Office of Artificial Intelligence Policy
UTAH DEPARTMENT OF COMMERCE

Division of Professional Licensing
UTAH DEPARTMENT OF COMMERCE

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.