



MINNESOTA BOARD OF PSYCHOLOGY

**June 12, 2026
Board Meeting**

Order of Business

PUBLIC SESSION:

- 1. Call to Order**
- 2. Adoption of Tentative Agenda**
- 3. Announcements**
- 4. Approval of the Board Minutes**
 - A. Board meeting minutes**
- 5. Consent Agenda**
 - A. Consent Agenda**
- 6. New Business**
 - A. AI in Psychology Practice**
 - B. ED Report**
 - C. Variance Request**
 - D. FARB Summit on Regulatory Excellence**
 - E. Board Administrative Terminations**
- 7. Committee Reports**
- 8. Adjournment**



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY:

TITLE: Board meeting minutes

INTRODUCTION TO THE TOPIC:

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description

Board meeting May Minutes

Upload Date Type

6/12/2026 Cover Memo

MINNESOTA BOARD OF PSYCHOLOGY
Board Meeting Minutes of May 15, 2026

Board Members and Staff in Attendance: Sonal Markanda, Sebastian Rilen, Michael Thompson, Daniel Hurley, Pamela Freske, Cesar Gonzalez, Lindsey Hickey, Michelle Zhao, Nancy Cameron, Jackie Cardinal, Sam Sands, and Trisha Hoffman.

Guests: Rebecca Huting

PUBLIC SESSION

1. Call to Order

Sonal Markanda called the meeting to order at 9:35AM. The meeting was held in a hybrid format with some individuals in attendance in person and others online. Voting was held by roll call.

A. Webex MeetingLink

2. Adoption of Tentative Agenda

Daniel Hurley moved, seconded by Michael Thompson Motion: to adopt the tentative agenda, provided that the variance request is considered before any other agenda items. There being 9 "ayes" and 0 "nays" the motion Passed.

3. Announcements

4. Approval of the Board Minutes

Seb Rilen moved, seconded by Nancy Camerson Motion: to adopt March 20, 2026, Board Meeting Minutes. There being 7 "ayes" and 0 "nays" the motion Passed

5. Consent Agenda

A. Staff Delegated Authority Report

B. HPSP Reports

6. New Business

A. AI in Psychology Practice

The Board discussed several articles relating to developments in Artificial Intelligence and the Practice of Psychology

B. Master's Level Licensure

Seb Rilen reported that the Legislative Committee has requested that Board staff gather information related to Master's level licensure, and after today's Board meeting, the committee will discuss the issues in light of the information gathered.

C. Executive Director's Report

Trisha Hoffman reported that the Licensure team has issued 1,032 Behavior Analyst licenses and continues to receive several applications per week. The team has also continued to follow up with applicants who haven't had recent movement in their applications, and prepared to implement the renewal process for BA licensees.

Sam Sands reminded the Board of the upcoming transition from NOVUS to PEAK Agenda. He also described presentations given at the ASPPB Mid-Year Meeting and obtained input from the Board on a potential training day featuring some of the presenters. He updated the Board on the state of revenue and expenses relative to the budget, and noted that the legislature will finish its session shortly and there is uncertainty about whether it may take further action on some bills related to the practice of Psychology.

D. CLEAR Training

The Board reviewed and discussed training made available by the Council on Licensure, Enforcement and Regulation (CLEAR). Michelle Zhao moved, seconded by Nancy Cameron Motion: to approve purchase of CLEAR training for any and all Board members who opt in to receive the training There being 9 "ayes" and 0 "nays" the motion Passed

E. Variance Request Application 22-0263

Motion by the Application Review Committee: to approve the CE Variance Request. There being 9 "ayes" and 0 "nays" the motion Passed.

F. Board Administrative Terminations

Daniel Hurley moved, seconded by Seb Rilen Motion: to approve the Board Administrative Terminations. There being 9 "ayes" and 0 "nays" the motion Passed.

7. Committee Reports

Sam Sands noted that Agreements for Corrective Action that were approved by the Complaint Review Committees since the last Board meeting have been provided to Board members for their information.

A. Agreements for Corrective Action Signed Since Last Board Meeting

8. Adjournment

Adjourned at 10:35 AM.



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY:

TITLE: Consent Agenda

INTRODUCTION TO THE TOPIC:

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description

Psychologist Licensure Consent Agenda
Behavior Analyst Licensure Consent Agenda
Compliance Report Consent Agenda

Upload Date Type

6/12/2026 Cover Memo
6/12/2026 Cover Memo
6/12/2026 Cover Memo

CONSENT AGENDA ITEMS: Staff Delegated Authority Report

Admission to Examination for Professional Practice in Psychology (EPPP)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Admission to the Examination for Professional Practice in Psychology (EPPP) pursuant to [Minnesota Rules 7200.0550](#).

Applicant(s) Granted Admission to the (EPPP) Exam
Leah Danson, Psy.D.
Tanner McCarthy, Psy.D.
Breanne Sande Russell, Ph.D.
Amanda Smith, Psy.D.
Haley Brush, Psy.D.
Amy Maslowski, Ph.D.
Douglas Sylvester, Psy.D.
Mych-Jarrodd Dallozzi, Psy.D.
Ashley Friesen-Janochoski, Ph.D.
Tanya Johnston, Ph.D.

Admission to Professional Responsibility Examination (PRE)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Admission to the Professional Responsibility Examination (PRE) pursuant to [Minnesota Rules 7200.0550](#).

Applicant(s) Granted Admission to the (PRE)
Adrienne Manbeck, Ph.D.
Ciara Baumann, Ph.D.
Amy Maslowski, Ph.D.
Cole Toovey, Ph.D.
Jessica Davis, Psy.D.
Julia Jordan, Psy.D.
Benjamin Bassier, Psy.D.
Ulysses Jimenez, Psy.D.
Anna Achter, Ph.D.
Natalia Cristina Montero Vazquez, Psy.D.

Licensed Psychologist (LP)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Licensed Psychologist (LP) licensure pursuant to [Minnesota Statutes, section 148.907](#) and the administrative rules of the [Psychology Practice Act](#).

License Number	Licensee
LP5624	Frederick Robinson, Psy.D.
LP7349	Raghad Hassabelnaby Psy.D.
LP7350	Melissa Schill, Ph.D.

LP7351	Melody Hannani, Psy.D.
LP7352	Dawn Clark-Plowman, Psy.D.
LP7353	Emily Wilson, Ph.D.
LP7354	Ashlan Mahan, Psy.D.
LP7355	Rita Taylor, Ph.D.
LP7356	Quincy Guinadi, Psy.D.
LP7357	Laura Cunningham, Psy.D.
LP7358	Brenda Sharp, Ph.D.
RL00101	Frederick Robinson

Guest Licensure (GL)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Guest Licensure (GL) pursuant to [Minnesota Statutes, section 148.916](#) and the administrative rules of the [Psychology Practice Act](#).

License Number	Licensee

Licensure for Voluntary Practice (L-VP)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Licensure for Volunteer Practice (LPV) pursuant to [Minnesota Statutes 148.909](#) and the administrative rules of the [Psychology Practice Act](#).

License Number	Licensee

Emeritus Registration (Em.)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Emeritus Registration pursuant to [Minnesota Statutes, section 148.9105](#).

License Number	Licensee
ER00218	Glenn Werner

Voluntary Terminations (VT)

Under delegated authority from the Board, Board staff terminated the following License's pursuant to [Minnesota Rules 7200.3700](#).

License Number	Licensee
LP2383	Barbara Martin
LP7057	Elise Gibbs
LP3083	Steven Carter
LP1890	Raymond Knutson

LP3475	David Olson
LP1090	Susan Lund

Continuing Education Variance Requests

Under delegated authority from the Board, Board staff approved the following licensee(s)' requests for a six (6) month continuing education variance pursuant to [Minnesota Rules 7200.3860, D.](#)

License Number	Licensee
LP6239	Casey Lankow
LP1297	Jane Mcnaught
LP1094	Monica Mandell
LP2358	Grace Hong

Licensure Progression Statistics

The following data is a summary of the length of time it takes for an applicant to obtain licensure with the Minnesota Board of Psychology. The starting point is staff review; when the applicant has submitted all required documents for the specific type of license application.

Number of Initial, Reciprocity and Mobility LP applications filed since last Board meeting:	12
--	-----------

Of applications filed, number of LP applications still in review:	1
--	----------

Reasons for continued review: Additional information needed.

Initial, Reciprocity, and Mobility applications days to license:	12
---	-----------

Number of Guest License applications filed since last Board meeting:	0
---	----------

Of applications filed, number of Guest License applications still in review:	0
---	----------

Reasons for continued review: N/A
--

Guest License applications days to license:
--

CONSENT AGENDA ITEMS: Staff Delegated Authority Report

Licensed Behavior Analyst (LBA)

Under delegated authority from the Board, Board staff approved the following applicant(s) for Behavior Analyst (LBA) licensure pursuant to MN Statute 148.9983.

License Number	Licensee
LBA1032	Marsha Thompson
LBA1033	Jannah Fields
LBA1034	Catherine Hurd
LBA1035	Emma Rosenthal
LBA1036	Kirstie Ruhland
LBA1037	Savannah Washburn
LBA1038	Garrick Leith
LBA1039	Alissa Conway
LBA1040	Erin Faulkner
LBA1041	Maetha Werner
LBA1042	Paige Ramsey
LBA1043	Amanda Langone
LBA1044	McKenzie Johnson
LBA1045	Jeffrey Smith
LBA1046	Brooke Van Horn
LBA1047	Dominique Goode
LBA1048	Matthew Russell
LBA1049	Christelle Byrd
LBA1050	Hilary Comstock
LBA1051	Amie Bomber

Licensure Progression Statistics

The following data is a summary of the length of time it takes for an applicant to obtain licensure as a Behavior Analyst with the Minnesota Board of Psychology.

Total Number of LBA Applications Filed Since Last Council Meeting: 13 (Plus 5 Pending)

Of applications filed, number of LBA applications that have satisfied all license fees: 13

Of these applications, number submitted to CBC program (anticipated timeline to process CBC is 30 days): 13

Of all applications filed (and paid fees), number in compliance review: 2

Average days for license to be granted (time counted from staff review to license application approved): 1 - 2

Of applications filed, number of Behavior Analyst License applications still in review: 7

Reasons for continued review: Applications are In Progress.

Minnesota Board of Psychology
Consent Agenda Items:
Staff Delegated Compliance Report
Data from 5/15/26 – 6/12/26

Complaint Data – Open/NJ/FWD
Data on complaints that have been received since the May Board meeting.

Total complaints received, and those that either were opened for investigative review or were non-jurisdictional or forwarded to another Health-Related Licensing Board (HLB). Forwarded and non-jurisdictional may overlap as a complaint may be non-jurisdictional because the respondent is licensed with another HLB.

Licensed Psychologists:

Total Complaints	Opened	Non-Jurisdictional	Forwarded
6	5	1	1

Behavior Analysts:

Total Complaints	Opened	Non-Jurisdictional	Forwarded
0	0	0	0

Complaint Data – Reviews
Compliance data that has been reviewed since the May Board meeting.

Licensed Psychologists:

Triage Cases	Dismissed Cases	Disciplinary Conferences	SCOs Offered	NOC Assigned	Disciplinary/ Corrective Action Assigned
7	4	2	1	2	0

Behavior Analysts:

Triage Cases	Dismissed Cases	Disciplinary Conferences	SCOs Offered	NOC Assigned	Disciplinary/ Corrective Action Assigned
0	0	0	0	0	0

Complaint Data

Amount of complaints currently under investigative review and the average (median) amount of days taken to open a complaint, meaning send to investigative review, and all complaints that were closed from 5/15/26 – 6/12/26.

Licensed Psychologists:

Currently Open Complaints (under review)	Days Taken (average amt of days since 5/15/26)	Closed Complaints
Investigative review: 48	<1	6
Committee review: 18		

Behavior Analysts:

Currently Open Complaints (under review)	Days Taken (average amt of days since 5/15/26)	Closed Complaints
Investigative review:7	<1	0
Committee review:2		



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY:

TITLE: AI in Psychology Practice

INTRODUCTION TO THE TOPIC:

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description	Upload Date	Type
Article 1	6/12/2026	Cover Memo
Article 2	6/12/2026	Cover Memo
Article 3	6/12/2026	Cover Memo
Article 4	6/12/2026	Cover Memo
Article 5	6/12/2026	Cover Memo



TOPICS

PUBLICATIONS & DATABASES

RESEARCH & PRACTICE

EDUCATION & CAREER

NEWS & ADVOCACY

UPDATE FROM THE CEO

Rethinking the model of psychology training

As AI reshapes psychological practice, we face an urgent question: Are we training psychologists for the work they will do, or the work we have always done?



By [Arthur C. Evans, Jr., PhD](#)

Date created: June 1, 2026 2 min read

Vol. 57, No. 4

[Print version: page 10](#)

Becoming a Psychologist

Artificial Intelligence

“(javascript:toggleCitation());

(javascript:toggleFeedback());

f (#)

X (javascript: openSocialShare('https://twitter.com/share?url=https%3a%2f%2fwww.apa.org%2fmonitor%2f2026%2f06%2fpopulation-health-ai-psychology-training-scientist-practitioner&via=APA&I

in (javascript: openSocialShare('https://www.linkedin.com/shareArticle?mini=true&url=https%3a%2f%2fwww.apa.org%2fmonitor%2f2026%2f06%2fpopulation-h
practitioner&title=From+the+CEO%3a+Rethinking+the+model+of+psychology+training&summary=As+AI+reshapes+psychological+practice%2c+we+face+an+urgent+question%3a+Are+we+training+psycholo

✉ (javascript:openEmail('English'));

🖨 (javascript:printThis());



The 1949 Boulder Conference established the scientist-practitioner model that has long guided the training of health services psychologists. Its core premise, that practice should be grounded in science, remains a defining strength of our field. But the world psychologists are preparing to enter has

changed in fundamental ways, raising an important question: has our training model kept pace?

The pressures are not new, but they have accelerated and converged. Health care financing is shifting from fee-for-service to value-based and integrated care models, requiring a deeper understanding of financial risk and clinical outcomes. Population health has reframed the goal of mental health care, moving beyond only treating individual illness toward proactive system-level interventions that improve mental health across communities. At the same time, demographic changes are reshaping both the populations we serve and the students entering the profession.

Artificial intelligence is amplifying all these trends. AI is increasingly capable of performing structured tasks central to psychological practice, including aspects of assessment, treatment planning, and psychoeducation. This does not diminish the importance of psychologists but does sharpen a central question: What is the distinctive contribution of a health services psychologist in an AI-augmented system of care?

Part of the answer lies in what has always made doctoral-level training in psychology so valuable. In my experience leading large public behavioral health systems, psychologists played critical roles navigating complexity, applying science in real time, and designing interventions at scale. These contributions were high value and often unique. Yet these capabilities were often developed on the job rather than intentionally cultivated through training grounded in science.

This raises a broader issue that predates the evolving health care financing landscape, demographic changes, and AI, but is now more urgent. If the world psychologists work in is changing so rapidly, and our traditional roles are evolving with it, should we reconsider how we train the next generation?

The Boulder model has served us well. But honoring its legacy may require us to do what its architects did: Step back, examine the realities of our time, and design training for the world psychologists will inhabit. What does it mean to be a health services psychologist in this evolving landscape? Are we preparing trainees for the work they will do, or the work we have traditionally done? Does our current model equip psychologists to lead in a population health framework and an AI-enabled system of care? And if not, is it time to begin envisioning a Boulder 2.0?

VIEW MORE "FROM THE CEO" ARTICLES >

([https://www.apa.org/monitor/search?](https://www.apa.org/monitor/search?osquery=%22from%20the%20ceo%22or%22update%20from%20the%20ceo%22)

[OSQUERY=%22FROM%20THE%20CEO%22OR%22UPDATE%20FROM%20THE%20CEO%22](https://www.apa.org/monitor/search?osquery=%22from%20the%20ceo%22or%22update%20from%20the%20ceo%22))

Arthur C. Evans Jr., PhD, is the chief executive officer of APA. You can follow him on [LinkedIn](https://www.linkedin.com/in/arthur-c-evans-phd) [↗](https://www.linkedin.com/in/arthur-c-evans-phd) (<https://www.linkedin.com/in/arthur-c-evans-phd>).

ADVERTISEMENT

ADVERTISEMENT

Recommended Reading



**Training and
Educating
Antiracist
Psychologists**

\$24.95



**Ethical and Legal
Issues in AI for
Organizational
Consultants and
Managers**

\$24.95



**The Psychology of
Technology**

\$87.99



**Deliberate
Practice in
Rational Emotive
Behavior Therapy**

\$43.99

Members may qualify for lower pricing

Six Things Psychologists
are Talking About

The APA Monitor on Psychology® sister e-newsletter offers fresh articles on psychology trends, new research, and more.

First Name *

Last Name *

Email Address *

SUBSCRIBE**Speaking of Psychology**

Subscribe to APA's audio podcast series highlighting some of the most important and relevant psychological research being conducted today.

Subscribe to *Speaking of Psychology* and download via:

Apple [↗](#)Spotify [↗](#)

CONTACT APA

Related and recent



Bringing psychological expertise into AI development



Americans are more concerned than excited about AI



A litmus test for AI in education



Top 10 principles for understanding artificial intelligence in psychology



What psychologists need to know about new federal student loan rules



Teaching academic writing in the age of AI

Advancing psychology to benefit society and improve lives

ABOUT PSYCHOLOGY

- Science of Psychology
- Psychology Topics

STUDENTS

- Accredited Psychology Programs
- Careers in Psychology
- More for Students

PUBLICATIONS & DATABASES

- APA Style
- Books
 - Children's Books
- Databases
- DVD/Streaming Video
- Journal Subscriptions
- APA PsycNet® Journal Articles
- More Publications & Databases

NEWS & ADVOCACY

- Monitor on Psychology Magazine
- Newsletters
- Press Room
- Advocacy from APA Services, Inc.

STANDARDS & GUIDELINES

- Standards and Guidelines
- Ethics

CAREERS

- Find a Job with PsycCareers
- Early Career Psychologists

EVENTS & TRAINING

- APA Annual Convention
- Continuing Education
- Events Calendar
- Training

ABOUT APA

- | | | |
|-------------------|----------------------|-----------------|
| Governance | Corporate Supporters | Donate |
| APA Divisions | Become a Partner | APA Merch Store |
| Teams and Offices | Advertise with Us | Refund Policy |
| Jobs at APA | Help | Contact Us |

MEMBERS

- APA Community
- Get Involved
- Membership Benefits
- More for Members

RENEW MEMBERSHIP

JOIN APA

Turn insight into action at APA 2026, August 6–8, in Washington, DC, and virtual. [Register now.](#)



CAREER

Bringing psychological expertise into AI development

Psychologists are uniquely positioned to shape AI. Here’s how to move from advisory roles into leadership positions

By [Chris Palmer, PhD](#) Date created: June 1, 2026 11 min read
Vol. 57, No. 4
[Print version: page 65](#)

Artificial Intelligence

“ (javascript:toggleCitation());

(javascript:toggleFeedback());

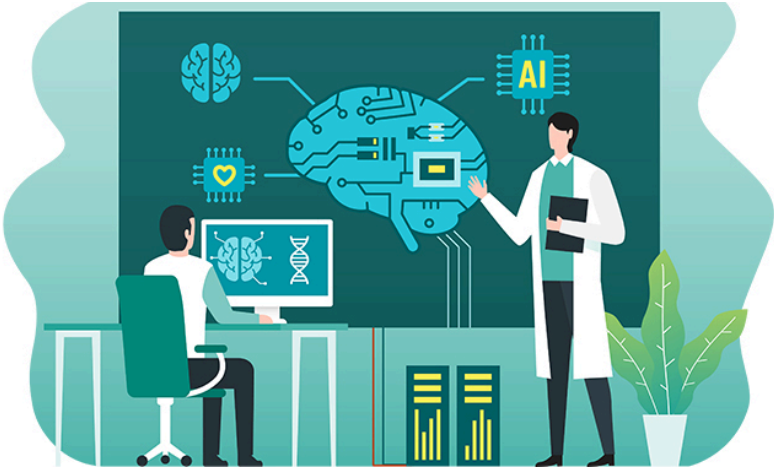
f (#)

✕ (javascript: openSocialShare('https://twitter.com/share?url=https%3a%2f%2fwww.apa.org%2fmonitor%2f2026%2f06%2fpsychological-expertise-ai-development&via=APA&text=Bringing+psychological+expertise+into+AI+development'))

in (javascript: openSocialShare('https://www.linkedin.com/shareArticle?mini=true&url=https%3a%2f%2fwww.apa.org%2fmonitor%2f2026%2f06%2fpsychological-expertise-ai-development&title=Bringing+psychological+expertise+into+AI+development&summary=Psychologists+are+uniquely+positioned+to+shape+AI.+Here%e2%80%99s+how+to+move+from+advisory+roles+into+'))

✉ (javascript:openEmail('English'));

🖨 (javascript:printThis());



When a well-known tech company attempted to build an AI-powered résumé-screening system several years ago, its developers discovered what a century of industrial and organizational (I/O) psychology research could have predicted from the start: Stripping obvious demographic markers like names from résumés didn't eliminate discriminatory bias. Information such as college affiliations, word choices, and even formatting preferences was enough to perpetuate discriminatory patterns.

"They spent a huge amount of money and time trying to sort out these models, and it never worked," said Richard Landers, an I/O psychologist at the University of Minnesota who has consulted with numerous tech companies. "[Psychological scientists] have spent decades upon decades trying to refine exactly what we're asking people for and why, specifically to avoid exactly that problem in hiring."

This scenario illustrates why psychological expertise needs to be brought into the AI development process from conception, before fundamental design decisions have been made. To do otherwise risks expensive failures that waste resources while potentially causing harm.

The AI revolution isn't just transforming work and communication—it's reshaping who gets to build the future. For psychological scientists, particularly those trained in understanding human behavior, cognition, and well-being, this moment presents what may be an unprecedented opportunity to play a more central role in technological innovation.

But seizing that opportunity requires understanding both the unique value psychologists bring to AI development and how to navigate the cultural chasm between psychological science and Silicon Valley.

What psychological scientists bring to AI

At its core, AI development is about understanding and predicting human behavior, making psychological science essential rather than optional. Psychological scientists' knowledge spans multiple critical domains that extend well beyond mental health applications and health services. Researchers and practitioners can contribute to better conception, design, implementation, and deployment of AI.

Understanding human complexity. Engineers are trained to identify specific problems, define key metrics, and iterate toward optimization. "The assumption from an engineering standpoint is: If we can't identify a metric for it, it doesn't really matter," Landers said. "Whereas a psychologist is immediately going to say, 'Well, there's this web of complexity that could erupt from what you're doing, and we need to understand that better before we charge forward.'"

This perspective has proved crucial in Landers's work developing an AI coaching chatbot for the U.S. Naval Academy. For example, when a computer scientist on his team suggested simply replicating coaching conversations to

train the model, Landers pushed back, as those exchanges are rarely recorded because of their private and sensitive nature. And even if the team did get some transcripts, coaching conversations have so much variability that thousands of examples would be required to create a data set that is broadly representative of “real” coaching.

Risk identification and mitigation. Psychological scientists’ training in ethics and their “do no harm” orientation make them attuned to potential dangers that technologists might undervalue or even dismiss altogether. When developing the Naval Academy coaching chatbot, Landers’s team proactively built a separate AI model to detect moderate risk scenarios—including indications of potential self-harm—so that trained professionals could intervene when necessary.

“From a tech perspective, they say, ‘Oh, that sounds expensive. Let’s just not do that until it’s a problem,’” Landers said. “And that’s how you end up with harm happening and people having to react to it rather than proactively trying to prevent it from the start.”

Notably, Landers’s technical partners took a back seat during this part of the project while the psychological scientists designed it themselves using an AI coding tool. “AI has become so accessible that we’ve actually been able to prototype on our own in a way that’s way faster than if we had to go through our tech team,” Landers said.

Alison Cerezo, chief science officer at mpathic, a conversation analytics company focused on AI safety, emphasizes that psychological scientists should apply the same scrutiny to AI systems as they would to human colleagues. “If I were going to add a new clinician to my clinical team, I would want to know about their background and training experiences,” they said. “We should be using that same framework with AI systems to ask, what is your training data? How are you being benchmarked? How are you being fine-tuned over time?”

Meaningful assessment. Margaret Morris, a clinical psychologist and affiliate professor in the Information School at the University of Washington who has worked with Intel, Amazon, and other companies on health technology, points to expertise in assessment as being particularly valuable. “Psychologists can find ways to embed assessments that used to be done infrequently and in the clinic into everyday interactions with technology.”

For example, social psychologists in particular can be experts on how humans tend to make biased decisions and how bounded human rationality is. Without their insights, AI models will more likely replicate such biases, leading to poor and discriminatory decision-making. This is applicable to AI decisions about things like job candidates, medical treatments, or even product choices.

The way AI responds to assessment results also benefits from psychology experts’ input. For example, AI chatbots may be particularly attuned to detecting negative emotions and encourage people to ignore or suppress

them. Cognitive psychologists understand this bias, as it replicates a human tendency to focus on negative information, where clinical psychologists understand that negative emotions are a part of daily life and that, in certain contexts, people can benefit from sitting with and even exploring such emotions.

Psychology across the AI life cycle

From framing the right research questions to evaluating real-world impact, psychological scientists play distinct roles across the AI development life cycle.

Initial design. At this stage, team members decide on a project's goals. Xuhai "Orson" Xu, a computer scientist at Columbia University's Department of Biomedical Informatics who regularly collaborates with psychological scientists, emphasizes involving experts from the start. For a project involving mental health prediction from wearable data, Xu's team consulted psychologists during study design to ensure they measured the right constructs with the right instruments. "What is the appropriate thing to measure? What are the right ecological momentary assessment short questionnaires? How can we make sure people can answer but are not also overburdened? There are many details we need to check with experts from the beginning."

Data processing. Once data collection begins, psychological expertise guides the data processing stage—the selection of information to extract. For the wearable devices used in his work, for example, Xu has to consider which variables to track, such as sleep and activity level. Rather than a purely data-driven approach, psychological scientists can draw on domain knowledge grounded in established theory.

Model interpretation. During this stage, team members evaluate their model's performance. When AI systems generate predictions, psychologists can help determine whether results align with scientific understanding. Experts can make comments such as "Okay, this is aligned with the existing literature or domain consensus," or "This is something that is misaligned but may be interesting to explore further," Xu said.

Implementation. The implementation stage is when research findings get translated to practical applications. Cerezo's work at mpathic illustrates how psychological principles can inform real-world deployment. Their team has developed more than 200 detection capabilities for monitoring AI conversations, from identifying clinical risk to assessing therapeutic rapport. The company works with about 100 psychologists with different areas of expertise to create training data, safety evaluation rubrics, and other critical tools to ensure AI safety systems are designed to handle the full range of human experience.

Ongoing monitoring. Morris emphasizes that psychological scientists' roles don't end at product launch but may involve monitoring use over time. She

argues that psychologists should check in on interactions between users and AI tools to prioritize long-term well-being over short-term engagement, to prevent users from developing dependency on the tools, and to ensure that the tools don't develop biases or violate user privacy.

Auditing, benchmarking, and red teaming

AI-related roles that psychologists can perform at technology companies include the following:

Psychological auditing involves evaluating AI systems against established psychological and ethical standards.

Human benchmarking compares AI outputs to expert human judgment rather than other AI systems. Landers's team on the Naval Academy project is bringing in clinical psychologists and professional coaches to rate the transcripts of its AI coaching chatbot, providing both quantitative and qualitative feedback that simple technical metrics cannot capture.

Red teaming involves deliberately simulating problematic or dangerous conversations to make AI systems fail or behave inappropriately. Originally a military term referring to the practice of a group taking on the role of the enemy to simulate attacks, red teaming requires a team member with the expertise and authority to know when to pause development and pressure test the system to say, "This is a conversation where human intervention is needed," Landers said. In the same training project, he used red teaming to identify danger signs in conversations with the chatbot. This led to the creation of a second AI model to monitor conversations between the chatbot and trainees to detect even moderate risk scenarios that should be handed over to human coaches.

Bridging the gap

Though still a nascent industry, a growing number of training programs are emerging to help both psychologists in training and working professionals gain the skills needed to contribute meaningfully to AI development. These range from certifications and microcredentials to master's degrees, postdoctoral programs, and continuing-education courses. Examples include initiatives at [Northwestern University](https://www.preventivemedicine.northwestern.edu/education/postdoctoral/digital-mental-health/index.html)

(<https://www.preventivemedicine.northwestern.edu/education/postdoctoral/digital-mental-health/index.html>), [Tilburg University](https://www.tilburguniversity.edu/education/masters-programs/artificial-intelligence-psychological-research/)

(<https://www.tilburguniversity.edu/education/masters-programs/artificial-intelligence-psychological-research/>), [Oregon State University](https://ecampus.oregonstate.edu/online-degrees/microcredentials/ai-fundamentals-psychology/)

(<https://ecampus.oregonstate.edu/online-degrees/microcredentials/ai-fundamentals-psychology/>), the [University of Pennsylvania](https://www.lps.upenn.edu/courses/ai-and-human-behavior/2026a/)

(<https://www.lps.upenn.edu/courses/ai-and-human-behavior/2026a/>), [Johns Hopkins University](https://ep.jhu.edu/courses/705640-cognitive-and-behavioral-foundations-for-artificial-intelligence/)

(<https://ep.jhu.edu/courses/705640-cognitive-and-behavioral-foundations-for-artificial-intelligence/>), [Brown University](https://carney.brown.edu/education-training/postdocs/training-program-computational-psychiatry-tpcp/)

(<https://carney.brown.edu/education-training/postdocs/training-program-computational-psychiatry-tpcp/>), [Harvard University](https://www.harvard.edu/psychology/department-of-psychology/)

(<https://execonline.hms.harvard.edu/artificial-intelligence-in-health-care-from-strategies-to-implementation>), and the [University of Zurich](https://www.translationalneuromodeling.org/cpcourse/) (<https://www.translationalneuromodeling.org/cpcourse/>). (These programs are provided for readers' awareness; their inclusion does not constitute endorsement.)

In addition, [APA's Ethical Emotionally Intelligent AI Chatbot Design \(/events/ethical-emotionally-intelligent-ai-training\)](#) program seeks to bridge the gap between psychological science and technology by giving chatbot designers psychological frameworks while equipping psychologists to take on stronger leadership roles within AI development. "With the right frameworks and shared language, psychologists don't have to remain in advisory roles," said Bernadette Han, PhD, an organizational psychologist at APA who develops enterprise-level learning initiatives for public-, organizational-, and technology-facing audiences. "They can serve as full partners—or even lead—by guiding teams through the psychological foundations of how AI systems are designed."

Practical advice for psychological scientists

For psychological scientists ready to engage with AI developers and succeed once they've begun a collaboration, here are some key recommendations:

Learn the basics. You don't need to become a programmer, but understanding fundamental concepts helps communication. Courses on different topics aimed at varying levels of depth abound on LinkedIn, YouTube, Coursera, and other platforms. In addition, most AI chatbots offer free versions with intuitive user interfaces.

Be flexible and curious. Morris advocates for "flexibility and curiosity" as essential qualities for psychologists to successfully collaborate with technologists, who tend to work in a nimbler iterative style that relies less on formal theory and more on tinkering. Likewise, ask questions when something is unclear or you hear an unfamiliar acronym, for example.

Calibrate expectations. Understand AI's current capabilities and limitations alike.

Stay current. AI technology evolves at a pace that can be surprisingly quick.

Network strategically. Xu's approach includes asking colleagues for connections, searching departmental websites for relevant specialists, and talking with people at conferences.

Present your expertise. Share relevant research at conferences and with computer science departments. Propose collaborations with university students doing capstone projects.

Find the applications that match your knowledge base. Psychological expertise can be valuable to everything from retail chatbots and educational tools that scaffold support and minimize extraneous mental load to digital assistants and talent acquisition platforms. Psychological scientists can contribute to several steps in the conception of these AI products, from their development and deployment to their continued evaluation.

Consider various roles. Opportunities range from academic collaborations to consulting gigs to full-time positions at tech companies.

Navigating projects. Insist on human-centeredness. Morris emphasizes that there should be a relentless focus on what's going to be meaningful to the end user.

Advocate for proper testing. Don't let pressure to ship a product quickly override safety concerns.

Balance rigor with pragmatism. Respect the complexity of the human mind and experiences—but not to the point of considering every possible contingency.

Document limitations. Be clear about what your psychological knowledge can and cannot determine.

The path forward

The integration of psychological science into AI development isn't just about making better products—it's about ensuring AI systems embody human values, respond to human complexity, and serve genuine human needs.

"I imagine what would've happened if in the early days of social media, psychologists had been on the original development teams asking, how will people benefit from interacting in this environment? What risks are there?" Landers said. "The road we're on would be quite different."

The stakes are similarly high for AI development, with AI systems increasingly mediating human interactions and even providing mental health support.

"We need more bold people doing that kind of work," Landers said. "There's never been a better time for psychologists to identify key drivers of human flourishing and develop platforms to improve it."

Cerezo sees immense potential. "I'm really excited about the future," they said. "If we lead with ethics and do our homework to understand how these systems are developed, there's no limit to what we can do."

ADVERTISEMENT

ADVERTISEMENT

Recommended Reading



**The Psychology of
Technology**
\$87.99



**Learning
Interventions for
Consultants**
\$49.99



**Ethical and Legal
Issues in AI for
Organizational
Consultants and
Managers**
\$24.95



**A Practitioner's
Guide to
Advancing
Behavioral Health
Care With
Artificial
Intelligence**
\$49.99

Members may qualify for lower pricing

**Six Things Psychologists
are Talking About**

The *APA Monitor on Psychology*® sister e-newsletter offers fresh articles on psychology trends, new research, and more.

First Name *

Last Name *

Email Address *

SUBSCRIBE

Speaking of Psychology

Subscribe to APA's audio podcast series highlighting some of the most important and relevant psychological research being conducted today.

Subscribe to *Speaking of Psychology* and download via:

Apple [↗](#)



Spotify [↗](#)



[CONTACT APA](#)

Related and recent



Rethinking the model of psychology training



Americans are more concerned than excited about AI



A litmus test for AI in education



Top 10 principles for understanding artificial intelligence in psychology



Teaching academic writing in the age of AI



As AI use surges among psychologists, so do concerns about risks

Advancing psychology to benefit society and improve lives

ABOUT PSYCHOLOGY Science of Psychology Psychology Topics STUDENTS Accredited Psychology Programs Careers in Psychology More for Students	PUBLICATIONS & DATABASES APA Style Books Children's Books Databases DVD/Streaming Video Journal Subscriptions APA PsycNet® Journal Articles More Publications & Databases	NEWS & ADVOCACY Monitor on Psychology Magazine Newsletters Press Room Advocacy from APA Services, Inc. STANDARDS & GUIDELINES Standards and Guidelines Ethics	CAREERS Find a Job with PsycCareers Early Career Psychologists EVENTS & TRAINING APA Annual Convention Continuing Education Events Calendar Training
ABOUT APA Governance APA Divisions Teams and Offices Jobs at APA	Corporate Supporters Become a Partner Advertise with Us Help	Donate APA Merch Store Refund Policy Contact Us	MEMBERS APA Community Get Involved Membership Benefits More for Members
			RENEW MEMBERSHIP JOIN APA



CASP The Council of Autism
Service Providers

PRACTICE PARAMETERS FOR ARTIFICIAL INTELLIGENCE USE IN APPLIED BEHAVIOR ANALYSIS

TABLE OF CONTENTS

Acknowledgments	3
Prefatory Notes	4
Preamble	5
• Purpose	5
• Scope	5
• Legal Disclaimer	5
Annotated Glossary of Terms	6
AI and ABA	10
• Federal and State Laws	10
• Obligations to Funding Sources	10
• Ethics	11
• Other Governance Guidelines	12
Organizational Oversight	13
• AI System Selection	13
• Transparency	14
• Social Significance	15
• AI Deployment	15
Communication	15
Timing	16
Training and Implementation	16
• Monitoring and Auditing	18
• Error Reporting	19
• AI System Deprecation	20
Conclusion	21
References	22

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*

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.

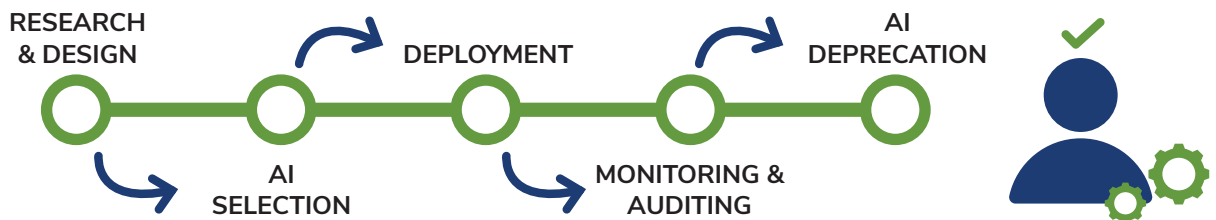
Behavior analysts, paraprofessionals, and RBTs may be collectively referred to as “provider(s)” when applicable.

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.



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.

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.

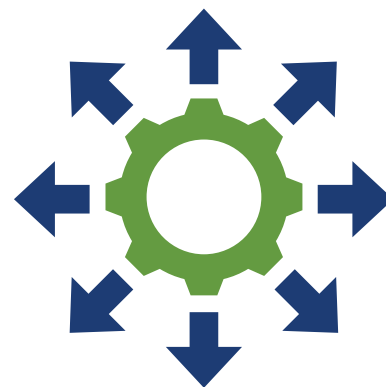
Example: A tool that is trained from characteristics from limited demographics may favor those demographics when screening resumes, leading to biased hiring decisions.

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.

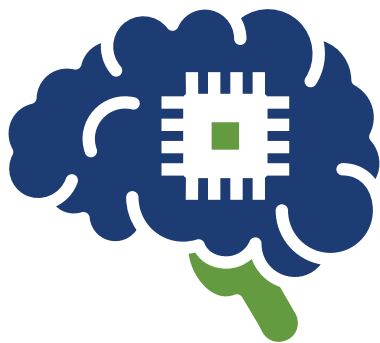
Example: In social media, computer vision is used to identify and automatically tag people in photos.

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.



- 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., Alsuehaby, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldaire, 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.pdf
- 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.bcbsook.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., Askill, 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>
- Cabitz, 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., Cows, 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>

- 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.

Guidance Letter

Best Practices for the Use of Artificial Intelligence by Mental Health Therapists

April 2025

Prepared for the Utah Office of Artificial Intelligence Policy and the Utah Division of Professional Licensing

Summary

Artificial intelligence (AI) technologies present significant opportunities to enhance mental health therapy while introducing novel ethical and practical challenges. This document establishes **guidelines for the responsible use of AI** by licensed mental health therapists within therapeutic, supervisory, and educational contexts.

Modern AI systems learn patterns from large training data sets to generate outputs (e.g., answers) for new inputs (e.g., questions). These systems can produce inaccurate or undesirable outputs. The likelihood of inaccurate or undesirable outputs can depend on the specific AI tool (e.g., some AI tools may learn wrong patterns) and the input (e.g., AI tools tend to be less reliable when either the input or the corresponding output is rare or unprecedented). We categorize AI technologies into predictive AI (i.e., systems that output specific labels or values) and generative AI (i.e., systems that produce flexible outputs like text, images, or conversational responses) and list themes to consider in weighing the advantages and disadvantages of their use.

The use of AI in mental health therapy offers several **potential benefits**. These include increased efficiency through automation of routine tasks, expanded capacity to process and analyze large amounts of patient data, and independent data-driven assessment to supplement therapist judgment. Patient-facing AI tools can provide personalized and engaging user experiences while improving the accessibility and availability of mental health support. However, **important risks** accompany these benefits. AI systems may produce inaccurate or biased outputs that lead to harmful decisions. They may respond inadequately to patients with diverse backgrounds and unique needs. Therapists may develop overconfidence in or overreliance on AI systems. Patient privacy and data security remain ongoing concerns. Technological barriers can create access inequities. Patients may experience over-immersion or overreliance on AI tools. Perhaps most fundamentally, AI interactions may lack the genuine emotional connection central to effective therapy.

The document examines various **applications of AI in mental health therapy**. These include AI-assisted risk assessment and diagnosis, monitoring of therapist performance, automated transcription and summarization of therapy sessions, treatment plan development, patient intake via conversational bots, bot-assisted therapy homework and learning experiences, and bot-administered talk therapy. Each use case presents unique benefits and challenges that require careful consideration.

Mental health therapists should adhere to several key guidelines for responsible AI implementation. Mental Health therapists must obtain **informed consent** from patients after clearly disclosing benefits, risks, and data practices. **Data privacy and safety** require verification of appropriate data handling standards and Business Associate Agreements with AI providers. **Therapist competence** demands ongoing education about AI capabilities, limitations, and proper use. **Patient safety** considerations include assessing digital literacy and risk factors for technology-related issues before implementing AI tools. Therapists should

carefully consider **unique patient needs** that may affect AI tool appropriateness and effectiveness. **Continuous monitoring** of AI outputs for accuracy and effectiveness should occur with frequency based on risk factors.

Therapists should maintain authorship transparency when interacting directly with AI systems by clearly disclosing when AI has contributed to documentation or communications. They should thoroughly **review AI-generated content** before including it in patient records or treatment decisions. Most importantly, therapists must evaluate **critically** and maintain professional judgment when considering AI-suggested diagnoses or treatments.

For patient-AI interactions, therapists should assess the **patient-technology relationship** before recommending AI tools. The **primacy of patient well-being** must be maintained, ensuring AI augments rather than diminishes the quality of care. **Contingency planning** prepares for potential service disruptions or malfunctions. Therapists must verify **ethical AI conduct** appropriate to system capabilities, understand **emergency response** limitations, and establish appropriate protocols. Regular **interaction monitoring** helps assess the effectiveness and safety of patient-AI exchanges.

Similar principles apply in educational contexts. **Student needs assessment** should precede AI incorporation in training. **Educational integrity** prioritizes effective learning outcomes over technological convenience. **A critical review** of AI-generated student assessments prevents overreliance. **Appropriate monitoring** of student-AI interactions ensures educational benefit.

These guidelines aim to help mental health therapists navigate the integration of AI technologies while upholding professional standards and prioritizing patient welfare. By implementing these best practices, therapists can harness AI's benefits while mitigating potential risks in therapeutic, supervisory, and educational settings.

Guidance Letter

Best Practices for the Use of Artificial Intelligence by Mental Health Therapists

April 2025

Prepared for the Utah Office of Artificial Intelligence Policy and the Utah Division of Professional Licensing

Summary

Artificial intelligence (AI) technologies present significant opportunities to enhance mental health therapy while introducing novel ethical and practical challenges. This document establishes **guidelines for the responsible use of AI** by licensed mental health therapists within therapeutic, supervisory, and educational contexts.

Modern AI systems learn patterns from large training data sets to generate outputs (e.g., answers) for new inputs (e.g., questions). These systems can produce inaccurate or undesirable outputs. The likelihood of inaccurate or undesirable outputs can depend on the specific AI tool (e.g., some AI tools may learn wrong patterns) and the input (e.g., AI tools tend to be less reliable when either the input or the corresponding output is rare or unprecedented). We categorize AI technologies into predictive AI (i.e., systems that output specific labels or values) and generative AI (i.e., systems that produce flexible outputs like text, images, or conversational responses) and list themes to consider in weighing the advantages and disadvantages of their use.

The use of AI in mental health therapy offers several **potential benefits**. These include increased efficiency through automation of routine tasks, expanded capacity to process and analyze large amounts of patient data, and independent data-driven assessment to supplement therapist judgment. Patient-facing AI tools can provide personalized and engaging user experiences while improving the accessibility and availability of mental health support. However, **important risks** accompany these benefits. AI systems may produce inaccurate or biased outputs that lead to harmful decisions. They may respond inadequately to patients with diverse backgrounds and unique needs. Therapists may develop overconfidence in or overreliance on AI systems. Patient privacy and data security remain ongoing concerns. Technological barriers can create access inequities. Patients may experience over-immersion or overreliance on AI tools. Perhaps most fundamentally, AI interactions may lack the genuine emotional connection central to effective therapy.

The document examines various **applications of AI in mental health therapy**. These include AI-assisted risk assessment and diagnosis, monitoring of therapist performance, automated transcription and summarization of therapy sessions, treatment plan development, patient intake via conversational bots, bot-assisted therapy homework and learning experiences, and bot-administered talk therapy. Each use case presents unique benefits and challenges that require careful consideration.

Mental health therapists should adhere to several key guidelines for responsible AI implementation. Mental Health therapists must obtain **informed consent** from patients after clearly disclosing benefits, risks, and data practices. **Data privacy and safety** require verification of appropriate data handling standards and Business Associate Agreements with AI providers. **Therapist competence** demands ongoing education about AI capabilities, limitations, and proper use. **Patient safety** considerations include assessing digital literacy and risk factors for technology-related issues before implementing AI tools. Therapists should

carefully consider **unique patient needs** that may affect AI tool appropriateness and effectiveness. **Continuous monitoring** of AI outputs for accuracy and effectiveness should occur with frequency based on risk factors.

Therapists should maintain authorship transparency when interacting directly with AI systems by clearly disclosing when AI has contributed to documentation or communications. They should thoroughly **review AI-generated content** before including it in patient records or treatment decisions. Most importantly, therapists must evaluate **critically** and maintain professional judgment when considering AI-suggested diagnoses or treatments.

For patient-AI interactions, therapists should assess the **patient-technology relationship** before recommending AI tools. The **primacy of patient well-being** must be maintained, ensuring AI augments rather than diminishes the quality of care. **Contingency planning** prepares for potential service disruptions or malfunctions. Therapists must verify **ethical AI conduct** appropriate to system capabilities, understand **emergency response** limitations, and establish appropriate protocols. Regular **interaction monitoring** helps assess the effectiveness and safety of patient-AI exchanges.

Similar principles apply in educational contexts. **Student needs assessment** should precede AI incorporation in training. **Educational integrity** prioritizes effective learning outcomes over technological convenience. **A critical review** of AI-generated student assessments prevents overreliance. **Appropriate monitoring** of student-AI interactions ensures educational benefit.

These guidelines aim to help mental health therapists navigate the integration of AI technologies while upholding professional standards and prioritizing patient welfare. By implementing these best practices, therapists can harness AI's benefits while mitigating potential risks in therapeutic, supervisory, and educational settings.

Contents

Summary.....	2
Contents.....	4
1. Introduction.....	6
2. Scope and Definitions.....	8
3. Background on Artificial Intelligence Technologies.....	10
3.1 Foundations of Modern Artificial Intelligence.....	10
3.2 Failure Modes for Modern Artificial Intelligence.....	11
3.3 Types of Artificial Intelligence.....	12
3.4 Implications for Data Privacy.....	14
4. Potential Benefits and Risks of the Use of Artificial Intelligence Technologies in Mental Health Therapy.....	15
4.1 Potential Benefits and Risks of the Use of Predictive Artificial Intelligence Technologies.....	15
4.2 Additional Potential Benefits and Risks of the Use of Generative Artificial Intelligence Technologies.....	17
5. Use Cases for Artificial Intelligence Technologies in Mental Health Therapy.....	22
5.1 Use Cases for Predictive Artificial Intelligence Technologies.....	22
<i>AI-assisted assessment of the risk of harm to self or others</i>	22
<i>AI-assisted diagnosis</i>	23
<i>AI-assisted monitoring of therapist performance</i>	25
5.2 Use Cases for Generative Artificial Intelligence Technologies.....	26
<i>Automated transcription of therapist-patient interactions</i>	26
<i>Automated summarization of therapist-patient interactions</i>	27
<i>Automated summarization of multiple patient data modalities</i>	29
<i>AI-assisted design of treatment plans</i>	30
<i>Patient intake via conversational bots</i>	31
<i>Bot-assisted patient homework</i>	33
<i>Bot-administered talk therapy</i>	34
<i>Bot-assisted learning experiences</i>	37
6. Best Practices for the Use of Artificial Intelligence Technologies by Mental Health Therapists.....	39
6.1 Best Practices for the General Use of AI Technologies.....	39

<i>Informed consent</i>	39
<i>Disclosure</i>	40
<i>Data privacy and data safety</i>	41
<i>Mental health therapist competence with AI technology</i>	41
<i>Patient safety and competence with AI technology</i>	43
<i>Unique patient needs</i>	44
<i>Continuous monitoring and reassessment</i>	45
6.2 Best Practices for Therapists' Interactions with Generative AI.....	45
<i>Authorship transparency</i>	45
<i>Review of AI-generated text</i>	46
<i>Reasonable doubt about AI-generated diagnosis and treatment proposals</i>	46
6.3 Best Practices for Patients' Interactions with Generative AI.....	46
<i>Patients' relationship with AI technology</i>	46
<i>Primary commitment to patients</i>	46
<i>Service interruptions and malfunctions</i>	47
<i>Conduct of generative AI</i>	47
<i>Responsibility in emergency situations</i>	47
<i>Continuous monitoring of patient-AI interactions</i>	48
<i>Mandated reporting of patient-AI interactions</i>	48
6.4 Best Practices for Supervisors' Interactions with Generative AI.....	48
<i>Review of AI-generated text</i>	48
<i>Reasonable doubt about AI-generated student assessments</i>	48
6.5 Best Practices for Students' Interactions with Generative AI.....	49
<i>Student's relationship with AI technology</i>	49
<i>Primary commitment to students</i>	49
<i>Service interruptions and malfunctions</i>	49
<i>Continuous monitoring of student-AI interactions</i>	49
7. Authorship Note and Acknowledgments.....	51
8. References.....	52

1. Introduction

The rapid advance of artificial intelligence (AI) technologies, particularly generative AI technologies, has the potential to substantially transform the behavioral health field. These tools may offer promising opportunities to enhance behavioral healthcare (Thakkar et al., 2024; Ettman and Galea, 2023; Olewade et al., 2024). Examples include improved diagnostics, increased accessibility to mental health services, and enhanced patient-therapist interactions. These and other developments may improve the experiences of patients and therapists. Patients may enjoy easier access to high-quality care. Since AI technologies tend to excel specifically at automating menial and repetitive tasks, therapists may find it particularly beneficial to transfer such tasks to AI and focus their efforts on fostering the human connection with their patients and building robust therapeutic relationships.

However, new challenges accompany the new opportunities that emerge with AI technologies. Integrating AI technologies into behavioral health introduces ethical, legal, and practical challenges. Concerns about data privacy and security arise in using AI technologies, as these technologies often involve recording, storing, and sharing data, which, in the context of behavioral health services, can include sensitive patient information. Issues around informed consent, transparency, and accountability become increasingly complex in the context of AI-assisted behavioral health services. Additionally, when a mental health therapist or a patient becomes over-reliant on the use of these technologies, it can have detrimental effects on the provided service and the patient's well-being. Generative AI technologies, specifically, have the potential to generate unpredictable and, thus, potentially inaccurate responses in conversations with mental health therapists and potentially insensitive and/or harmful responses in discussions with patients. These and other risks of the use of AI technologies can create conflicts between the desire to enjoy the benefits of the new technology and a therapist's duty to align their practice with the ethics code of their profession (American Association for Marriage and Family Therapy, 2015; American Mental Health Counselors Association, 2020; American Psychological Association, 2016; American Psychiatric Association, 2013; Association of Social Work Boards, 2014; National Association for Alcoholism and Drug Abuse Counselors and National Certification Commission for Addiction Professionals, 2021; National Association of Social Workers, 2021; National Board for Certified Counselors, 2023).

On the one hand, the great potential for benefits motivates a thorough exploration of AI technologies and their various use cases in behavioral health. On the other hand, the substantial risks associated with using these technologies in behavioral health create a responsibility to prevent and mitigate harm that may arise from such use. Healthcare organizations, healthcare providers, and providers of AI tools all carry a part of this responsibility. How should a mental health therapist who wishes to use AI tools navigate the new technologies and the associated responsibilities? This document aims to establish clear guidelines on the ethical and responsible use of AI by licensed professionals in behavioral health services.

To this end, this document includes a primer on the current state of AI, a general risk-benefit comparison for different types of AI, a collection of use cases of AI technologies in mental health therapy, and a list of guidelines for using AI technologies by mental health therapists. These guidelines are meant to help mental health therapists of behavioral healthcare navigate the complexities of introducing AI technologies while safeguarding patient well-being and upholding professional standards.

2. Scope and Definitions

This guide is concerned with the use of generative AI technologies that can assist a mental health therapist in a licensed profession in offering services within the scope of their professional license to patients, supervision to colleagues in their profession, and education and training to existing licensees (i.e., continuing education), as well as students seeking licensure. It does not concern the use of AI tools outside a dedicated therapeutic relationship between a patient and their licensed mental health therapist. The use of conversational bots aimed at improving mental health and/or well-being falls only within the scope of this guide when a patient has an ongoing therapeutic relationship with a licensed mental health therapist who recommended or prescribed the use of the conversational bot to the patient.

For some concepts that appear frequently in this document, one can find different definitions in different sources. This report assumes the following definitions of the concepts, most of which are in accordance with the Utah State Code.

Artificial intelligence

“Artificial intelligence” means a machine-based system that makes predictions, recommendations, or decisions influencing real or virtual environments. (Utah State Code 13-72-101)

Artificial intelligence technologies

“Artificial intelligence technology” means a computer system, application, or other product that uses or incorporates one or more forms of artificial intelligence. (Utah State Code 13-72-101)

Generative artificial intelligence

“Generative artificial intelligence” means an artificial system that (i) is trained on data; (ii) interacts with a person using text, audio, or visual communication; and (iii) generates non-scripted outputs similar to outputs created by a human, with limited or no human oversight. (Utah State Code 13-2-12)

Mental health therapist

"Mental health therapist" means an individual who is practicing within the scope of practice defined in the individual's respective licensing act and is licensed as:

1. a physician and surgeon or osteopathic physician engaged in the practice of mental health therapy;
2. an advanced practice registered nurse specializing in psychiatric mental health nursing;
3. an advanced practice registered nurse intern specializing in psychiatric mental health nursing;
4. a psychologist qualified to engage in the practice of mental health therapy;
5. a certified psychology resident qualifying to engage in the practice of mental health therapy;

6. a physician assistant specializing in mental health care under Section 58-70a-501.1 of the Utah State Code;
7. a clinical social worker;
8. a certified social worker;
9. a marriage and family therapist;
10. an associate marriage and family therapist;
11. a clinical mental health counselor;
12. an associate clinical mental health counselor;
13. a master addiction counselor or
14. an associate master addiction counselor.

(Utah State Code 58-60-102)

Patient / Client

"Patient" means an individual who consults or, is examined, or is interviewed by an individual licensed under this chapter who is acting in the individual's professional capacity (Utah State Code 58-60-102).

Supervisor / Teacher

"Supervisor" or "teacher" means a licensed mental health therapist offering formal training and/or expert or clinical supervision to a student of a licensed behavioral health profession or to another licensed mental health therapist.

Student / Trainee

"Student" or "trainee" means an individual receiving formal training and/or expert or clinical supervision of their service in a licensed behavioral health profession by a licensed mental health therapist.

3. Background on Artificial Intelligence Technologies

Several times in human history, a new technology has transformed how we work, live, and participate in society. The invention of the combustion engine, for example, has led to the development of motorized vehicles and thus transformed how we travel and transport goods. The discovery of semiconductor technology has led to the development of digital calculators and, later, personal digital computers, and thus, has transformed how we calculate and store information. The rise of such transformative technologies tends to be followed by a period of uncertainty during which

- entrepreneurs test the scope and limitations of using the new technology to create new products and services and/or make the production and delivery of existing products and services more efficient via automation,
- small-business owners and service providers test how they can integrate these new products and services in their business operations and interactions with their clients in a beneficial way and
- educators and workers try to identify what knowledge one should acquire to participate effectively in the economy and society as it is changed by new technology.

The current period of uncertainty follows the development of new AI technologies, which were made possible by a collection of computing hardware and algorithm innovations. Debates on the level of technological knowledge needed to effectively and safely use AI technologies in one's professional and personal life are ongoing. The remainder of this section offers concise explanations of the foundations of modern AI technologies, types of AI technologies, and the data-privacy implications of using these technologies.

3.1 Foundations of Modern Artificial Intelligence

Most attempts to assess human intelligence in an educational setting or to use human intelligence in a professional setting are based on finding correct answers for some specific questions. For example, in an elementary school math exam, students are given sets of math problems and tasked to find a correct solution for each problem. A part of the intellectual work of a mental health therapist is to offer a considerate and contextually sensitive response to a patient's communications. Another part is to consider a patient record and formulate a correct or at least plausible diagnosis. In machine learning, the question, including the information on which the answer should be based (e.g., a patient record), is called an *input*. The answer (e.g., a diagnosis) is called an *output*. Modern machine-learning algorithms are based on showing a computer program many pairs of inputs and the corresponding desired outputs. The computer program can learn a *pattern* that connects the inputs shown with their corresponding outputs. A computer program that has learned a pattern in such a way is often called *artificial intelligence*

(AI). The set of input-output pairs shown to an AI to facilitate its learning of a pattern is called *training data*. Once the AI has learned the pattern, it can use the learned pattern to generate an output for an input that has not been shown before. This means that a user of the AI can prompt it with a question for which the AI has never seen the correct answer, and the AI will respond with a plausible answer based on its learned pattern.

3.2 Failure Modes for Modern Artificial Intelligence

This paradigm of training a machine or computer by showing it a training data set is the foundation of modern AI technologies. It is a powerful paradigm, but the trained AIs can generate wrong or undesired outputs in individual cases. These failures of AIs are typically associated with one of the following failure modes:

1. **The learned pattern includes errors.** The machine-learning paradigm described above makes it possible for correlational relationships among features of inputs and outputs to be learned as causal relationships. For example, a computer may be shown a set of patient records and corresponding diagnoses. The training data may contain many examples of women in the 20-40 years age range who have been diagnosed with postnatal depression. This may prompt a computer to learn that women in the 20-40 age range are likely to have postnatal depression, regardless of whether they have any children or have ever been pregnant. This error in the learned pattern may lead a computer to propose postnatal depression as a diagnosis for a childless 28-year-old woman who has no children and has never been pregnant.
2. **An input is outside the scope of an AI's intended use.** Many "AI-powered" products are designed for a specific and narrow use. For example, a developer may develop an AI assistant to summarize talk therapy sessions for mental health therapists. With this purpose in mind, the developer uses training data that includes transcripts of therapy sessions and corresponding high-quality summaries of these sessions. The developer then makes the AI assistant available to mental health therapists, one of whom discovers that it is possible to submit patient records instead of therapy transcripts as inputs to the AI assistant and receive output. While the output may look like a proper summary of the patient record, this summary may be missing critical information about the patient because the computer has never seen a high-quality summary of a patient record and thus does not know how to generate one.
3. **The correct output requires case-specific information that is not included in the input.** Sometimes, the input does not include case-specific information important for generating the correct output. This lack of information can lead AI tools to generate inaccurate outputs. Consider an AI tool designed to make a diagnostic recommendation based on transcripts of a patient's conversation with their therapist. The transcripts may fail to capture the patient's mannerisms, gestures, tone, and mood during the conversation. The therapist who has noticed these contextual features during the

conversation may thus develop a more accurate diagnosis than the AI tool that has no access to the contextual information.

4. **An input is rare, at least in the training data.** If some input scenarios are rare or missing in the training data, the computer may have seen few or no instances of this input during training. The learned pattern then does not apply to the rare input scenario, and consequently, the computer generates inaccurate outputs for these input scenarios. For example, an AI-powered chatbot designed to assist individuals with anxiety may have been primarily trained on conversations related to social anxiety and generalized anxiety disorder. Suppose a user presents with a rare anxiety subtype, such as health anxiety or obsessive-compulsive disorder (OCD). In that case, the chatbot might fail to provide appropriate support or misinterpret the user's concerns.
5. **An output is rare, at least in the training data.** A computer may have seen few or no examples of this output in the training data if a particular diagnosis, treatment plan, or intervention is uncommon. The learned pattern then does not connect inputs to these rare outputs, and it may be that the computer generates these rare outputs with a much lower frequency than desired or not at all. Patients with rare conditions or who would benefit from uncommon interventions may thus be disadvantaged when the output of an AI tool solely determines diagnoses or treatment plans.
6. **An input or output is under or overrepresented in the training data.** An AI tool may generate inaccurate outputs when some input-output pairs are under or overrepresented in its training data, even if all possible inputs and outputs are included in the training data. For example, suppose the training data for an AI-powered depression screening tool primarily consists of cases from urban populations. In that case, the AI's learned pattern may put too much focus on environmental stressors and not give enough consideration to the lack of access to resources as a risk factor for depression.
7. **Inputs include incorrect assertions.** Some conversational bots and other generative AI tools sometimes provide “hallucinated” outputs that include decidedly wrong information. Hallucinations tend to happen when generative AI tools are asked to provide information they cannot access or when the query includes wrong assertions. For example, when asking an AI to “summarize the patient's relationship with their sister” for a patient who does not have siblings, the AI may describe the patient's relationship with a made-up sibling.

3.3 Types of Artificial Intelligence

To understand the landscape of AI products, it can be helpful to distinguish between two types of AI based on the types of outputs that they generate: predictive AI and generative AI (Narayanan and Kapoor, 2024).

Predictive AI is an umbrella term for AI tools whose output is a prediction of a label for an input. Examples include

- diagnostic tools that recommend one of several possible diagnoses given a patient's data,
- sentiment analysis tools that label a patient's mood (e.g., a value out of the five words "happy," "sad," "angry," "neutral," or "anxious") based on a transcript of their conversation with their mental health therapist, and
- risk assessment tools that output a value (e.g., a value chosen out of the three labels "high," "medium," "low," or a number ranging from 0 to 1) to describe a person's risk of harming themselves or others.

Because these AI tools have substantial limitations on what outputs they generate, it is easy (or at least easier than in some other cases) to enumerate all possible failure scenarios and develop mitigation strategies for each scenario.

Generative AI is an umbrella term for AI tools that are less restricted in their outputs. Outputs can include free-form text, images, audio, or video. Examples include

- summarization tools that, when given a long text (e.g., a transcript of a talk therapy session) as input, generate a short summary of the input text,
- recommendation tools that generate recommended personalized treatment plans for a therapist based on a patient's data and
- chatbots or other conversational bots that engage in conversation with patients outside of therapy sessions with their therapist.

The flexibility of generative AI tools makes it impossible to account for every output they can generate and how that output may aid or harm therapists and patients. In this sense, the use of generative AI tools is more similar to using a car than predictive AI. There are simply too many ways in which accidents can happen to prepare for every single one of them individually. The skill of safely driving a car is obtained through many hours of driving experience and a driving instructor's guidance. Similarly, users of generative AI tools can learn to anticipate the likelihood of failure and quickly respond to failures of the AI tool appropriately through extensive experience with the AI tool in a controlled or assisted setting. Accordingly, our guidance on the safe use of AI technologies in mental health therapy includes sections specific to generative AI use.

Drawing a clear line between predictive AI and generative AI becomes difficult when predictive AI tools have many possible outputs or when generative AI tools are flexible enough to be used to predict labels. For example, it is in principle possible—although not recommended as it has the potential to violate HIPAA and related privacy laws—to share a patient record with a general-purpose language model (e.g., OpenAI's ChatGPT, Google's Gemini, Microsoft's

Copilot, or Anthropic's Claude) and ask it to generate a diagnosis, thereby effectively performing the task of a predictive AI. While the language model will likely comply, it is important to remember that these general-purpose AI tools are not designed for this use and that the quality and accuracy of diagnostic recommendations obtained in such a way can vary strongly. Some AI tools also combine a predictive AI for a specialized purpose with a generative AI (e.g., a language model), which lets users interact with the predictive AI via written or spoken language instead of requiring them to learn how to use a specific graphical user interface.

3.4 Implications for Data Privacy

The modern machine learning paradigm requires large amounts of data to continue improving the capabilities of AI tools. The demand for large data sets creates a strong incentive for providers of AI tools to collect data from their customers' interactions with their AI tools. Specifically, providers of AI tools for mental health services and mental-health-related services may be incentivized to collect and use patient data to improve their AI systems and sell these datasets to other entities for AI development purposes.

The practice of collecting and selling/sharing sensitive data is always associated with risks, such as

- data accidentally being made publicly available through human error,
- data being misused by one of the persons charged with data handling or
- data being stolen and misused or published by an external person.

In addition to these general risks, the collection and selling/sharing of sensitive data for training AIs bear further privacy risks because it may be possible for users to retrieve sensitive information about individual people whose data was included in the AI's training via *adversarial queries* (Nasr et al., 2023; Schwarzschild et al. 2025). Typically, these adversarial queries are inputs that users have meticulously and purposefully designed to trick AIs into revealing sensitive data. In rare cases, however, it is also possible that the benign use of an AI leads to the disclosure of sensitive data.

The likelihood of adversarial queries to an AI successfully leading to the disclosure of sensitive data tends to be low. Still, it changes frequently as AI developers improve AI security and hackers strengthen their capabilities. This complex landscape of privacy risks makes it more important than ever for mental health therapists to

- carefully review service agreements, business associate agreements, and other agreements that govern the collection and use of patient data and

- obtain a patient's informed consent before using AI tools that record or transcribe therapy sessions or involve interactions between a patient and a conversational bot beyond patient-intake procedures.

4. Potential Benefits and Risks of the Use of Artificial Intelligence Technologies in Mental Health Therapy

Emerging AI technologies developed and/or used in mental health therapy can be associated with a host of potential benefits and risks. While the concrete, complete risk-benefit profiles of each AI tool will depend on the specific tool, use case, and context, it is possible to identify general themes with which the potential benefits and risks of many AI tools in mental health therapy align.

4.1 Potential Benefits and Risks of the Use of Predictive Artificial Intelligence Technologies

Use cases for predictive AI in mental health therapy involve analyzing data to help therapists make decisions (Lee, 2021). Accordingly, **potential benefits** tend to fall into one of three categories:

- **Increased efficiency through automation.** The AI-assisted automation of decision-making tasks can help therapists and therapist supervisors to work more efficiently. For therapist and therapist supervisors working long hours, saving time on tasks that can be made more efficient through the use of AI technologies can improve their work-life balance and personal health, and it can improve the quality of the therapeutic, educational, and supervisory services that they offer to their patients, students, and trainees. The increased efficiency can also allow therapists to serve an increased number of patients and therapist supervisors to serve an increased number of students and trainees.
- **Increased scope or capacity through automation.** AIs can process large amounts of data quickly, thereby allowing therapists and therapist supervisors to base their assessments, recommendations, and decisions on holistic analyses that would be impractical to do otherwise. This increase in capacity has the potential to lead to more accurate diagnoses, a higher quality of care, and, in some cases, greater transparency in the therapist's decision-making process.
- **Independent data-driven assessment.** Therapists can compare their assessments of patient conditions, therapy sessions, etc., to those provided by an AI. This data-driven "second opinion" can provide a therapist with additional insights that can improve

decision-making. Similarly, an independent data-driven perspective can potentially aid therapist supervisors in assessing their trainees' performance and progress more holistically and/or more accurately.

Risks tend to fall into one of six categories:

- **Risk of inaccurate or wrong AI outputs.** Many commercially available AI tools advertise their high accuracy and low failure rates. Nonetheless, it is almost always possible for AIs to generate inaccurate or wrong outputs, for example, via one of the failure modes listed under [Failure Modes for Modern Artificial Intelligence](#).
- **Risk of inadequate response to patient backgrounds and unique needs.** Several examples of the use of predictive AI on large populations of people have highlighted that, in such cases, subpopulations can be negatively affected. Some AI tools provide accurate predictions and good recommendations on average but consistently lead to undesirable outcomes for patients with specific backgrounds. These systematic deviations occur when some patient backgrounds or unique needs are not adequately represented in the AI's training data (see [Failure Modes for Modern Artificial Intelligence](#)). The resulting harms can be hard or even impossible to detect by individual patients or therapists because they may only become evident when comparing several large patient groups with different backgrounds and/or needs. Undetected and unmitigated, these inadequate responses to patient backgrounds and unique needs can lead to large groups of patients receiving inadequate care.
- **Risk of human overconfidence in AI outputs.** Some of the potential harm from wrong or implausible AI outputs can be prevented or mitigated through appropriate human oversight. When therapists or therapist supervisors become overconfident in their AI tools, however, they may decide
 - not to review an AI's output or
 - to uncritically trust an AI's judgment over their own, even when they notice discrepancies.

Therapists who neglect their responsibility to review AI outputs critically may fail to prevent harm to patients. In educational settings, inaccurate AI outputs may harm a trainee's education if their supervisor does not detect and correct them.

- **Risk of human overreliance on AI assistance.** A therapist who offloads many of their analysis and decision-making tasks on AI tools can become so reliant on these tools that they feel unequipped to perform these tasks without AI tools. Being able to analyze patient data and make therapy-related decisions independently ensures that therapists can
 - critically review AI outputs,
 - detect mistakes made by AIs, and

- produce accurate diagnostic analyses, well-supported therapeutic decisions, and treatment plans when an AI tool fails or becomes unavailable.

AI tools may become unavailable for various reasons, including but not limited to computer viruses, internet or power outages, or product discontinuation by or bankruptcy of the providing company.

- **Privacy risks.** The use of AI tools on patient or student data requires this data to be collected and motivates its digital storage. The collection and storage of such data are associated with the risk of breaches of confidentiality through accidental or purposeful data mishandling. If the collected data is used to train AIs, an additional privacy risk arises because AIs can leak training data in response to adversarial queries (see [Implications for Data Privacy](#)).

4.2 Additional Potential Benefits and Risks of the Use of Generative Artificial Intelligence Technologies

Generative AI has facilitated the development of capable and flexible conversational bots. These conversational bots include chatbots, audiobots, and “digital humans” (i.e., animated videobots). The various potential use cases of such conversational bots can involve therapists, patients, therapist supervisors, and therapist trainees as users. For these user groups, using such conversational bots tends to be associated with the potential benefits and risks specific to human-machine interactions in which a machine mimics human responses. **Potential benefits** of such interactions tend to fall into one of five categories:

- **Ease of use.** When querying or adding data to a patient record or other data collection would otherwise involve a user interface with various dropdown menus, checkboxes, and so on, sharing the data request or the new data with a conversational bot can make the process easier or more pleasant for the user.
- **Engaging user experiences.** Due to the immersive nature of interactions with conversational bots, users who feel motivated or energized by interactions with others may be more engaged in tasks that involve conversational bots than in tasks that do not involve any conversational element.
- **Personalized service.** Conversational bots may have access to detailed information about their users through memory of their conversations or other means. A conversational bot developed to customize its service based on such data may use this information to create personalized user experiences that match a user’s individual preferences or needs.
- **Increased sensitivity to a user’s background and unique needs.** If developed accordingly, a conversational bot can adjust its responses to consider a user’s unique

needs that may arise from their economic, racial, ethnic, cultural, or other backgrounds. In a therapeutic setting, a conversational bot sensitive to its patient's background can help patients feel validated and understood, especially when it is difficult for their therapists to relate to the patient's personal experiences. In an educational setting, a conversational bot sensitive to a student's background can potentially provide more effective learning experiences.

- **Increased availability of support.** A conversational bot whose purpose is to support a patient in maintaining or achieving mental health and well-being can, at least in principle, be available anywhere and at all times via a smartphone application or similar device or service. This creates an opportunity for improved care for patients who would benefit from immediate support on demand and/or continuous 24-hour support. A conversational bot developed for educational purposes can support students in their learning whenever the students decide to devote time to their studies.

Risks tend to fall into one of six categories:

- **Technological barriers for users.** Access to technology, including reliable internet connectivity and suitable devices (e.g., smartphones or computers), is not evenly distributed across all populations. This can create a barrier for individuals from low-income communities, rural areas, and certain marginalized groups, potentially exacerbating existing inequalities in access to mental healthcare. Additionally, positive and successful user experiences may require users to have
 - a certain level of trust in a technology's capability and alignment and
 - a certain level of technical proficiency.

Users with a low level of trust in the technology or technical proficiency may face challenges in effectively interacting with these tools. They may feel intimidated or frustrated by the experience. This tech-induced frustration can make therapy for these patients less effective. For therapists in training, tech-induced frustration can lead to poor learning outcomes.

- **User over-immersion.** Some conversational bots can mimic human behavior in very immersive ways. A user of such a bot may form an emotional connection with the bot and may prefer excessive interactions over healthy human social connections, negatively affecting their emotional well-being (Fang et al., 2025). In such situations, users may
 - neglect meaningful real-life relationships,
 - fail to develop healthy coping mechanisms for challenges that they experience in social interactions,
 - suffer a decline in essential social skills and a diminished capacity for genuine human connection,

- develop a diminished sense of agency and autonomy through overreliance on the conversational bot's guidance and support,
 - experience anxiety and other withdrawal symptoms when access to the AI tool is limited.
- **Lack of emotional connection.** While AI tools can mimic human language and emotional expressions, they cannot fully replicate a human therapist's nuanced emotional intelligence and empathy. This can lead to a lack of genuine emotional connection and understanding, potentially hindering the therapeutic process. An AI tool may struggle to accurately interpret and respond to subtle emotional cues, such as tone of voice, body language, and micro-expressions, potentially leading to a user feeling isolated and misunderstood.
 - **Lack of sensitivity to a patient's background and unique needs.** AI tools may not be adequately equipped to understand and respond appropriately to the unique cultural nuances and values of diverse patient populations. This can lead to culturally insensitive or inappropriate responses from conversational bots, potentially alienating and offending users from different cultural backgrounds. Specifically, if the datasets used to train AI models lack diversity (see [Failure Modes for Modern Artificial Intelligence](#)), the AI may not be able to effectively address the specific mental health needs of patient populations that are underrepresented or not represented in those data sets.
 - **Undesired consequences from out-of-scope use.** AI tools can be misused or abused, potentially leading to harmful consequences. Over-reliance on generative AI tools could potentially erode professional standards in mental healthcare. This could lead to a decline in the quality of care therapists provide and a devaluing of the essential human element in the therapeutic process.
 - **Declining engagement over time.** Digital mental health interventions, including AI-powered tools, often face challenges with user retention and sustained engagement. Initial enthusiasm may quickly wane as users experience intervention fatigue, leading to decreasing usage patterns and eventual abandonment. This dropout phenomenon can result in
 - incomplete therapeutic processes that fail to deliver meaningful clinical outcomes,
 - false impressions of treatment efficacy when only engaged users are measured in evaluations,
 - frustration and self-blame among users who interpret their disengagement as personal failure rather than a design limitation,
 - resource misallocation when healthcare systems invest in technologies that remain underutilized beyond initial adoption phases.

The challenge is particularly pronounced for users with certain mental health conditions that inherently impact motivation and follow-through, potentially creating a paradoxical situation where those who might benefit most struggle the most to maintain consistent engagement with digital interventions.

- **Hallucinations and harmful advice.** Some conversational bots and other generative AI tools sometimes provide “hallucinated” outputs that include decidedly wrong information (see [Failure Modes for Modern Artificial Intelligence](#)). If unchecked, these hallucinated outputs may lead therapists to adopt ineffective or even dangerous treatment plans or make decisions based on false or misleading information. Generative AI tools interacting with patients may directly provide harmful advice to a patient, instill or strengthen harmful beliefs and opinions that a patient holds, or encourage patients to act in ways that are harmful to themselves or others.

Together with the risk categories listed under [Potential Benefits and Risks of the Use of Predictive Artificial Intelligence Technologies](#), the risk categories listed here outline the risk landscape for the use of generative AI tools. A comparison with the list of categories of potential benefits of generative AI tools suggests that many categories of potential benefits and risks can be paired as two sides of the same coin. For example:

- Conversational bots create **new ways to interact with information technology**. For some, these new ways may improve the ease of use. For others, these new ways to interact with technology can cause tech-induced frustration.
- **Immersive interactions with conversational bots** can have positive effects on users. However, users who become over-immersed may experience various adverse effects.
- Generative AI tools can create **personalized user experiences** using personal user data. These personalized experiences can make patients feel understood and validated. However, patients can feel misunderstood or offended when the user experience is not adequately personalized. For these patients, AI-augmented therapy may be less effective than a traditional therapeutic approach that focuses on building an emotional connection with a therapist.
- Generative AI tools may be equipped to converse appropriately (sometimes even more so than a therapist) with **patients of various economic, racial, ethnic, cultural, or other backgrounds** if the diverse challenges associated with these backgrounds were addressed during the AI’s development. If they were not addressed during the AI’s development, the AI tool may be particularly insensitive to a patient’s background and unique needs and thereby risk alienating or offending the patient.

Assessing whether potential benefits outweigh the risks when introducing a generative AI tool in therapy can be difficult. The outcome of such a risk-benefit analysis can depend on how a specific AI tool was developed, a patient’s background, relationship to technology, mental health

conditions, and how the therapist uses (or instructs the patient to use) the AI tool to augment the patient's therapy.

5. Use Cases for Artificial Intelligence Technologies in Mental Health Therapy

This section includes a non-exhaustive list of use cases for AI in mental health therapy and their risk and benefit profiles. Many AI products currently advertised for this use tend to address one or several of these use cases.

5.1 Use Cases for Predictive Artificial Intelligence Technologies

AI-assisted assessment of the risk of harm to self or others

AI tools can analyze patient data, such as past medical records, therapy notes, and - with appropriate consent - even social media activity to predict the likelihood of self-harm or harm to others.

Potential benefits

- **Increased efficiency through automation.** Automating parts of the risk assessment process can reduce the time it takes a therapist to conclude a risk assessment. Identifying patients with an elevated risk of self-harm or violence more quickly can create a greater window of opportunity for timely interventions that can potentially prevent tragic outcomes.
- **Increased scope or capacity through automation.** Time saved through automating part of the risk assessment process can free a therapist to focus more on other complex aspects of patient care.
- **Independent data-driven assessment.** By analyzing vast amounts of data, AI can potentially improve the accuracy of risk assessments compared to traditional methods, which may be subjective and prone to human bias.

Risks

- **Risk of inaccurate AI outputs.** AI tools may generate inaccurate predictions, leading to unnecessary interventions (i.e., false-positive predictions) or missed opportunities for intervention (i.e., false-negative predictions). Inaccurate assessment can occur, for example, when conversations include language nuances such as metaphors, idioms, irony, or sarcasm, which may be challenging for an AI to interpret correctly. Another risk factor is situations where the correct interpretation of a conversation requires important context (e.g., a patient's cultural or religious background) that may affect how they express themselves and what cultural norms give the context for their relationships with others. These and other factors can lead to inaccurate risk assessments, and some of

these inaccurate assessments can have detrimental and potentially even fatal consequences.

- **Risk of human overconfidence in AI assistance.** A therapist's expert knowledge and personal understanding of a patient's experiences can be necessary for an accurate risk assessment. When a therapist uncritically trusts an AI's risk assessment over their own, these essential factors are not included in the risk assessment. The resulting risk assessment may be less accurate and ultimately detrimental to the patient and the people around them.
- **Risk of human overreliance on AI assistance.** Like any other digital tool, an AI-driven risk-assessment tool can become unavailable anytime for various reasons (see [Potential Benefits and Risks of the Use of Predictive Artificial Intelligence Technologies](#)). **Individual patients may also** choose to opt out of AI tool usage in their therapy. In such cases, therapists must realize all aspects of their responsibilities, including assessing a patient's risk of self-harm and harm to others without the AI tool.
- **Risks to patient privacy.** These AI technologies can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries. Risk-assessment tools used by mental health therapists may collect past medical records, therapy notes, and other personal data for their assessment. Using an AI-driven risk-assessment tool thus bears the risk of a severe breach of confidentiality.

AI-assisted diagnosis

AI-assisted diagnostic tools can analyze patient data to aid therapists in a clinical setting in identifying mental health conditions.

Potential benefits

- **Increased efficiency through automation.** A therapist's diagnosis of a patient's mental health conditions may depend on the therapist's impressions gathered from interactions with their patients and various patient data, such as intake forms, medical records, and referral notes from other therapists. Reviewing these and other patient data can be partially or fully automated using AI tools. Automating the data review can increase efficiency in a therapist's diagnostic work.
- **Increased scope or capacity through automation.** Automating diagnostic tasks can allow therapists to diagnose a larger number of patients. Additionally, therapists may consider a larger amount of patient information in each diagnosis because automated data review allows therapists to analyze large amounts of data quickly.

- **Independent data-driven assessment.** A therapist or an AI (or both) can arrive at inaccurate diagnoses for various reasons. A discrepancy between the therapist's diagnosis and a diagnosis obtained from an AI's review of the patient's data indicates that at least one of these diagnoses is incomplete or inaccurate. The discrepancy thus motivates a critical review of both diagnoses. Using the output of a diagnostic AI tool to contrast a therapist's diagnoses of their patient's mental health can increase the accuracy of the therapist's diagnoses.

Risks

- **Risk of inaccurate AI outputs.** Diagnoses provided by AI tools can be inaccurate. The risk of an inaccurate diagnosis can be elevated when
 - the accurate diagnosis would include rare conditions or conditions that may have been underrepresented during an AI tool's training and development or
 - demographic, cultural, religious, or other aspects of a patient's background may affect how a condition presents itself and may have been underrepresented during an AI tool's training and development.
- **Risk of human overconfidence in AI assistance.** A therapist's expert knowledge and personal understanding of a patient's experiences can be necessary for arriving at an accurate and complete diagnosis of a patient's mental health conditions. When a therapist uncritically trusts an AI's diagnosis over their own, these essential factors are not included in the diagnosis, which may thus be less accurate and ultimately detrimental to the patient.
- **Risk of human overreliance on AI assistance.** Like any other digital tool, an AI-driven diagnostic tool can become unavailable anytime for various reasons (see [Potential Benefits and Risks of the Use of Predictive Artificial Intelligence Technologies](#)). **Individual patients may also** choose to opt against the use of such an AI tool in their therapy. In such cases, it is important that therapists can realize all aspects of their responsibilities, including their diagnostic work, without the AI tool.
- **Risks to patient privacy.** These AI technologies can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries. Diagnostic AI tools used by mental health therapists may collect past medical records, therapy notes, and other personal data for their assessment. Using an AI-driven diagnostic tool thus bears the risk of a severe breach of confidentiality.

AI-assisted monitoring of therapist performance

AI tools can evaluate therapist's performance based on transcriptions or recordings of their interactions with patients. Tools can calculate performance scores according to performance metrics set by a therapist for self-monitoring or a supervisor for training.

Potential benefits

- **Increased efficiency through automation.** AI-driven performance-monitoring tools may help supervisors provide performance feedback to trainees more efficiently than before by automating several aspects of the review and analysis of the trainees' therapist-patient interactions. In self-monitoring and training settings, AI-driven performance monitoring can improve treatment fidelity by evaluating how faithfully or accurately a therapist implements a treatment plan.
- **Increased scope or capacity through automation.** The efficiency gained through automating some aspects of therapist supervision may allow a supervisor to supervise a larger number of trainees simultaneously. It may also allow trainees to receive more detailed feedback on their performance than from a supervisor who has to divide their time across several trainees and other job responsibilities.
- **Independent data-driven assessment.** An AI-driven performance monitoring tool can provide a data-driven performance assessment that is not subject to the same biases that might affect assessments from human supervisors, which might, in some cases, rely too heavily on their judgment of a trainee's character rather than their most recent performance. In addition to efficiency gains, if a supervisor selects one or several performance measures to be used over an extended supervision period, AI-calculated performance measures can improve the consistency of their trainees' performance scores. If the supervisor also shares this decision and the selected performance measures with their trainees, trainees can benefit from a more transparent assessment process with clear avenues for self-improvement.

Risks

- **Risk of inaccurate AI outputs.** AI tools can produce inaccurate performance scores. Such inaccurate scores can result from the AI tool misinterpreting aspects of the therapist-patient interaction and/or the therapist's approach to treatment. These misinterpretations are particularly common when the therapist-patient interactions include language nuances such as metaphors, idioms, irony, or sarcasm, which may be challenging for an AI to interpret correctly. Another risk factor is situations where the correct interpretation of a conversation requires important context (e.g., a patient's cultural or religious background) that may affect how they express themselves and react to their therapist's suggestions.
- **Risk of human overconfidence in AI assistance.** A supervisor or trainee who is overconfident in the AI tool may decide to trust an AI's assessment of a trainee's

performance over their assessment, even when these two assessments conflict. This may cause inaccurate performance assessments to detrimentally influence trainees' self-perception, learning strategy, and professional skill development.

- **Risk of human overreliance on AI assistance.** Like any other digital tool, an AI-driven performance assessment tool can become unavailable anytime for various reasons (see [Potential Benefits and Risks of the Use of Predictive Artificial Intelligence Technologies](#)). It is also possible that individual trainees or their patients may choose to opt against the use of such an AI tool in their therapy. In such cases, a supervisor needs to be prepared to fulfill their supervisory responsibilities in any such case.
- **Risks to patient privacy.** These AI technologies can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on trainees' and patients' data may reveal data in response to adversarial queries. Performance-assessment tools used by mental health therapists or their supervisors may collect
 - copies of text interactions between therapists and their patients,
 - transcripts, or audio or video recordings of person-to-person therapist-patient interactions,
 - and other sensitive data.

Using an AI-driven performance-assessment tool thus bears the risk of exposing sensitive trainee or patient data.

5.2 Use Cases for Generative Artificial Intelligence Technologies

Automated transcription of therapist-patient interactions

Speech-to-text technology enables therapists to transcribe spoken words during therapist-patient interactions (e.g., talk therapy sessions) into written text automatically.

Potential benefits

- **Increased efficiency through automation.** Automated transcription alleviates the need to take manual notes during a therapy session. The transcripts or AI-generated summaries can help therapists efficiently write reports on patients' status and progress.
- **Increased scope or capacity through automation.** Reducing the labor associated with note-taking during therapy sessions frees therapists to fully focus on the interaction with their patients while the interactions occur.

- **Independent data-driven assessment.** Automated transcripts may lead to more comprehensive and accurate patient records. Therapists can use the computerized transcripts to review the exact dialogue of a therapy session when designing a personalized treatment plan.

Risks

- **Risk of inaccurate AI outputs.** AI-generated transcripts can include transcription errors that can misrepresent the conversation. Such inaccuracies can arise, for example, due to algorithms with better speech recognition for certain accents or forms of expression. The inaccurate information in transcripts can harm patients and therapists in various ways. For example, inaccurate transcripts can lead to inaccurate patient records and, in turn, to patients receiving inaccurate diagnoses and inefficient treatment plans from their therapist and/or other healthcare providers. Additionally, if transcripts are subpoenaed or divulged via discovery, courts may assume that the information in these transcripts is accurate and allow them as evidence in legal disputes. In extreme cases, AI hallucinations can lead to completely fabricated and unsubstantiated details being included in a patient's record, which may be used to convict a patient of a crime that they have not committed.
- **Risk of human overconfidence in AI assistance.** Overconfidence in AI-generated transcripts may cause therapists to include transcription errors in a patient's permanent records, thereby failing to mitigate the harms that can arise from inaccurate AI outputs (see above).
- **Risk of human overreliance on AI assistance.** A therapist who becomes over-reliant on AI-generated transcripts may refrain from taking personal notes and consulting their memory of therapy sessions. The practice of memorizing the contents of a therapy session and revisiting those memories after the session is necessary to critically review AI-generated transcripts and identify and correct inaccuracies in such transcripts.
- **Risks to patient privacy.** These AI technologies can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries. The existence of such transcripts also opens up the possibility for them to be subpoenaed or divulged via discovery. (For example, therapists are mandated reporters for certain things divulged in therapy, such as child abuse. A victim could subpoena the transcript as evidence that the therapist had information that they should have reported but did not.)

Automated summarization of therapist-patient interactions

Language models can assist therapists by generating summaries of therapist-patient interactions based on therapy session transcripts, therapist notes, and/or patients' electronic records.

Potential benefits

- **Increased efficiency through automation.** Automated summarization can streamline documentation practices for therapists. Improving the efficiency of writing reports on patients' status and progress can, for example, decrease therapists' workload, potentially improving a therapist's work-life balance. Patients may also benefit from this efficiency gain because an energized and well-rested therapist may be able to offer an improved quality of care.
- **Increased scope or capacity through automation.** The AI-generated summaries can offer a comprehensive overview of therapy sessions, potentially exceeding the level of detail that one would reasonably expect from a therapist's manually taken notes. They may thus improve a therapist's ability to identify key themes and issues in their patient's therapy.
- **Independent data-driven assessment.** The AI-generated summaries can offer a data-driven alternative to a therapist's evaluation of a patient's condition and progress. Therapists can use AI-generated summaries to review and potentially refine their assessments critically.

Risks

- **Risk of inaccurate AI outputs.** AI-generated summaries can include hallucinations, misinterpretations, and other errors that can misrepresent a patient's electronic records. Such errors can arise, for example, when the correct interpretation of a patient's condition would require important context like their facial expression, tone of voice, and gestures that they made during their interactions with their therapist. That information is not included in the AI's inputs. An AI-driven summarization tool may emphasize long and detailed record entries more than short and concise entries (or vice versa). Such biases can lead to a skewed AI-generated patient electronic record summary. They are particularly prone to arise if patient records
 - include entries from multiple therapists and/or other healthcare providers with different documentation styles
 - include some entries that a therapist generated with AI assistance and some entries that the therapist created with assistance from AI tools.

The inaccurate information in AI-generated summaries can harm patients and therapists. For example, they can lead to patients receiving inaccurate diagnoses and inefficient treatment plans from their therapist and/or other healthcare providers. Additionally,

suppose a patient's electronic records are subpoenaed or divulged via discovery. In that case, courts may assume that the information in these records is accurate and allow them as evidence in legal disputes. In extreme cases, AI hallucinations can lead to completely fabricated and unsubstantiated details being included in a patient's record, which may be used as evidence in a legal dispute.

- **Risk of human overconfidence in AI assistance.** Overconfidence in AI-generated summaries may cause therapists to include inaccurate information in a patient's permanent records, thereby failing to mitigate the harms that can arise from inaccurate AI outputs (see above).
- **Risk of human overreliance on AI assistance.** Over-reliance on AI-generated summaries may cause therapists to refrain from documenting their observations in their patients' records. Their observations can be different and sometimes more accurate than the information in an AI-generated summary. This can be the case, for example, when vital context on the patient's background is missing from the information provided to the AI tool. In such a case, the therapist must review and amend the AI-generated summary as necessary.
- **Risks to patient privacy.** These AI technologies can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries. Moreover, AI-generated summaries included in a therapist's notes or a patient's electronic records can be subpoenaed or divulged via discovery. (For example, therapists are mandated reporters for certain things divulged in therapy, such as child abuse. A victim could subpoena the therapist's notes as evidence that the therapist had information that they should have reported but did not.)

Automated summarization of multiple patient data modalities

Language models can be used to consolidate a patient's history from various records, which may include different types of data.

Potential benefits

- **Increased efficiency through automation.** Automated data summarization of patient data can substantially improve a therapist's workflow for generating comprehensive overviews of a patient's condition and progress.
- **Increased scope or capacity through automation.** AI-generated data summaries can facilitate the summarization of data types that a therapist cannot process and summarize on their own. For example, AI-generated summaries could incorporate information from

many transcripts of therapy sessions, therapy notes, user activity data from a patient's use of phone apps that offer mental health support, etc.

- **Independent data-driven assessment.** Summaries considering large and diverse data sources may provide a cohesive overview and holistic, data-driven understanding of a patient's condition. Such a summary may uncover patterns or essential details that a therapist may have missed.

Risks

Compare risks of [Automated summarization of therapist-patient interactions](#).

AI-assisted design of treatment plans

AI tools can provide treatment recommendations and generate treatment plans based on patient data and established clinical guidelines.

Potential benefits

- **Increased efficiency through automation.** The AI-assisted design of treatment plans can substantially improve a therapist's workflow for developing treatment plans for their patients.
- **Increased scope or capacity through automation.** If AI tools are developed to consider an extensive catalog of professional sources in creating treatment recommendations and/or treatment plans, such tools can enable therapists to base their treatment plans on a wide range of evidence-based practices. If these tools are also frequently updated by their developers to incorporate state-of-the-art knowledge, the AI-generated treatment recommendations may include new therapeutic approaches that a therapist has not yet been able to include in their continued education.
- **Increased sensitivity to a user's background and unique needs.** If developed accordingly, an AI tool can adjust its treatment recommendations to consider a patient's unique needs that may arise from their economic, racial, ethnic, cultural, or other backgrounds. For example, the AI tool might recommend treatment approaches that are demonstrated to be most effective with elderly populations if the patient is a senior citizen. Such tailored interventions have the potential to improve the efficacy of therapy and thereby improve patient outcomes.

Risks

- **Risk of inaccurate AI outputs.** AI can generate unhelpful treatment recommendations based on inaccurate data in a patient's record or a misinterpretation of a patient's unique needs. AI hallucinations may cause an AI to assume a patient's record includes

- **Risk of human overconfidence in AI assistance.** Overconfidence in AI-generated treatment recommendations may cause a therapist to opt for an AI-generated treatment plan instead of the treatment plan the therapist would have developed independently. Information that should inform a treatment plan but is not included in a patient's electronic records can then not influence the development of the treatment plan.
- **Risk of human overreliance on AI assistance.** Depending on AI tools to develop treatment plans can diminish a therapist's own critical judgment and expertise. This judgment and expertise is important for critically assessing the accuracy of AI-generated treatment recommendations and to provide treatment recommendations when AI tools are not available or a patient has opted out of the use of such tools in their therapy.
- **Risks to patient privacy.** These AI technologies can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries.

Patient intake via conversational bots

A conversational bot can interact with a patient during intake and collect medical information and other relevant data about the patient through conversation with the patient. It can also assist in the procurement of a patient's informed consent.

Potential benefits

- **Increased efficiency through automation.** Automating some or most tasks related to patient intake via conversational AI can reduce administrative burdens for a therapist or their administrative co-workers or staff. Conversational bots can take on time-consuming tasks like gathering basic patient information (contact details, demographics, insurance), scheduling appointments, and obtaining consent.
- **Engaging user experiences.** By incorporating interactive elements, such as personalized questions and multimedia content, the AI can make the intake process interactive and enjoyable for patients. The augmented intake procedure can create a comprehensive and detailed patient history without causing the anxiety, stress, or frustration that some patients may feel when being asked to fill out a large number of long questionnaires and forms.
- **Increased sensitivity to a user's background and unique needs.** A conversational bot may communicate with a patient in a way that accommodates the patient's background and unique needs. For example, many AIs are multilingual and can converse with users in a language of the user's choice. They may also accommodate varying literacy levels, cultural backgrounds, and preferred communication methods (e.g., voice, text). Aspects of the communication format, such as volume of voice or font

size of text, can be adjusted to accommodate a patient's preferences and/or hearing or visual impairments.

- **Increased availability of support.** An AI tool for patient intake can operate 24/7. Its around-the-clock availability allows patients to complete their intake at any time before their first appointment at their preferred pace.

Risks

- **Risk of inaccurate AI outputs.** AI-generated intake documentation may include inaccuracies that arise from a conversational bot missing or misunderstanding information provided by a patient. Additionally, a generative AI may hallucinate new, inaccurate information and include it in the intake documentation without the patient's knowledge.
- **Technological barriers for users.** Some patients may face technological barriers, such as limited access to technology, lack of digital literacy, or difficulty using technology due to age, disability, or other factors. These patients may find completing their intake with a conversational bot difficult or impossible. They may have a negative intake experience and may even decide not to seek mental health support because of this negative experience.
- **Lack of sensitivity to a patient's background and unique needs.** A conversational bot may fail to detect or misinterpret signs that indicate a patient's unique needs and may thus not appropriately accommodate such needs. Patients may have a negative intake experience and may even decide not to seek mental health support because of this negative experience.
- **Undesired consequences from out-of-scope use.** Patients may try to talk to an intake bot about a topic that is unrelated to their intake. A conversational bot without appropriate guardrails may engage in a conversation that is outside the scope of its intended use. These out-of-scope conversations can lead the intake bot to produce inappropriate comments or harmful advice. Patients may ask the bot for real-time therapeutic support in emergencies, which a bot designed for patient intake would not be equipped to offer. Especially when an AI intake bot operates outside of standard office hours, it may be hard for a healthcare organization's therapist or staff to detect such an emergency situation and intervene appropriately. Some patients may also try to deliberately deceive an intake bot to gain access to sensitive information about a therapist or their practice. Without appropriate guardrails, an intake bot can fall for such phishing attacks.
- **Risk of human overconfidence on AI assistance.** Hallucinations by generative AI, tech-induced frustration, and other factors may lead to intake documentation being inaccurate or incomplete. A therapist who does not confirm with patients that the

information collected by an intake bot is correct, risks basing their analyses and their patients' treatment on wrong information.

- **Risks to patient privacy.** The use of conversational bots can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries.

Bot-assisted patient homework

Assigning homework that involves patients interacting with conversational bots can reinforce therapeutic techniques between sessions. For example, such bots can guide patients through CBT exercises, meditation, or day planning.

Potential benefits

- **Increased capacity through automation.** Using conversational bots for therapy homework allows therapists to include guided exercises in treatment plans without allocating a substantial amount of time within therapy sessions to these exercises. As a result, the therapist can use the time within therapy sessions to focus on other aspects of their patient's treatment.
- **Personalized service.** Conversational bots can adjust their responses to a user's personal preferences, which they deduce from user data collected over time. Customized user experiences can make the interactions between a patient and the conversational bot more engaging for the patient. Therapy homework with a conversational bot's assistance may thus be more effective than therapy homework without the bot.
- **Engaging user experiences.** In addition to offering a personalized user experience, some conversational bots can incorporate interactive elements, such as multimedia content and gamified progress updates. These elements can make the therapy homework fun for some patients.
- **Increased sensitivity to a user's background and unique needs.** A conversational bot may communicate with a patient in a way that accommodates the patient's background and unique needs. For example, many AIs are multilingual and can converse with users in a language of the user's choice. They may also accommodate varying literacy levels, cultural backgrounds, and preferred communication methods (e.g., voice, text). Aspects of the communication format, such as volume of voice or font size of text, can be adjusted to accommodate a patient's preferences and/or hearing or visual impairments.

- **Increased availability of support.** A conversational bot can operate 24/7, thereby allowing the patient to complete their guided homework at their preferred time. Specifically, therapy homework can involve practicing coping strategies with the bot's guidance in real-time scenarios.

Risks

- **Risk of inaccurate AI outputs.** A conversational bot's responses can be inaccurate if it misses or misunderstands information provided by a patient. Additionally, a conversational bot may sometimes provide incorrect, inappropriate, or non-evidence-based responses due to generative AI hallucinations or other limitations of the bot's development. These inaccuracies can lead to discrepancies between the conversational bot's guidance and the therapist's advice, potentially causing confusion and/or hindering a patient's progress.
- **Technological barriers for users.** Some patients may face technological barriers, such as limited access to technology, lack of digital literacy, or difficulty using technology due to age, disability, or other factors. These patients may find it difficult or impossible to complete their therapy homework with a conversational bot. Their negative experiences with their homework assignments can diminish the effectiveness of their homework assignments and their therapy.
- **Lack of sensitivity to a patient's background and unique needs.** A conversational bot may fail to detect or misinterpret signs that indicate a patient's unique needs and may thus not appropriately accommodate such needs. Patients may have negative experiences when engaging with their bot-assisted homework assignments. Their negative experiences with their homework assignments can diminish the effectiveness of their homework assignments and their therapy.
- **Undesired consequences from out-of-scope use.** Patients may try to talk to a conversational bot about a topic unrelated to their homework assignment. A conversational bot without appropriate guardrails may engage in a conversation that is outside the scope of its intended use. These out-of-scope conversations can lead a conversational bot to produce inappropriate comments or harmful advice.
- **Risk of human overconfidence on AI assistance.** Tech-induced frustration, a conversational bot's lack of sensitivity to a patient's background and unique needs, and other factors may lead to a bad user experience for patients. When a therapist does not check with patients that their bot-assisted therapy homework is impacting the patient positively, patients may have frustrating experiences in doing their homework. These negative experiences can diminish the effectiveness of the homework assignment and their therapy.
- **Risks to patient privacy.** The use of conversational bots can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of

the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries.

Bot-administered talk therapy

Several providers of AI tools offer conversational bots that aim to mimic a therapist in a virtual talk therapy session. Initial research results motivate cautious optimism about these bots' efficacy (Aggarwal et al., 2023; Schillings et al., 2024). Some of these conversational bots are advertised directly to patients as consumers. Others are advertised to health organizations and/or clinicians who may recommend or prescribe the use of these bots to their patients. This document only considers using such bots as prescribed or recommended by their therapist.

Potential benefits

- **Increased capacity through automation.** Using conversational bots for additional talk therapy sessions allows therapists to design treatment plans that include talk therapy for the patient at a higher frequency than their therapist can provide. Patients can benefit from frequent mental health support that can be flexibly adjusted to accommodate a patient's schedule and preferences. As a result, the patient may be able to manage some challenges of their day-to-day life without the assistance of their therapist. The therapist can then use the time within therapy sessions to focus on other, possibly more complex, aspects of their patient's treatment.
- **Personalized service.** Conversational bots can adjust their responses to a user's personal preferences, which they deduce from user data collected over time. Personalized user experiences can make the interactions between a patient and the conversational bot more engaging for the patient. A conversational bot may be able to converse with a patient in the patient's preferred language and/or conversation style. For patients seeking to improve their mental health through long and deep discussions of a special-interest topic, a conversational bot may be an apt conversation partner.
- **Increased sensitivity to a user's background and unique needs.** The conversational bot may communicate with a patient in a way that accommodates the patient's background and unique needs. For example, many AIs are multilingual and can converse with users in a language of the user's choice. They may also accommodate varying literacy levels, cultural backgrounds, and preferred communication methods (e.g., voice, text). Aspects of the communication format, such as volume of voice or font size of text, can be adjusted to accommodate a patient's preferences and/or hearing or visual impairments.
- **Increased availability of support.** A conversational bot can operate 24/7, thereby allowing the patient to receive support at their preferred time or in particularly stressful situations.

Risks

- **Risk of inaccurate AI outputs.** A conversational bot's responses can be inaccurate if it misses or misunderstands information provided by a patient. Additionally, a conversational bot may sometimes provide incorrect, inappropriate, or non-evidence-based responses due to generative AI hallucinations or other limitations of the bot's development. These inaccuracies can lead to discrepancies between the conversational bot's guidance and the therapist's advice, thereby potentially causing confusion and/or hindering a patient's progress.
- **Technological barriers for users.** Some patients may face technological barriers, such as limited access to technology, lack of digital literacy, or difficulty using technology due to age, disability, or other factors. These patients may find it difficult or impossible to have helpful conversations with a conversational bot. Their negative experiences with the conversational bot can diminish the overall effectiveness of their therapy.
- **Lack of sensitivity to a patient's background and unique needs.** The conversational bot may fail to detect or misinterpret signs that indicate a patient's unique needs and may thus not appropriately accommodate such needs. Patients may have negative experiences when engaging with their bot-assisted homework assignments. Their negative experiences with their homework assignments can diminish the effectiveness of their homework assignments and their therapy.
- **Undesired consequences from out-of-scope use.** Patients may try to talk to a conversational bot about topics that are either unrelated to their mental well-being or address their mental well-being in harmful ways (e.g., a patient with an eating disorder asking for advice on how to lose weight quickly). A conversational bot without appropriate guardrails may engage in a conversation that is outside the scope of its intended use. These out-of-scope conversations can lead a conversational bot to produce inappropriate comments or harmful advice.
- **Risk of human overconfidence in AI assistance.** Tech-induced frustration, a conversational bot's lack of sensitivity to a patient's background and unique needs, and other factors may lead to a bad patient user experience when a therapist who is overconfident in the capabilities of AI-administered talk therapy may decide not to check with patients that the conversations with the bot are impacting the patient positively. In such a case, negative experiences with the conversational bot can detrimentally affect the patient's progress and well-being. Additionally, a patient can grow overconfident in the capabilities of the conversational bot. This overconfidence can affect the patient negatively when, for example, a patient trusts the conversational bot's advice over their judgment and the advice from their therapist.
- **Risk of human overreliance on AI assistance.** Frequent or excessive use of a conversational bot or a feeling of insecurity or helplessness in situations in which the bot

is unavailable can be signs of a patient becoming overreliant on a conversational bot (Hu et al., 2023). Such an overreliance can

- cause a patient to experience a diminished sense of agency and autonomy,
 - lower the patient's self-esteem and sense of security,
 - develop unhealthy coping mechanisms for navigating stressful situations and other mental-health-related challenges,
 - impair the patient's relationships with other people,
 - and ultimately have detrimental effects on the patient's mental health.
- **Risks to patient privacy.** The use of conversational bots can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the therapist or one of their business associates. Additionally, AIs trained on a patient's data may reveal that data in response to adversarial queries.

Bot-assisted learning experiences

Conversational bots can guide or mentor students and trainees of mental health therapy to enhance their learning experience. For example, a conversational bot can be a source of relevant information that can be accessed via conversation. A conversational bot may guide a student through an exercise, help students test their skills through simulating test scenarios, evaluate a student's responses, and assist in developing personal learning plans.

Potential benefits

- **Increased efficiency through automation.** A conversational bot that serves as a source of information for studying mental health therapy can quickly provide students with the most relevant information for their queries. Such bots may thus make information retrieval for students more efficient than manually searching a textbook and more accurate than an internet search.
- **Personalized service.** Conversational bots can adjust their responses to a user's personal preferences, which they deduce from user data collected over time. Personalized user experiences can make the interactions between a student and the conversational bot more engaging for the student. The student may feel more motivated to learn and have better learning outcomes than without the conversational bot's assistance.
- **Engaging user experiences.** In addition to offering a personalized user experience, some conversational bots can incorporate interactive elements, such as multimedia content and gamified progress updates. These elements can make the learning more fun for some students.

- **Increased sensitivity to a user's background and unique needs.** A conversational bot may be able to communicate with a student in a way that accommodates the student's background and unique needs. For example, many AIs are multilingual and can converse with users in a language of the user's choice. They may also accommodate varying literacy levels, cultural backgrounds, and preferred communication methods (e.g., voice, text). Aspects of the communication format, such as volume of voice or font size of text, can be adjusted to accommodate a student's preferences and/or hearing or visual impairments.
- **Increased availability of support.** A conversational bot can operate 24/7, allowing students to learn at their preferred time.

Risks

- **Risk of inaccurate AI outputs.** A conversational bot's responses can be inaccurate if the bot misses or misunderstands information a student provides. Additionally, a conversational bot may sometimes provide incorrect, inappropriate, or non-evidence-based responses due to generative AI hallucinations or other limitations of the bot's development. These inaccuracies can lead to discrepancies between the conversational bot's guidance and the teacher's advice, potentially causing confusion and/or hindering the student's learning progress.
- **Technological barriers for users.** Some students may face technological barriers, such as limited access to technology, lack of digital literacy, or difficulty using technology due to age, disability, or other factors. These students may find using a conversational bot to improve their learning experience and outcomes difficult or impossible. When a teacher suggests that students use a conversational bot to augment their learning experience, but the student has negative experiences with the bot, the continued interactions can diminish their learning outcomes.
- **Lack of sensitivity to a patient's background and unique needs.** A conversational bot may fail to detect or misinterpret signs that indicate a student's unique needs and may thus not appropriately accommodate such needs. Negative experiences with such a bot due to inappropriate responses can distract and sometimes harm a student's learning outcomes.
- **Undesired consequences from out-of-scope use.** Students may try to talk to a conversational bot about a topic unrelated to their studies. A conversational bot without appropriate guardrails may engage in a conversation that is outside the scope of its intended use. These out-of-scope conversations can lead a bot to produce inappropriate comments or harmful responses.
- **Risk of human overconfidence in AI assistance.** Tech-induced frustration, a conversational bot's lack of sensitivity to a student's background and unique needs, and other factors may lead to a bad user experience for students. When a teacher does not

check with students that their bot-assisted learning activities positively impact the student and/or their learning process, the teacher may remain unaware of a student's negative experiences, which can be detrimental to learning outcomes.

- **Risks to patient privacy.** The use of conversational bots can lead to breaches of confidentiality through accidental or purposeful mishandling or unauthorized access of the collected data by the teacher, supervisor, or one of their business associates. Additionally, AIs trained on a student's data may reveal that data in response to adversarial queries.

6. Best Practices for the Use of Artificial Intelligence Technologies by Mental Health Therapists

6.1 Best Practices for the General Use of AI Technologies

Informed consent

- A mental health therapist who uses or intends to use AI technologies that involve
 - recording or transcribing therapy sessions or
 - interactions between a patient and a conversational bot beyond patient-intake procedures

in their practice should ensure that patients are aware of and consent to use these technologies.

- The mental health therapist should inform patients about (1) the potential benefits and risks of the use of these technologies, (2) what data is collected, stored, or shared with the provider of the AI tool through the use of these technologies, and (3) the risks associated with this data collection and dissemination. For AI technologies involving interactions between a patient and a conversational bot beyond patient intake, the mental health therapist should also inform patients about the form of monitoring that the mental health therapist uses in their ongoing assessment of the effectiveness of these technologies.
- The mental health therapist should ensure that (1) this information is presented to each patient in a manner that is accessible to the patient, (2) the patient understands the presented information, and (3) the patient understands that they have the option to refuse to give consent.
- If a patient asks to review the uses of AI technology to which they consented, the mental health therapist should provide that information.
- A patient may revoke their consent at any time. Once notified about their patient's withdrawn consent, a mental health therapist should act promptly to ensure that all uses of AI technologies that the patient does not consent to anymore cease. The therapist should inform the patient that (1) the therapist has stopped the use of these AI technologies, (2) data shared with providers of AI tools before the revocation of consent may still be stored and used by those providers in accordance with the data-sharing agreement between the therapist and the AI providers.
- Patients must be allowed to refuse to consent to any or all uses of AI technologies in the service they receive from the mental health therapist. If a patient refuses the use of any or all AI technologies, the mental health therapist should make a reasonable effort to

provide the service using only the AI technologies to which the patient has consented. If the mental health therapist can't provide the service to the patient under these circumstances, the mental health therapist should make a reasonable effort to connect the patient with a mental health therapist who can.

Disclosure

- A mental health therapist who uses or intends to use AI technologies that are not covered under *Informed Consent* and that
 - include interactions between a patient and a conversational bot during patient-intake procedures,
 - generate text to be included in patients' therapy notes, progress notes, or other parts of their electronic health records,
 - generate individual treatment recommendations or individual treatment plans for patients

should disclose the use or intended use of these AI technologies to their patients.

- The intake bot can disclose the use of a conversational bot during patient-intake procedures at the beginning of the interaction. The disclosure should include instructions on how the patient can contact a human provider to conduct their intake if they object to conducting their intake with the conversational bot.
- The mental health therapist should inform patients about (1) the potential benefits and risks of the use of these technologies, including the risk of inadequate responses to patient backgrounds and unique needs (see *Risk of inadequate response to patient backgrounds and unique needs*), (2) what data is collected, stored, or shared with third parties through the use of these technologies, and (3) the risks associated with this data collection and dissemination.
- The mental health therapist should ensure that this information is presented to each patient in a manner that is accessible to the patient.
- If a patient objects to using any or all of the disclosed AI technologies, the mental health therapist may offer to provide their therapeutic services without using them. If the mental health therapist cannot or does not offer this option to the patient, the mental health therapist should make a reasonable effort to connect the patient with a mental health therapist who will offer their therapeutic services in a way that aligns with the patient's stated preferences on the use of AI technologies in their therapy.

Data privacy and data safety

- Using AI technologies in behavioral health care introduces risks of sharing sensitive patient data with third parties, intentionally or unintentionally. Before using any AI tool, a mental health therapist should ascertain what data the provider collects.
- If the information that the AI tool collects includes patients' personally identifiable information (PII)—which may include but is not limited to personal health information (PHI)—the therapist should further ascertain
 - how long collected PII is stored,
 - if, how, and for what purpose PII is shared with or sold to other parties,
 - and what precautions the provider of the AI tool has taken to protect collected data from theft and accidental leakage.
- If the provider's practices for data collection, storage, sharing, and safety meet appropriate safety standards (e.g., as stated in the code of ethics and administrative guidance of the licensed mental-health therapist's profession; see Utah Administrative Code R156-60c), the therapist or the healthcare organization that employs the therapist should establish a Business Associates Agreement (BAA) with the provider of the AI tool to verify the provider's adherence to these standards. At the minimum, the BAA should clarify
 - concrete requirements on the provider's data handling procedures to meet appropriate safety standards
 - notification procedures and liability in the event of a data breach.
- Therapists should refrain from using an AI tool if the provider of the AI tool
 - follows practices for data collection, storage, sharing, and safety that do not meet appropriate safety standards or
 - refuses to sign a BAA verifying the provider's adherence to these standards.
- An AI tool's "HIPAA certification" does not alleviate the need for a BAA between its provider and the therapist or the healthcare organization that employs the therapist.

Mental health therapist competence with AI technology

- Mental health therapists should maintain a high level of competence with the AI technologies they employ. This practice involves continuous education and training to understand these AI technologies' capabilities, limitations, and proper use. A thorough understanding of an AI tool's capabilities and limitations includes knowledge about how

frequently and under what circumstances one should expect the AI tool to produce inaccurate or undesirable outputs.

- A thorough understanding of the capabilities of an AI tool used in diagnosing or treating a mental health therapist's patients includes knowledge of its scope, i.e., the demographic, cultural, and other population groups for which the AI tool is expected to work as intended. Before incorporating an AI tool in the diagnosis or treatment of their patients, therapists should inform themselves about the scope of the AI tool. When this information is not publicly available, the therapist should inquire directly with the provider of the AI tool if the AI tool has been developed considering the needs associated with the population characteristics of the therapist's patients. For example, AI tools used for diagnosing and treating patients in Utah should generally be developed considering the needs, expectations, and everyday environment of US populations. Other relevant population characteristics to consider may include
 - age ranges (e.g., an AI tool developed using only training data from individuals with ages between 21 and 50 may be unsuitable for the diagnosis or treatment of minors and seniors),
 - socio-economic statuses (e.g., an AI tool developed using only training data from predominantly medium-income households may be unsuitable for the therapist's work with low-income individuals or families),
 - cultural, racial, and religious backgrounds (e.g., an AI tool developed using only training data from people for whom religion plays a minor or no role in their lives may be less accurate or effective for the diagnosis or treatment of patients with a strong religious identity),
 - mental health conditions and their severity (e.g., an AI tool developed to augment the treatment of patients with mild depression may be unsuitable for treating patients with severe depression or mild depression with severe comorbidities), and
 - mental or physical disabilities (e.g., a conversational bot that has no consideration for the accommodations that some physical disabilities require may be unhelpful to patients with mobility disabilities).
- Mental health therapists should consider that
 - more information about the performance, efficacy, and safety of an AI tool (including accuracy and incidence rates for false-positive and false-negative outputs of tools using predictive AI technologies) may become available with time,
 - this information can come from various sources, including (but not limited to) internal or external audits of AI tools or their providers, evaluations conducted by professional societies and healthcare organizations, clinical studies carried out by AI tool providers, universities, and other institutions,

- capabilities, limitations, and proper use of these AI technologies can be substantially affected by new information about the efficacy and safety of an AI tool,
- capabilities, limitations, and guidelines for proper use can also be considerably affected by software updates to AI technology and interacting software and hardware products.
- Mental health therapists should make a reasonable effort to stay informed about the efficacy, safety, capability, limitations, and proper use of AI tools. When new information indicates that an AI tool is no longer efficacious or safe for the diagnosis or treatment of all or some patients, therapists should change their use of the AI tool to prioritize the well-being and safety of their patients.
- Mental health therapists may find it helpful to regularly consult information technology specialists to ensure their competence on these issues adjusts to relevant technology updates.

Patient safety and competence with AI technology

- Before incorporating patient-facing AI tools in a patient's treatment, a mental health therapist should weigh potential benefits against potential harms that the patient may experience from interacting with AI tools.
- In assessing potential harms, therapists should consider
 - the patient's digital literacy,
 - the patient's propensity for AI over immersion, overreliance, or addiction, and
 - physical, economical, logistic, and time constraints that may make it difficult or impossible for the patient to interact with the AI tool as intended.
- The therapist may talk to them about their technology experiences to assess a patient's digital literacy. It may be useful to demonstrate the patient-facing AI tool to the patient and ask them to rate their confidence in using the technology appropriately.
- To assess a patient's potential for harm via AI over immersion, overreliance, or addiction, the therapist may consult the patient's record and ask the patient about prior experiences with over-immersion in, overreliance on, and addiction to technology. Examples may include online social media, video games, smartphones, and other devices or applications that may or may not use AI technologies.
- When a therapist decides to incorporate an AI tool in treating a patient with a history of technology-related behavioral issues—because the likely benefits outweigh the risks—the therapist should adjust their procedures accordingly. For such a patient, a

customized approach to continuously monitoring patient-AI interactions may be implemented to address the increased risk of harm (see [Continuous monitoring of patient-AI interactions](#)).

- A patient's physical constraints may become relevant if patient-facing AI tools require a certain mode of interaction. For example, an AI tool that uses a graphical user interface incompatible with screen readers may be unsuitable for some patients with visual impairments, and an AI tool that interacts with patients via speech may be unsuitable for patients with auditory or speech impairments.

Unique patient needs

- AI technologies trained on data sourced predominantly from one demographic group are likely to be less well-suited for individuals with differing unique needs, which may depend on factors related to different cultures, ways of learning, abilities, educational levels, and economic circumstances. When using generative AI technologies, mental health therapists should consider patients' unique needs.
- The considerations should include
 - the patient's social, economic, racial, ethnic, cultural, or other backgrounds, and
 - the patient's mental health conditions, physical health conditions, and temporary or permanent disabilities.
- If a patient's background or condition falls outside the scope of an AI tool (see [Mental health therapist competence with AI technology](#)), therapists should generally not incorporate the AI tool in the patient's diagnosis or treatment.
- These considerations apply to AI technologies used in patient-AI interactions (e.g., an assistant for therapy homework) and AI technologies used in mental health therapist-AI interactions (e.g., a diagnostic assistant) because both use cases can cause harm to patients whose unique needs differ from the source of training data.
- If the mental health therapist assesses that using AI technology will benefit a patient despite a discrepancy between the AI technology's trained capacity and the patient's unique needs, the therapist should inform the patient about the potential benefits and risks arising from this discrepancy. If the patient consents to the use of the AI technology after receiving this information, the mental health therapist may proceed cautiously with the use of the AI technology.

Continuous monitoring and reassessment

- To the best of their abilities, a mental health therapist should ensure that interactions facilitated by AI are safe and do not cause harm.

- A mental health therapist should continuously monitor and critically challenge AI outputs for inaccuracies and biases and intervene promptly if the AI produces incorrect, incomplete, or inappropriate content or recommendations.
- The frequency with which a therapist monitors an AI tool's outputs should reflect
 - the therapist's familiarity with the AI tool,
 - the therapist's knowledge of evidence on the accuracy, failure rates, and safety of the AI tool,
 - considerations related to the background and unique needs of individual patients.
- Considerations related to the background and unique needs of individual patients should include that AI tools can be less accurate and more prone to fail in cases of patients with rare backgrounds, rare unique needs, or rare mental-health conditions due to shortcomings of their training (see [Failure Modes for Modern Artificial Intelligence](#)).

6.2 Best Practices for Therapists' Interactions with Generative AI

Authorship transparency

- When AI technologies contribute to documentation or communication, mental health therapists should disclose their use of these technologies to maintain transparency. Such disclosure can, for example, be included as part of a signature (e.g., similar to an automated email signature) or in the form of an authorship note (e.g., see [Authorship note and Acknowledgments](#)) at the end of a text or in the form of a disclosure (e.g., the statement "The following text was generated using with the AI tool [name of AI tool]" at the beginning of the text.
- Acknowledging AI involvement helps preserve integrity and allows for appropriate attribution of sources in professional records. It also facilitates an accurate assessment of the veracity of information in patient records subpoenaed or divulged in discovery. (For example, it may be essential to know that AI-generated text to understand that words or sentences that seem out of context or very unusual could result from AI hallucinations.)

Review of AI-generated text

- Therapists should review all AI-generated text about a patient, the patient's condition, or the patient's progress before they
 - include the text in a patient's electronic health record,

- share the text with another healthcare provider who provides healthcare to the patient or
- use the text to determine a diagnosis or treatment decision.

This ensures accuracy, relevance, and appropriateness, allowing mental health therapists to correct any errors or omissions the AI may have made.

Reasonable doubt about AI-generated diagnosis and treatment proposals

- Mental health therapists should critically evaluate any diagnoses or treatment suggestions AI technologies offer. Recognizing that AI is a supplement—not a replacement—for professional judgment helps prevent overreliance on technology and supports evidence-based practice.

6.3 Best Practices for Patients' Interactions with Generative AI

Patients' relationship with AI technology

- Before suggesting the use of AI technology to a patient, a mental health therapist should consider the patient's views about technology and how they use it, including strengths, needs, risks, and challenges. These considerations are important for assessing whether the potential benefits outweigh the risks of using this AI technology in the mental health therapist's service to the patient.

Primary commitment to patients

- Patients' well-being should always precede the use of AI technology. Mental health therapists should ensure that (1) AI technologies augment their services for the benefit of their patients and (2) the potential benefits outweigh the risks for each patient in the context of their unique needs.

Service interruptions and malfunctions

- Mental health therapists should prepare for potential AI service disruptions by having backup plans in place. Communicating these contingencies to patients reassures them that their care will continue smoothly even if technical issues arise.

Conduct of generative AI

- Conversational bots should interact with patients ethically and professionally. They should provide accurate information and appropriate responses.
- The more the capabilities or appearance of AI technology in patient-AI interactions make them similar to interactions with a licensed mental health therapist, the more the technology can be expected to comply with the ethical guidelines and standards of the mental health therapist's profession (e.g., confidentiality, competence, beneficence).
- Before incorporating a conversational bot into their therapeutic practice or recommending its use to patients, a mental health therapist should examine and verify that the bot displays awareness and alignment with ethical and professional best practices commensurate with its capabilities and appearance.

Responsibility in emergency situations

- A generative AI technology that interacts with a patient similarly to how a mental health therapist would interact with the patient can be expected to show a similar level of responsibility in an emergency situation. Mental health therapists should inform themselves about AI technology's capability to act responsibly in emergencies and escalate situations if needed.
- If an AI technology does not have the capabilities to act responsibly in an emergency, the mental health therapist should not proceed with the use of the AI technology or proceed only after having taken appropriate precautions. These precautions include (1) considering the risk of their patient seeking help from the AI technology in an emergency based on their patient's vulnerability and relationship to technology, (2) informing the patient about the limits of the AI technology's capability in emergencies, (3) informing the patient about how to seek appropriate help in emergencies.

Continuous monitoring of patient-AI interactions

- The mental health therapist should monitor interactions between patients and conversational bots that occur as part of the patient's treatment to continuously assess the effectiveness of the use of AI technologies in the service provided to the patient.
- The mental health therapist should select a form, frequency, and practice of monitoring patient-AI interactions that adequately reflects the mental health therapist's experience with the AI technology, the AI technology's robustness and safety, and the patient's unique needs. (In medium or higher risk scenarios, this may involve a timely review of

entire chat logs. In low-risk scenarios, this may affect the mental health therapist, who initializes regular discussions about the patient's experience with AI technology.)

- The mental health therapist should explain the frequency and practice of monitoring patient-AI interactions to the patient and receive the patient's informed consent prior to using the AI technology and the mental health therapist's proposed monitoring scheme.
- Changes in technology or a patient's needs may justify changes to the monitoring scheme. If a mental health therapist decides that a change is appropriate, they should discuss the proposed change with the patient and only proceed with the change after receiving the patient's informed consent.

Mandated reporting of patient-AI interactions

- A mental health therapist's mandated reporting duties include patient-AI interactions to the extent that the therapist monitors them. The therapist should inform the patient about the therapist's reporting duties in this context.

6.4 Best Practices for Supervisors' Interactions with Generative AI

Review of AI-generated text

- Supervisors should thoroughly review all AI-generated text before using it to assess a student's performance or the well-being of a student's patients.

Reasonable doubt about AI-generated student assessments

- Mental health therapists should critically evaluate any student assessment offered by AI technologies. Recognizing that AI is a supplement—not a replacement—for professional judgment helps prevent overreliance on technology and ensures that students receive fair, accurate, and helpful training and feedback.

6.5 Best Practices for Students' Interactions with Generative AI

Student's relationship with AI technology

- Before suggesting the use of AI technology to a student, a supervisor should consider the student's views about technology and the ways in which they use technology,

including strengths, needs, risks, and challenges. These considerations are important for assessing whether the potential benefits outweigh the risks of using this AI technology in the training and supervision of students and trainees. Risks for the student and risks for the student's patients should be considered.

Primary commitment to students

- The well-being of students and the positive outcome of their training should always take precedence over the use of AI technology. Supervisors should ensure that (1) AI technologies augment the educational experience of their students and (2) the potential benefits for the student outweigh the risks for each student and their patients.

Service interruptions and malfunctions

- When using AI technology in the education and training of mental health therapists involves a risk of overreliance or dependence on AI technology, supervisors should discuss these risks with their students. They should also prepare for potential AI service disruptions by having backup plans to mitigate harms that may befall students or students' patients. Communicating these contingencies to students reassures them that their care will continue smoothly even if technical issues arise.

Continuous monitoring of student-AI interactions

- Interactions between students and conversational bots that occur as part of the student's training and are overseen by a supervisor should be monitored by the supervisor to continuously assess the effectiveness of using the employed AI technologies in the education and supervision provided to the student.
- Supervisors should select a form, frequency, and practice of monitoring patient-AI interactions that adequately reflects the supervisor's experience with the AI technology, the student's experience with the AI technology, and the AI technology's robustness and safety. When the supervisor or the student has no or little experience with AI technology, monitoring may involve a timely review of entire chat logs. When the supervisor and the student have substantial experience with AI technology, regular discussions may be held between the supervisor and the student about the student's experience with AI technology.
- The supervisor should explain the frequency and practice of monitoring student-AI interactions with the student and obtain the student's informed consent prior to using the AI technology and the supervisor's proposed monitoring scheme.

7. Authorship Note and Acknowledgments

This document was prepared by Alice C. Schwarze and Zachary M. Boyd for the Utah Office of Artificial Intelligence Policy (OAIP) and the Utah Division of Professional Licensing (DOPL).

Google Gemini and Anthropic's Claude AI were used in the preparation of this document. All AI-generated content was reviewed and/or revised by the authors.

The authors would like to thank Jared Ferguson, Aaron J. Fisher, Cole Hooley, Zac E. Imel, Jana Johanson, Nina de Lacy, Tamara Moores Todd, Frederic Reamer, Mark Steinagel, Sarah Stroup, Greg Whisenant, Brady Young, the American Psychological Association, and the Utah Behavioral Health Board for their helpful suggestions and comments.

8. References

- Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence–based chatbots for promoting health behavioral changes: systematic review. *Journal of medical Internet research*, 25, e40789. URL: <https://doi.org/10.2196/40789>
- American Association for Marriage and Family Therapy. (2015). Code of Ethics. URL: https://www.aamft.org/AAMFT/Legal_Ethics/Code_of_Ethics.aspx
- American Mental Health Counselors Association. (2020). Code of Ethics. URL: <https://www.amhca.org/events/publications/ethics>
- American Mental Health Counselors Association. (2023). Code of Ethics Addendum: Addressing Artificial Intelligence. URL: <https://www.amhca.org/viewdocument/amhca-code-of-ethics-addendum-addr#:~:text=AMHCA%20Code%20of%20Ethics%20Addendum%3A%20Addressing%20Artificial%20Intelligence%3A%202023&text=CMHCs%20clearly%20disclose%20to%20clients,person%2Dto%2Dperson%20contact>
- American Nurses Association. (2014). *Psychiatric-Mental Health Nursing: Scope and Standards of Practice* (2nd ed.).
- American Psychiatric Association. (2013). Principles of Medical Ethics With Annotations Especially Applicable to Psychiatry. URL: <https://www.psychiatry.org/psychiatrists/practice/ethics>
- American Psychological Association. (2016). Ethical Principles of Psychologists and Code of Conduct. URL: <https://www.apa.org/ethics/code>
- Association of Social Work Boards. (2014). Model Regulatory Standards for Technology and Social Work Practice. URL: <https://www.aswb.org/wp-content/uploads/2015/03/ASWB-Model-Regulatory-Standards-for-Technology-and-Social-Work-Practice.pdf>
- Ettman, C. K., & Galea, S. (2023). The Potential Influence of AI on Population Mental Health. *Journal of Medicine, Surgery, and Public Health*, 10, Article e49936. URL: <https://doi.org/10.2196/49936>
- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., and Agarwal, S. (2025). How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Controlled Study. arXiv:2503.17473. URL: <https://arxiv.org/abs/2503.17473>
- Hu, B., Mao, Y., & Kim, K. J. (2023). How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. *Computers in Human Behavior*, 145, Article 107760. URL: <https://doi.org/10.1016/j.chb.2023.107760>
- Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H. C., Paulus, M. P., Krystal, J. H. & Jeste, D. V. (2021). Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9), 856-864. URL: <https://doi.org/10.1016/j.bpsc.2021.02.001>
- Narayanan, A., & Kapoor, S. (2024). *AI Snake Oil. What Artificial Intelligence can do, what it can't, and how to tell the difference*. Princeton University Press.

Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., & Lee, K. (2023). Scalable Extraction of Training Data from (Production) Language Models. arXiv:2311.17035. URL: <https://arxiv.org/abs/2311.17035>

National Association for Alcoholism and Drug Abuse Counselors / National Certification Commission for Addiction Professionals. (2021). Code of Ethics. URL: https://www.naadac.org/assets/2416/naadac_code_of_ethics_06012025.pdf

National Association of Social Workers. (2021). Code of Ethics. URL: <https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English>

National Board for Certified Counselors. (2023). Code of Ethics. URL: <https://nbcc.org/assets/ethics/nbcccodeofethics.pdf>

National Association of Social Workers, Association of Social Work Boards, Council on Social Work Education, & Clinical Social Work Association. (2017). Standards for Technology in Social Work Practice. URL: <https://www.socialworkers.org/Practice/NASW-Practice-Standards-Guidelines/Standards-for-Technology-in-Social-Work-Practice>

Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health*, 3, Article 100089. URL: <https://doi.org/10.1016/j.glmedi.2024.100099>

Schillings, C., Meißner, E., Erb, B., Bendig, E., Schultchen, D., & Pollatos, O. (2024). Effects of a chatbot-based intervention on stress and health-related parameters in a stressed sample: randomized controlled trial. *JMIR Mental Health*, 11(1), e50454. URL: <https://doi.org/10.2196/50454>

Schwarzschild, A., Feng, Z., Maini, P., Lipton, Z., & Kolter, J. Z. (2025). Rethinking LLM memorization through the lens of adversarial compression. *Advances in Neural Information Processing Systems*, 37, 56244-56267. URL: https://proceedings.neurips.cc/paper_files/paper/2024/hash/66453d578afae006252d2ea090e151c9-Abstract-Conference.html

Thakkar, A., Gupta, A., & de Sousa, A. (2024). Artificial intelligence in positive mental health: A narrative review. *Frontiers in Digital Health*, 6, Article 1280235. URL: <https://doi.org/10.3389/fdgth.2024.1280235>

Utah State Code. URL: <https://le.utah.gov/xcode/code.html>

Utah Administrative Code R156-60c. URL: <https://adminrules.utah.gov/public/rule/R156-60c/Current%20Rules?>



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY:

TITLE: ED Report

INTRODUCTION TO THE TOPIC:

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description

ED Report

Upload Date Type

6/12/2026 Cover Memo



Minnesota Board of Psychology Executive Director Report

June 12, 2026

Introduction

The mission of the Board is to protect the public through licensure, regulation, and education to promote access to safe, competent, and ethical psychological services. The work of the Board is strategically aligned to accomplish this mission, including prioritization of Board action and the assignment of resources (both human and financial).

The work of the Board has focused on the following since the last Board meeting:

I. Administrative Updates

a. Assistant Executive Director Licensing Update

The Licensure Team has continued to support the Mission and Vision of the Board by processing Psychologist and Behavior Analyst license applications. The Board has now issued over 1,050 Behavior Analyst licenses and continues to receive paid applications with a consistent number of at least four per week. Board staff are working with IT to develop the Behavior Analyst license renewal process.

The team has updated the Board's website with the focus of making the steps to licensure clearer. First the team has posted the checklist for applicants trained in an educational institution outside the United States to the FAQ section for international applicants. Second staff have provided more clarity to the requirements for the guest licensure application types posted on the website. The Board has two guest license application types, one for applicants concurrently seeking temporary and full licensure and one for psychologists who only plan to practice in Minnesota for more than thirty calendar days and for less than nine consecutive months. The main revision was to clearly state all the requirements as they appear in MN Statute 148.916, subd 1, which illustrates that all guest license applicants need to take and pass the PRE, the state jurisprudence exam.

Finally, the Licensure Team presented to a group of post-doctoral supervisees on the topics of supervision, licensure documentation requirements, and the supervision agreement.

II. Executive Director's Report

- a. Education: The licensure team presented to a group of post-doctoral students. One change to this presentation was to have the licensure team present at the training. This affords applicants or future applicants to begin identifying Board staff who will be

assisting them with navigating the licensure process with the Board. The presentation itself was very well received, and it was noted that the attendees appreciated putting a face to name of staff that are behind the scenes responding to emails.

The complaint team presented to post-doctoral students at a local agency. The team presented on the complaint process and Board investigations. As with the licensure team it is important for our applicants and licensees to see how Board staff and Board members apply statute and rule to complaints.

- b. The use artificial intelligence in behavioral health services: The Executive Director continues to meet with his counterparts at the other behavioral health boards to understand how boards can collaborate to better address the artificial intelligence questions all the boards are receiving. The Executive Director is working toward the goals set by the board of exploring presentation possibilities for Board members on AI use in behavioral health as well as best practices in regulating AI in behavioral health settings.
- c. Legislative Updates: [HF3825](#) made it unlawful for “[A]ny person who is not a natural person to engage in the practice of psychology or misrepresent themselves as a psychologist or psychological practitioner.” It also codified that the practice of psychology is a gross misdemeanor for natural persons practicing psychology or misrepresenting themselves as a psychologist.

[HF3893](#) – Regulating artificial intelligence in psychotherapy bill did not pass this session. The bill proponents have indicated that it will be back next year.

[HF2590](#) – Licensure for Clinical Art Therapists did not pass this session. The bill proponents have indicated that it will be back next year.

[HF2906](#) – Psilocybin Registration Program did not pass this session. The bill proponents have indicated it will be back next year.

- d. Financial Updates: the end of FY26 will be June 30, 2026. Board staff are ensuring all the Board’s outstanding invoices have been paid and are working to begin the close of the fiscal year. Budgets for FY27 have been completed and approved. Staff are finalizing authorizations for purchases orders that require special approval to be ready to utilize as of July 1.



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY:

TITLE: Variance Request

INTRODUCTION TO THE TOPIC:

I have experienced significant economic hardship due to the ongoing delay in my licensure process. I initiated this application last year after relocating to the state for medical care. In the interim, I have been limited to seeing a small number of clients through my New York license, which I have held unrestrictedly since November 21, 2024.

Reasons for variance or waiver requested I would like to clarify a specific challenge regarding the board's requirement that all hours before independent licensure be attested. Following my postdoctoral fellowship, I was employed as a Level I psychologist. During that period, my permit was signed by a supervisor who was not required to track supervision hours, as the hours necessary for licensure had already been fully satisfied during my two-year postdoctoral fellowship.

Requiring this supervisor to attest to hours they were not originally mandated to track has created a significant obstacle. Given that my licensure requirements were already met before this role, the supervisor signed this permit for all Level I psychologists at that location to be able to practice independently, even if they did not sign off on any hours.

Not adversely affect the public welfare

Approving this request will benefit public welfare by providing access to a qualified doctoral-level mental health and wellness professional.

I have made multiple unsuccessful attempts to communicate with the second supervisor on the following dates:

Written plan of alternative practices

- April 26, 2026, at 1:27 PM
- May 4, 2026, at 6:03 PM
- May 11, 2026, at 8:23 AM

Despite my attempts to establish a clear path forward, I am currently left with no conclusive action or way to proceed with the existing requirements.

BOARD ACTION REQUESTED:



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY: Assistant Executive Director

TITLE: FARB Summit on Regulatory Excellence

INTRODUCTION TO THE TOPIC:

Summit materials and cost proposal are submitted for your review.

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description

FARB Summit Agenda Summary

FARB Cost Proposal

Upload Date Type

6/12/2026

Cover Memo

6/12/2026

Cover Memo



2026 FARB Summit on Regulatory Excellence

July 23-25, 2026
The Marquette Hotel
710 Marquette Ave
Minneapolis, MN

Cultivating Courageous Leadership: Leading with Conviction, Collaboration, and Clarity

The FARB Summit on Regulatory Excellence is the premier gathering for regulatory law professionals, licensure board members and staff, and leaders of national and international regulatory associations. Designed to explore the intersection of law, policy, leadership, and innovation, the Summit offers timely insights and practical strategies to strengthen professional regulation in a rapidly evolving environment.

As scrutiny of licensure intensifies and public expectations continue to shift, courageous leadership is essential to uphold integrity, transparency, and trust. This year's Summit will focus on developing the mindset and skills needed to lead with confidence, whether navigating political pressures, guiding boards through change, addressing complex enforcement issues, or shaping national policy. Participants will leave empowered to champion regulation that serves the public interest with foresight, conviction, collaboration, and clarity.

Summit Format and Learning Tracks

The Summit features a combination of plenary sessions and concurrent breakout tracks designed to deliver targeted learning experiences for FARB's key audiences. Across both tracks, sessions are designed to strengthen courageous leadership and equip attendees to lead with conviction, collaboration, and clarity in service of the public interest.

Track 1: Regulatory Leadership & Practice (CAE Track)

Designed for licensure boards, regulatory agencies, and national and international regulatory associations. Track 1 sessions are structured to support continuing education for Certified Association Executives. CAE accreditation details will be confirmed closer to the Summit.

Track 2: Regulatory Law & Enforcement (CLE Track)

Designed for regulatory law professionals and includes continuing legal education (CLE) content to support attorney licensure renewal requirements. Track 2 sessions are structured to support CLE credit. CLE accreditation details will be confirmed closer to the Summit.

****Please note:*** Session titles, presenters, and descriptions are subject to change without notice. We appreciate your understanding as we work to provide the most timely, relevant, and engaging program possible.

Thursday, July 23
Pre-Summit Workshops
2:00 - 5:00 p.m.

Pre-summit workshops are optional professional development sessions designed to support regulators and legal counsel in their ongoing growth and pursuit of regulatory excellence. Each in-depth learning opportunity is available to Summit attendees for an additional fee of \$150 per session. As the workshops are offered concurrently, attendees may select one to attend.

Track 1: Regulatory Excellence Training: Best Practices for Board Members and Staff
Presenter: Dale Atkinson, The Atkinson Firm

This interactive workshop provides a practical overview of core regulatory responsibilities and proven best practices for effective board and agency operations. Participants will explore key governance principles, decision-making frameworks, and strategies for strengthening consistency, accountability, and public trust. Designed for both new and experienced regulators, this session offers tools and insights that can be applied immediately in day-to-day regulatory work, supporting leaders in making confident decisions with clarity and purpose.

Track 2: Ethics, Wellness, and Professionalism for Regulatory Law Professionals

This workshop explores the ethical responsibilities and professional challenges faced by regulatory attorneys and legal staff, with a focus on maintaining sound judgment under pressure. Participants will examine how implicit bias, stress, and burnout can influence legal decision-making and professional conduct. The session will provide practical strategies for strengthening professionalism, resilience, and ethical integrity, reinforcing the role of courageous leadership in high-stakes regulatory environments.

Optional Dinner Circles
6:00 p.m.

Meet new regulatory colleagues or join old friends in one of our optional dinner circles. FARB leadership will serve as hosts at a variety of local restaurants, allowing you to connect with fellow Summit attendees. Sign up for the restaurant of your choice at the registration desk throughout the day.

Friday, July 24
2026 FARB Summit on Regulatory Excellence

8:00 – 8:50 a.m. Networking Breakfast with Sponsors

9:00 – 9:50 a.m. Opening Plenary & Keynote - Leading Courageously During Times of Regulatory Challenge

Presenter: Donald A. Balasa, JD, MBA, American Association of Medical Assistants

Courageous leadership includes the willingness to challenge oneself and those around us to do better. Leaders also provide foresight on ways to navigate during times of challenge while being humble and open to critiques. Recent public debate over occupational licensing has intensified as arguments are presented stating that licensing restricts workforce participation and increases costs for consumers.

In this keynote session, Donald Balasa will examine major critiques of occupational licensing, acknowledging where reform conversations may have merit while also highlighting where key assertions are not supported by law, fact, or economic theory. The session will explore how regulators can courageously and thoughtfully respond to the growing movement toward deregulation and articulate the value of regulation and its critical role in public protection.

This presentation will help regulators lead with clarity and conviction in policy discussions about professional regulation, while remaining grounded in the public protection mission.

9:50 – 10:10 a.m. Networking Break

10:10 – 11:00 a.m. Breakout Sessions (Concurrent)

Track 1: AI: Charting the Course from Principle to Practice

This session will examine how establishing principles of use can help regulators navigate the introduction of artificial intelligence into professional practice and their own work. Then, this session will highlight the difficulty of regulating a technology that intersects with professional practice but is quickly evolving outside of current occupational licensing frameworks. Discussion will include recommendations for responsible use by licensees and regulators, misuse by unlicensed individuals, and implications for enforcement, consumer protection, and evolving standards. Participants will gain practical insights into how boards can better anticipate AI-related issues and respond to them more proactively and consistently—all while maintaining transparency and public trust as technology rapidly evolves. Participants will also gain a better understanding of the limits of current regulatory frameworks and the need for new policies to address emerging threats and the largely unregulated digital professional services.

Track 2: Secondary Trauma in Investigations and Trauma-Informed Prosecution

This session will examine the impact of secondary trauma on regulatory attorneys and investigative teams and how trauma-informed approaches can strengthen outcomes. Discussion will include wellness considerations, effective communication strategies, and how trauma-informed principles

intersect with licensing questions, enforcement, and due process. Participants will gain practical tools for supporting both professional resilience and fair regulatory practice, reinforcing the importance of leading with integrity and sound judgment in difficult circumstances.

11:10 a.m. – Noon Breakout Sessions (Concurrent)

Track 1: Coalition Building and Public Engagement: Strategies for Building Trust and Influence

This session will explore practical strategies for building effective coalitions and engaging the public in ways that strengthen trust in regulation. Participants will examine how regulatory boards, agencies, and associations can collaborate with stakeholders, communicate with clarity, and navigate competing interests while remaining grounded in the public protection mission. The session will provide tools and real-world approaches for building alignment, increasing transparency, and strengthening credibility in today's evolving regulatory environment.

Track 2: Speech vs. Conduct: Supreme Court and Constitutional Considerations

This session will explore recent legal developments affecting the regulation of professional conduct, with attention to constitutional boundaries and First Amendment considerations. Attendees will review key case examples and discuss how courts interpret speech versus conduct in professional regulation. The session will provide practical guidance for legal counsel advising boards and agencies as they navigate complex legal landscapes with clarity and confidence.

Noon – 1:00 p.m. Networking Lunch with Sponsors

1:10 – 2:00 p.m. Partner Breakout Sessions (Concurrent)

Track 1: Exam Development: Getting it Right for Regulation

This partner-led session will provide insight into examination development and the role of assessment in supporting competency-based licensure. Participants will explore best practices in test design, validity, security, and modernization strategies. The session is designed for regulators and board staff seeking a deeper understanding of the licensure pipeline and how strong assessment systems support public trust and professional accountability.

Track 2: The Legal Issues Everyone's Watching

This session will provide an in-depth exploration of a timely legal issue affecting professional regulation and board operations. Content will be designed to support regulatory attorneys and agency counsel with practical guidance and emerging trends. Additional details will be announced as the Summit program is finalized.

2:10 – 3:00 p.m. Breakout Sessions (Concurrent)

Track 1: Public Board Member Success Profile

This session will focus on what makes an effective public board member and how boards can better recruit, onboard, and support public participants. Discussion will explore expectations, governance dynamics, and strategies for maximizing public member contributions to board decision-making. Participants will leave with practical tools for strengthening board performance and public trust by fostering accountability, clarity, and shared leadership.

Track 2: Unlicensed Practice: Enforcement Tools and Legal Strategies

This breakout will focus on the legal and practical challenges of addressing unlicensed practice, including investigative approaches, enforcement options, and litigation considerations. Participants will explore strategies for improving compliance while ensuring due process and appropriate public communication. The session is designed to strengthen consistency and confidence in enforcement decision-making under increasing public and political scrutiny.

3:00 – 3:20 p.m. Networking Break

3:20 – 4:30 p.m. Day 1 Closing Plenary: Policy Maker Panel Discussion

This moderated panel will bring together policymakers and regulatory stakeholders to discuss current legislative trends affecting professional licensure and oversight. Panelists will explore emerging reform efforts, political pressures, and the evolving expectations placed on regulatory bodies. Participants will leave with greater clarity on how to engage effectively and lead with conviction while remaining grounded in the public protection mission.

5:30 – 6:30 p.m. Networking Reception

Join fellow Summit attendees for an evening networking reception designed to foster connection and collaboration across the regulatory community. This informal gathering offers an opportunity to build relationships with regulatory peers, exchange ideas, and strengthen cross-jurisdiction partnerships that support courageous leadership and regulatory excellence.

Saturday, July 25
2025 FARB Summit on Regulatory Excellence

8:00 – 8:50 a.m. Networking Breakfast with Sponsors

9:00 – 9:50 a.m. Atkinson Lecture

Presenter: Courtney Arbour, Texas Department of Licensing and Regulation

This lecture series offers perspectives on the state of regulation to inspire and provoke important discussions throughout and beyond the summit.

9:50 – 10:10 a.m. Networking Break

10:10 – 11:00 a.m. Breakout Sessions (Concurrent)

Track 1: The Burden and the Benefit: Balancing the Occupational Regulation Equation

This moderated fireside chat will feature two leading researchers examining both the benefits and the costs of occupational regulation. Through an engaging dialogue, the session will explore how licensure supports public protection and professional accountability, while also considering the societal and workforce impacts of restrictive entry requirements. Participants will gain a balanced, evidence-based perspective to support thoughtful policy discussions and confident decision-making.

Track 2: Who Can Represent? Representation Challenges and Pro Se Issues

This session will examine common issues related to representation in administrative proceedings, including unauthorized practice of law concerns and challenges when respondents appear pro se. Participants will discuss procedural fairness, risk management, and strategies for maintaining orderly and defensible hearings. The session will provide guidance for navigating complex representation scenarios in regulatory settings while upholding integrity and due process.

11:10 a.m. – Noon Partner Breakout Sessions (Concurrent)

Day 2 partner sessions will focus on industry leaders presenting ideas and solutions that address current legal and regulatory challenges that all regulators face. Learn from current and potential future partners to help you solve your most pressing concerns.

Noon – 1:00 p.m. Networking Lunch with Sponsors

1:10 – 2:00 p.m. Breakout Sessions (Concurrent)

Track 1: The Future of Workforce Wellbeing

Professions across many sectors are experiencing notable change as workforce demands, training pathways, and public expectations evolve. This session will explore key trends shaping the future of professional practice and discuss considerations for regulatory leaders as they respond to workforce and policy developments.

Track 2: Collateral Consequences of Disciplinary Databases

This session will explore the downstream impacts of public disciplinary reporting, including how databases influence licensee mobility, employment outcomes, and public perception. Discussion will include legal and policy considerations related to transparency, accuracy, and fairness in reporting. Participants will examine how boards can balance public protection with proportional consequences.

2:10 – 3:30 p.m. Closing Plenary Session: Agency Leadership Panel

This plenary session will feature senior agency leaders sharing lessons learned from managing regulatory systems through reform, modernization, and heightened public accountability. Discussion will focus on governance models, organizational leadership challenges, and strategies for maintaining trust and operational effectiveness. The session is designed to provide practical takeaways for both board leadership and staff.

3:30 – 5:00 p.m. Agency Leadership Roundtable (Invitation Only)

The Agency Leadership Roundtable is an invitation-only session convening agency directors and senior regulatory staff for candid discussion on shared challenges and emerging issues impacting regulatory agencies nationwide. This facilitated dialogue provides a confidential space to exchange perspectives, explore innovative approaches, and identify opportunities for collaboration across jurisdictions. Participation is limited to invited agency leaders to support focused peer-to-peer engagement.

3:30 – 5:00 p.m. FARB Governing Council Meeting (Invitation Only)

The FARB Governing Council Meeting convenes representatives from FARB's member organizations to conduct official Council business and advance shared priorities. This working session provides an opportunity to discuss emerging regulatory issues, coordinate cross-organizational collaboration, and guide strategic initiatives supporting regulatory excellence. The meeting is intended for Governing Council members and invited participants.



- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 6/12/2026

SUBMITTED BY: State Program Administrator

TITLE: Board Administrative Terminations

INTRODUCTION TO THE TOPIC:

The Board shall terminate the license of a licensee whose license renewal is at least 60 days overdue and to whom notification has been sent as provided in the administrative rules. Failure of a licensee to receive notice is not grounds for later challenge of the termination.

Licensees are provided several opportunities to renew the license prior to Board termination. Licensees are sent a notice within 30 days after the renewal date when they have not renewed the license. This letter is sent via certified mail to the last known address of the licensee in the file of the board. This notifies the licensee that the license renewal is overdue and that failure to pay the current renewal fee and the current late fee (\$250.00) within 60 days after the renewal date will result in termination of the license. A second notice is sent to the licensee at least seven days before a board meeting (which occurs 60 days or more after the renewal date).

Minn. R. 7200.3510.

BOARD ACTION REQUESTED:

License Number	Name	License Expiration Date
LP6515	Ann Haynos	2/28/2026
LP6226	Rosaire Daigle	2/28/2026
LP3425	Judith Peters	2/28/2026
LP0826	George Lindall	2/28/2026
LP3952	Mary Glomb	2/28/2026
LP3949	Mary Emanuelson	2/28/2026