committees with sufficient expertise in AI functionality and ABA practice standards. When possible, audits should be conducted independently from the developers or vendors of the system being evaluated, with the results posted publicly (e.g., Gerhards & Siemer, 2016; Hibbard et al., 2005; Nordstrom et al., 1991).

Findings from AI audits can inform risk mitigation strategies, deprecation decisions, staff training needs, and future procurement decisions (e.g., Mökander & Floridi, 2024). To help support a culture of responsible innovation, audit results should be shared with relevant stakeholders—including clinical leadership and compliance teams—and tied to mechanisms and policies for actionable change. Organizations may also consider external or third-party audits for high-risk or high-impact systems (Funda, 2025).

Error Reporting

Error reporting refers to the structured process whereby AI-related malfunctions, unexpected outputs, or harmful consequences are identified, documented, and communicated within an organization (e.g., Cattell et al., 2024; Cabrera et al., 2021). A clear, accessible reporting mechanism enables providers, paraprofessionals, and other stakeholders to flag concerns without fear of reprisal. Error reports may include instances of incorrect data interpretation, biased recommendations, or system failures that disrupt care delivery (e.g., Cattell et al., 2024; Cabrera et al., 2021). Effective error reporting processes should include mechanisms for triaging urgent risks and procedures to review, remediate, and provide feedback to those who reported the issue. For example, an organization might implement a Coordinated Flaw Disclosure framework that includes documentation, adjudication panels, and automated verification processes to systematically address reported AI errors (Cattell et al., 2024).

Organizations should incorporate AI error reporting into existing incident reporting or quality assurance frameworks. Additionally, vendors and AI system developers should provide timely access to documentation and support when investigating system-related concerns. Transparency in how errors are handled builds trust among users and contributes to the safe and ethical integration of AI technologies in ABA (Winecoff & Bogen, 2024). Recurring issues flagged through this process can inform updates to monitoring protocols, staff training, and system configuration. In turn, this enhances the reliability and safety of AI-assisted tools.

For example, consider a situation where a behavior analyst reports that an AI-driven recommendation suggests an intervention that does not align with GASC. The behavior analyst should be able to easily communicate their concerns to the vendor. The vendor should provide comprehensive documentation detailing the AI system's decision-making process (e.g., access to model cards, information about underlying data sources, algorithmic logic, methods used for values alignment). The vendor should also provide support through dedicated channels such as help desks, ticket submission systems, or technical customer success specialists. Such quick, efficient, proactive engagement can ensure concerns raised by clinicians are addressed efficiently while maintaining the quality and integrity of clinical services provided.

AI System Deprecation

AI system deprecation refers to the planned phaseout or discontinuation of a specific AI model, algorithm, or software version. This typically occurs due to performance limitations, outdated training data, security vulnerabilities, or the release of more advanced alternatives (Brundage et al., 2020; Lepri et al., 2018).

For ABA organizations, this process has at least two critical implications for clinical decision making and organizational operations. First, deprecated systems may no longer receive updates, security patches, or compliance certifications. This increases the risk of inaccurate outputs, ethical violations, or breaches of confidentiality (Cabitza et al., 2017). Second, if vendors automatically switch a product to a new version (because they deprecated an older version), the AI system outputs may differ from expectations developed by using the previous version (e.g., different types of goals are recommended, session note auditing catches different things). Given this potential disruption, ABA organizations and vendors should proactively communicate about when deprecation will occur and known changes to expected outputs.

ABA practitioners should regularly assess whether their AI tools remain supported and aligned with current ethical standards and regulations as well as how expected outputs change over time. Vendor contracts should include clauses mandating notification of deprecation plans. They should also include access to documentation for evaluating system reliability and appropriateness for client use when changes are made.

CONCLUSION

IN CONCLUSION, AI CAN POSITIVELY SHAPE SERVICE DELIVERY WHEN THOUGHTFULLY DEPLOYED, anchored in science, and ethically committed. From reducing administrative burdens to enhancing learning experiences, AI is quickly contacting every workflow facet. CASP remains committed to supporting its members by providing resources to assist with this rapidly evolving technology.

REFERENCES

ABA Coding Coalition. (2022, January). *Model coverage policy for adaptive behavior services* (2nd ed.). https://abacodes.org/wp-content/uploads/2022/01/Model-Coverage-Policy-for-ABA-01.25.2022.pdf

AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Saleh, K. B., Badreldin, H. A., Yami, M. S. A., Harbi, S. A., & Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1). https://doi.org/10.1186/s12909-023-04698-z

Artificial intelligence definitions. (2025, February 10). UNF Ofe. https://www.unf.edu/ofe/ai/definitions.html

Artificial Intelligence (AI) in Health Care. (2025, April 25). https://www.congress.gov/crs-product/R48319

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91–97. https://doi.org/10.1901/jaba.1968.1-91

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *20*(4), 313–327. https://doi.org/10.1901/jaba.1987.20-313

Bayram, F., Ahmed, B. S., & Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *arXiv*:2203.110700. https://doi.org/10.48550/arXiv.2203.11070

Behavior Analyst Certification Board. (2024, July). *BACB Newsletter: July 2024*. Behavior Analyst Certification Board. https://www.bacb.com/wp-content/uploads/2024/07/BACB_July2024_Newsletter-241021-a.pd

Behavior Analyst Certification Board (BACB) (2020). Ethics Code for Behavior Analysts. https://www.bacb.com/ethics-code/

Behavior Analyst Certification Board. (2021). RBT ethics code (2.0). https://www.bacb.com/wp-content/rbt-ethics-code

Berner, E.S. (2009) Clinical Decision Support Systems: State of the Art. AHRQ Publication, 4-26.

Blue Cross and Blue Shield of Oklahoma. (2023). *Applied Behavior Analysis Description (Policy No. CPCP011)*. https://www.bcbsok.com/docs/provider/ok/standards/cpcp/cpcp011-aba-06-12-23.pdf

Brinkerhoff, D. W. (2004). Accountability and health systems: Toward conceptual clarity and policy relevance. *Health Policy and Planning, 19*(6), 371–379. https://doi.org/10.1093/heapol/czh052

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., Fong, R., Maharaj, T., Koh, P. W., Hooker, S., Leung, J., Trask, A., Bluemke, E., Lebensold, J., O'Keefe, C., Koren, M., Ryffel, T., Rubinovitz, J. B., Besiroglu, T., Carugati, F., Clark, J., Eckersley, P., de Haas, S., Johnson, M., Laurie, B., Ingerman, A., Krawczuk, I., Askell, A., Cammarota, R., Lohn, A., Krueger, D., Stix, C., Henderson, P., Graham, L., Prunkl, C., Martin, B., Seger, E., Zilberman, N., Ó hÉigeartaigh, S., Kroeger, F., Sastry, G., Kagan, R., Weller, A., Tse, B., Barnes, E., Dafoe, A., Scharre, P., Herbert-Voss, A., Rasser, M., Sodhani, S., Flynn, C., Gilbert, T. K., Dyer, L., Khan, S., Bengio, Y., & Anderljung, M. (2020). *Toward trustworthy AI development: Mechanisms for supporting verifiable claims*(arXiv:2004.07213). arXiv. https://arxiv.org/abs/2004.07213

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended Consequences of Machine Learning in Medicine. *Journal of the American Medical Association, 318*(6), 517–518. https://doi.org/10.1001/jama.2017.7797

Cabrera, Á. A., Druck, A. J., Hong, J. I., & Peter, A. (2021). Discovering and validating AI errors with crowdsourced failure reports. *arXiv*:21.09.11690. https://doi.org/10.48550/arXiv.2109.11690

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), Article 832. https://doi.org/10.3390/electronics8080832

Cattell, S., Ghosh, A., & Kaffee, L. A. (2024). Coordinated flaw disclosure for AI: Beyond security vulnerabilities. *arXiv*:2402.07039. https://doi.org/10.48550/arXiv.2402.07039

Centers for Medicare & Medicaid Services (CMS). (2024). Complying with Medical Record Documentation Requirements [Fact Sheet]. https://www.cms.gov/files/mln909160-complying-with-medical-record-documentation-requirements.pdf

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*, 1421273. https://doi.org/10.3389/fhumd.2024.1421273

Chowdhary, K.R. (2020). Natural Language Processing. In: Chowdhary, K.R., Ed., Fundamentals of Artificial Intelligence, Springer, Berlin, 603-649. https://doi.org/10.1007/978-81-322-3972-7_19

Cigna. (2025). *Intensive Behavioral Interventions (Coverage Policy No. EN0499)*. https://static.cigna.com/assets/chcp/pdf/coveragePolicies/medical/en_mm_0499_coveragepositioncriteria_intensive_behavioral_interventions.pdf

Council of Autism Service Providers (CASP). (2025). Evidence About ABA Treatment for Young children with Autism: The impact of Treatment Intensity on Outcomes. Author. https://assets-002.noviams.com/novi-file-uploads/casp/pdfs-and-documents/evidenceaboutabatreatment.pdf

DAMA International (2017). DAMA-DMBOK: Data Management Book of Knowledge: 2nd Edition. Technics Publications, LLC. ISBN: 978-1-63462-234-9

Drori, I., & Te-eni, D. (2024). Human-in-the-loop AI reviewing: Feasibility, opportunity, and risks. *Journal of the Association for Information Systems*, 25(1), 98-109. https://doi.org/10.17705/1jais.00867

European Commission (2021). Proposal for a Regulation on Artificial Intelligence. https://artificialintelligenceact.eu/

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Funda, V. (2025). A systematic review of algorithm auditing processes to assess bias and risks in AI systems. *Journal of Infrastructure, Policy and Development*, 9(2), 11489. https://doi.org/10.24294/jipd11489

Gerhards, L., & Siemer, N. (2016). The impact of private and public feedback on worker performance—Evidence from the lab. *Economic Inquiry*, 54(2), 1188-1201. https://doi.org/10.1111/ecin.12310

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). Aligning AI with Shared Human Values. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 35(17), 14660-14669. https://arxiv.org/pdf/2008.02275v3

Hibbard, J. H., Stockard, J., & Tusler, M. (2005). Hospital performance reports: Impact on quality, market share, and reputation. *Health Affairs,* 24(4), 1150-1160. https://doi.org/10.1377/hlthaff.24.4.1150

Howe, J. L., Adams, K. T., Hettinger, A. Z., & Ratwani, R. M. (2018). Electronic health record usability issues and potential contribution to patient harm. *Journal of the American Medical Association*, *319*(12), 1276-1278. https://doi.org/10.1001/jama.2018.1171.

IBM. (2025, February 13). Data labeling: What is data labeling? https://www.ibm.com/think/topics/data-labeling?mhsrc=ibmsearch_a&mhq=data%20labeling

Jennings, A. M., & Cox, D. J. (2024). Starting the conversation around the ethical use of artificial intelligence in applied behavior analysis. *Behavior Analysis in Practice*, *17*, 107-122. https://doi.org/10.1007/s40617-023-00868-z

Kadzielski, M. A., Reynolds, M. B. (1993). Legal review: auto-authentication of medical records raises verification concerns. Perspectives in Health information Management, 14(1), 77-82.

Kiseleva, A. (2020a). AI as a medical device: Is it enough to ensure performance transparency and accountability? *European Pharmaceutical Law Review*, 4(1), 5–16. https://doi.org/10.21552/eplr/2020/1/4

Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations. *Frontiers in artificial intelligence*, 5, 879603. https://doi.org/10.3389/frai.2022.879603

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(18). https://doi.org/10.3390/e23010018

Martins, M. R. (2023). *Adapting change management strategies for the AI era: Lessons from large-scale IT integrations. World Journal of Advanced Research and Reviews*, *19*(3), 1604–1629. https://doi.org/10.30574/wjarr.2023.19.3.1556

Mello, M. M., & Guha, N. (2024). Understanding Liability Risk from Using Health Care Artificial Intelligence Tools. *The New England Journal of Medicine*, *390*(3), 271–278. https://doi.org/10.1056/NEJMhle2308901

REFERENCES

Merkow, J., Dorfner, F. J., Yang, X., Ersoy, A., Dasegowda, G., Kalra, M., Lungren, M. P., Bridge, C. P., & Tarapov, I. (2024). Scalable drift monitoring in medical imaging AI. *arXiv*:2410.13174. https://doi.org/10.48550/arXiv.2410.13174

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence 1*, 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mökander, J. (2023). Auditing of AI: Legal, ethical, and technical approaches. *Digital Society*, 2, 49. https://doi.org/10.1007/s44206-023-00074-y

Mökander, J., & Floridi, L. (2024). Operationalising AI governance through ethics-based auditing: An industry case study. *arXiv*:2407.06232. https://doi.org/10.48550/arXiv.2407.06232

Nordstrom, R., Lorenzi, P., & Hall, R. V. (1991). A Review of Public Posting of Performance Feedback in Work Settings. *Journal of Organizational Behavior Management*, *11*(2), 101–124. https://doi.org/10.1300/J075v11n02_07

OECD (2019). OECD Principles on Artificial Intelligence. https://oecd.ai/en/ai-principles

Pascoe, J. L., Lu, L, Moore, M. M., Blezek, D. J., Ovalle, A. E., Linderbaum, J. A., Callstrom, M. R., & Williamson, E. E. (2024). Strategic considerations for selecting artificial intelligence solutions for institutional integration: A single-center experience. *Mayo Clinic Proceedings: Digital Health*, 2(4), 665-676. https://doi.org/10.1016/j.mcpdig.2024.10.004.

Publication manual of the American Psychological Association (7th ed.). (2019). In *American Psychological Association eBooks*. https://doi.org/10.1037/0000165-000

Rieder, B., & Hofmann, J. (2020). Towards platform observability. *Internet Policy Review*, 9(4), Article 1535. https://doi.org/10.14763/2020.4.1535

Secinaro, S. Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in health care: a structured literature review. BMC Medical Informatics and Decision Making, 21, 125-148. https://doi.org/10.1186/s12911-021-01488-9.

Shen, T., Jin, R., Huang, Y., Liu, C., Dong, W., Guo, Z., Wu, X., Liu, Y., Xiong, D. (2023). Large Language Model Alignment: A survey. *arXiv*:2309.15025. https://doi.org/10.48550/arXiv.2309.15025

Smith, A. L., Greaves, F., & Panch, T. (2023). Hallucination or confabulation? Neuroanatomy as metaphor in large language models. *PLOS Digital Health*, 2(11), e0000388. https://doi.org/10.1371/journal.pdig.0000388

Sokolow, B. (2024, February 21). Back to basics: How to pick the right AI solution. *Technology Solutions That Drive Health Care*. https://healthtechmagazine.net/article/2024/02/back-basics-how-pick-right-ai-solution

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanities and Social Sciences Communications, 11(1). https://doi.org/10.1057/s41599-024-03811-x

Walz, R. (2024, January 12). *Navigating the Ethical Terrain of Generative AI in Behavior Analysis: a Three-Part series*. https://science.abainternational.org/2024/01/12/6836/

What is data modeling? (2024, October 9). Sigma Computing. https://www.sigmacomputing.com/blog/what-is-data-modeling#:~:text=Consider%20a%20social%20network%20where,on%20and%20are%20great%20for:

Why confabulation, not hallucination, defines AI errors. (n.d.). https://www.integrative-psych.org/resources/confabulation-not-hallucination-ai-errors

White House OSTP (2022). Blueprint for an AI Bill of Rights. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/

Winecoff, A. A. (2024). Improving governance outcomes through AI documentation: Bridging theory and practice. *arXiv*.2409.08960. https://doi.org/10.48550/arXiv.2409.08960

World Health Organization (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Licence: CC BY-NC-SA 3.0 IGO.