# Responsible AI Adoption in Community Mental Health Organizations:

# A Study of Leaders' Perceptions and Decisions

Mabel Yiu

Rossier School of Education

University of Southern California

A dissertation submitted to the faculty

in partial fulfillment of the requirements for the degree of

Doctor of Education

May 2025

The Committee for Mabel Yiu certifies the approval of this Dissertation

Ruth Chung

Albert Rizzo

Patricia Tobey, Committee Chair

Rossier School of Education

University of Southern California

2025

## Abstract

This research examined the technological, organizational, and environmental factors impacting how mental healthcare leaders responsibly adopt artificial intelligence (AI) technologies. Drawing from interviews with mental health leaders, the paper identifies key factors, including leaders' perceptions of AI's capabilities and limitations, the organization's technological readiness and change management processes, and broader industry pressures. While ethical AI principles focused on patient autonomy, avoiding harm, and promoting well-being are essential, this paper explored the meaning of "responsible" in AI adoption that requires mental health leaders to meet ethical standards and take direct accountability for outcomes, anticipate potential challenges, obtain stakeholders' buy-in, and maintain the adaptability as healthcare needs evolve. The findings reveal that while most interviewed mental health leaders are not technical experts in AI, they share a cautiously optimistic outlook about AI's potential to enhance treatment effectiveness and reduce provider burnout, tempered by concerns about the absence of comprehensive guidelines, data integrity, and algorithmic accuracy. Leaders specifically expressed apprehension about AI's cultural competency limitations, the risk of over-reliance and ethical dissonance, funding inequities, and potential impacts on therapeutic rapport in mental healthcare delivery. By understanding the multifaceted influences on leaders, the research aims to provide a practical guide for harnessing the benefits of AI while preserving the essential human elements of compassionate, patient-centered care, ensuring AI augments rather than replaces clinical expertise.



*Keywords*: artificial intelligence, mental health, responsible AI, technology-organization-environment, leadership, digital transformation, ethics

## Dedication

To my husband, Randy, thank you for your unwavering love and support throughout this journey. I could not have accomplished this without you.

To my children, Sophie and Daphne, may you always reach for the stars, even if you sometimes land on someone's roof. Your curiosity and enthusiasm are a constant inspiration.

To Yvonne, Ingrid, and my mom, thank you for your guidance and for believing in me. I hope I have made you proud.

To my heavenly dad, there is another Dr. Yiu in the house after this dissertation is published. No, it is not a medical doctor; deal with it. If you are indeed smiling down on me, I would prefer if you smile somewhere that I cannot see you. I do not need your supernatural parental gazes upon me.

To my classmates, you have each played a role in pushing me to be my best. I am honored to call you colleagues and friends.

To Isaac Takeuchi, thank you for listening to me and pushing me along the way with loads of humor. I will always cherish the moments when we were called on to speak for our community. It was uncomfortable but interestingly enriching.

To Chris Thomas, you inspired me to be a better person through your passion and compassion for restorative justice (and everything in life). I am honored to have called you a friend. I will forever miss our banters and shared cat memes. Though you are no longer here, your spirit lives on and fights on at USC—especially now that you are immortalized in my dissertation here.

## Acknowledgments

I want to express my sincere gratitude to my dissertation committee for their invaluable guidance and support throughout this research process.

Thank you, Dr. Patricia Tobey, for your insightful feedback and thought-provoking questions during our discussions. Your expertise in mental health and research has been instrumental in shaping the theoretical foundation of this work.

Thank you, Dr. Ruth Chung, for your willingness to lend your comprehensive knowledge and analytical skills to this project. Your constructive critiques have helped strengthen the rigor and depth of my analysis.

Thank you, Dr. Albert Rizzo, for your unique VR and AI perspective in this interdisciplinary study. Your expertise has challenged me to consider novel approaches and broaden the implications of my findings. And your wise words, “don’t boil the ocean,” saved me more times than you knew.

A special thank you goes to the mental health leaders who graciously agreed to be interviewed for this research. Your willingness to share your perspectives, experiences, and insights has been invaluable. As the frontline innovators and decision-makers in this field, your participation has provided crucial, first-hand knowledge that has shaped the findings and recommendations of this dissertation. I sincerely appreciate the time you dedicated amid your demanding schedules and openness in discussing the complex challenges and opportunities surrounding the responsible adoption of AI technologies in mental healthcare. I am sincerely grateful that this research would not have been possible without your meaningful contributions.

## Table of Contents

## List of Tables

## List of Figures

# Chapter One: Overview of the Study

Artificial intelligence (AI) has the potential to revolutionize mental health by automating routine tasks and administrative processes, allowing healthcare providers to focus more on patient relationships and care delivery (Familoni, 2024; Marr, 2023; World Health Organization, 2021). The World Health Organization (2021) defined an AI system as a machine-based system capable of making predictions, recommendations, or decisions in response to a set of specific human-defined objectives. Although it has many definitions, AI involves computational technologies that can mimic the way humans sense, learn, reason, create, analyze, predict, and take action (IEEE, 2019). As AI adoption accelerates in mental healthcare, concerns grow about its impact on workforce roles and care quality (Moradi & Levy, 2020; World Health Organization, 2021).

While AI offers clear benefits in streamlining patient care and documentation, its implementation requires careful leadership guidance (Alami et al., 2020). This study addressed a critical challenge facing mental healthcare organizations: the adoption of AI without adequate planning may compromise patient safety, public trust, and the essential human elements of therapeutic care. To mitigate this challenge, this research investigated how these organizations can responsibly implement AI in a way that maintains the integrity of therapeutic relationships, ensures patient privacy, and upholds ethical standards in healthcare delivery through the technology-organization-environment (TOE) framework.

While at least 84 public-private initiatives published AI principles from different AI governance organizations, they struggle to translate abstract principles to practical implementation (Mittelstadt, 2019). According to Akbarighatar (2024), organizations need to balance both technical and organizational aspects to implement AI responsibly. To do so, these organizations need practical guidance, not abstract principles (Akbarighatar, 2024). Using the TOE framework to analyze mental health leaders' (MHLs) perspectives on AI adoption, this study examined how technological capabilities, organizational readiness, and external factors influence responsible implementation decisions in mental health care settings. This paper identifies that the concept of responsible AI adoption goes beyond ethics and provides a concrete process that enables these organizations to leverage AI's benefits while maintaining human-centered care through specific technological, organizational, and environmental capabilities required to integrate AI responsibly. This study sought to reveal gaps in leadership knowledge, resource planning, and other barriers to ethically implementing AI innovations. This insight across individual, organizational, and environmental levels may offer recommendations and strategies to promote adoption readiness, mitigate risks, allocate resources, and shape company policy to enable effective AI adoption.

It is important to acknowledge that this research, begun in 2023, captures MHLs' perspectives during the early stages of AI adoption in healthcare. By the time this paper is completed in 2025, both AI capabilities and leaders' understanding and attitudes toward AI have evolved considerably. When this research began, the American Psychological Association (APA, 2024) was in early discussions about AI's ethical implications and responsible use in mental health. By the time this study was completed, the APA had developed preliminary guidelines for AI implementation in mental healthcare settings, demonstrating the field's rapid progress in addressing these emerging challenges (Stringer, 2025). Nevertheless, the fundamental insights about leadership, ethics, and organizational readiness provide enduring value for mental healthcare organizations navigating AI adoption.

## Context and Background of the Problem

While AI adoption accelerates across clinical and administrative mental health care, current abstract ethical frameworks fail to provide the practical implementation guidance needed to protect patient safety and preserve the human-centered nature of care. Nonetheless, AI has made substantial advances in both clinical and administrative aspects of mental health care (Nguyen et al., 2022). On the patient side, AI applications enhance direct patient care through psychoeducation, therapeutic interventions, and predictive analytics (Sinha et al., 2023). For example, AI-powered chatbots like Woebot demonstrate some capability in suicide risk assessment and predictive analysis (Prochaska et al., 2021), while virtual reality platforms offer innovative therapeutic approaches (Rizzo et al., 2021). These patient-facing technologies increase accessibility to mental health support and help engage hard-to-reach populations while potentially reducing treatment stigma (Fiske et al., 2019; Stix, 2018).

On the administrative side, AI streamlines operational processes and enhances organizational efficiency by assisting with assessment and documentation automation, scheduling optimization, insurance prior authorization, and claim processing (Bickman, 2020; Lenert et al., 2023). The technology helps maintain regulatory compliance through automated audit trails, supports resource allocation through predictive staffing models, and enhances revenue cycle management through improved billing accuracy (Fiske et al., 2019). These administrative applications free up clinicians’ time, allowing them to focus more on direct patient care while improving operational efficiency.

While AI’s adoption in healthcare offers clear benefits, the risks and stakes exceed those of conventional technological implementations (Graham et al., 2019). However, there are abstract ethical principles but no concrete steps to help organizations adopt AI responsibly

(Mittelstadt, 2019). Concerns span both clinical and administrative domains, from the potential displacement of human roles to risks in patient care quality (Allyn, 2023; Mendes-Santos et al., 2022). The challenge lies in balancing human expertise and empathy with technological advancement, ensuring that AI augments rather than replaces the essential human elements of mental health care delivery (Carr, 2020; Zidaru et al., 2021).

## Purpose of the Study and Research Questions

The purpose of this study was to investigate the individual, organizational, and external factors that influence the acceptance of AI tools and provide recommendations that facilitate the responsible integration of AI. This research aimed to provide a practical guide for MHLs and developers of AI mental health products navigating the meaning of responsible implementation of AI and how it is different from ethical implementation, even as the dynamic field continues to evolve rapidly. It is akin to attempting to fix an airplane while it is still in flight, a challenging but necessary task given the potential benefits and risks of AI in mental healthcare. For MHLs with limited technical expertise, this study sought to provide a preparation roadmap for AI adoption. For AI developers, it sought insights into organizations' holistic needs and operational realities, enabling more effective technology development and implementation. By examining the balanced approaches to implementation, the study sought to bridge the gap between AI's technological capabilities and mental health care's fundamental human elements. The goal was to provide mental health organizations with a framework for responsible AI adoption that enhances service delivery while preserving therapeutic relationships, ultimately benefiting both healthcare providers and patients.

This study focused on leaders with backgrounds in community mental health; however, its findings may also be useful to other professionals in the field, professional associations, patient advocacy groups, educators, and students who are attempting to adopt AI responsibly in a manner that maintains ethical standards and complements clinical expertise with technology. The research questions guiding this study are as follows:

1. What are the MHLs' perceptions of adopting AI technologies in their organizations?
2. What organizational factors influence MHLs to adopt AI technologies responsibly in their organization?
3. What external factors are MHLs concerned about the most when evaluating their organization's readiness to adopt AI technologies?

By addressing the research questions, this study sought to develop an understanding of MHLs' perceptions, intentions, and decisions regarding adopting AI technologies in their organizations. The first question focuses directly on leaders' views of AI adoption, including perceived benefits, risks, barriers, and motivations. Understanding these perceptions provides insight into their openness and attitudes toward these tools. The second question looks at organizational factors that may influence adoption, such as resources, workflows, leadership support, and training capabilities. Examining the organizational context reveals facilitators and obstacles at a systems level. The third question explores how external environmental factors influence leaders' decisions on AI adoption. These include pressures and influences from regulatory policies, competition, professional standards, social influences, and patient needs and preferences. Analyzing the environmental context highlights external forces that may motivate or hinder organizational adoption of AI technologies. Together, these questions provide a comprehensive framework grounded in technology acceptance and diffusion of innovations theory to study multiple levels of influence on AI adoption decisions.

## Importance of the Study

Implementation of AI without considering responsibility in mental health care exacerbates healthcare inequities, harming vulnerable patients, employee engagement, and organizational effectiveness (Ahmed et al., 2023; Hoffman & Podgurski, 2020). This research examined how to ethically integrate AI while preserving the essential human elements of care delivery. By focusing on the perspectives of MHLs, this study aimed to develop guidelines for adoption that enhance care quality and efficiency while maintaining the therapeutic relationship at the core of effective treatment. The goal is to ensure AI enhances rather than replaces human clinical expertise and connections, supporting a balanced approach that improves service delivery while preserving the compassionate, human-centered nature of mental health care.

## Overview of Theoretical Frameworks

Integrating AI in mental healthcare presents challenges that require careful consideration of multiple factors (Dos Santos & Rosinhas, 2023). This study applied the TOE framework to develop a comprehensive understanding of the factors influencing the responsible integration of AI in mental healthcare settings. The TOE framework (Tornatzky et al., 1990) provides a lens to analyze the technological, organizational, and environmental dimensions that shape AI adoption, including organizational culture, resources, regulations, and ethical landscape.

While existing research on the TOE framework in ethics and mental health settings is limited, its principles can be applied to understand AI adoption in complex mental health environments. While AI tools like natural language processing show promise in diagnosis and treatment, their success hinges on organizational preparedness, staff training, and adherence to mental health regulations (Henry et al., 2021). The framework provides a multidimensional perspective on AI integration, illuminating both the human and contextual drivers of acceptance

(Chatterjee et al., 2021). The TOE framework is adaptive and relevant to multiple industries and may also aid in navigating issues like patient privacy, algorithmic biases, and transparency in AI-driven solutions. (Baker, 2011). Past research found that technological factors, along with individual and external variables, positively influence the perceived usefulness of AI systems (Na et al., 2022). The multidimensional TOE provides a robust framework for investigating the adoption of AI in this context.

## Definitions

*Algorithmic bias*: systematic and unfair discrimination that emerges in AI systems due to biased training data, leading to unjust or inequitable outcomes (Turner Lee et al., 2023).

*Artificial intelligence* is a broad field focused on creating systems that can perform tasks and typically mimic human intelligence, such as reasoning, problem-solving, perception, and language understanding (World Health Organization, 2021). It involves developing machine learning algorithms and technologies that enable computers to perform tasks that typically require human intelligence, such as problem-solving, learning from experience, recognizing patterns, and making decisions (IEEE, 2019).

*Community mental health* refers to a system of services and support provided in a community-based setting rather than in centralized facilities, such as large psychiatric hospitals (Sharfstein, 2019). Community mental health services often include counseling, therapy, crisis intervention, medication management, and rehabilitation, along with supportive services that address housing, employment, and social connections.

*Electronic medical records* and *electronic health records* (EHRs) both store patient information and medical records in digital formats, allowing physicians to improve the quality of care and manage costs more effectively. Electronic medical records are typically used within a single healthcare organization, serving as an internal system. In contrast, EHRs are designed for use across multiple organizations, making them an inter-organizational system that supports the sharing of patient data between healthcare providers (Heart et al., 2016). This research uses the acronym of EHR.

*General Data Protection Regulation* (GDPR): A comprehensive EU data privacy law enacted in 2018, granting EU citizens control over their personal data and applying to any company that processes their data, regardless of location. Although the United States lacks a federal equivalent, GDPR's global reach impacts U.S.-based companies with international clients, driving significant academic and legal analysis (Niebel, 2021). As a global standard for privacy, GDPR has influenced emerging laws worldwide, including the California Consumer Privacy Act in the United States (Bukaty, 2019).

*Healthcare inequalities* are disparities in access to healthcare services, treatments, and outcomes, often influenced by social, economic, and cultural factors, which ethical AI implementation should strive to address (Patton-López, 2022). These fall under environmental, as it relates to systemic biases and disparities in the broader healthcare sector that AI solutions should aim to address.

*Health Insurance Portability and Accountability Act* (HIPAA) is a U.S. federal law enacted in 1996 to protect sensitive patient health information from unauthorized access and disclosure. The HIPAA sets national standards for the security and privacy of protected health information and applies to healthcare providers, insurance companies, and other entities handling it. It also includes provisions to ensure the secure electronic transfer of health information and gives patients certain rights over their health data (Krager & Krager, 2016).

*Human-centered care* or *patient-centered care*: A holistic approach to providing services that prioritizes the well-being, preferences, and needs of individuals, placing them at the center of decision-making and care delivery (Engle et al., 2019). This approach spans both organizational and environmental facets. An organization must focus on patient-centered values and care while also recognizing the larger landscape of ethical clinical relationships and standards.

*Machine learning system*: A subset of AI that specifically involves training algorithms to learn from data and make predictions or decisions without being explicitly programmed for each task (Laijawala et al., 2020; Liu, 2021). Opposite of rule-based systems, instead of relying on predefined rules, the AI learns patterns, relationships, and decision criteria directly from large datasets. Doing so allows it to adapt and improve over time as it processes more data, making it flexible and capable of handling complex, unpredictable scenarios without explicit programming for each condition (Goodfellow et al., 2016)

*Managed care*: A healthcare approach that focuses on coordinating and managing healthcare services, treatment plans, and costs for individuals, often involving collaboration between mental health providers, organizations, and insurers (Borghouts et al., 2021).

*Mental health consumers* (consumers): individuals seeking mental health support, treatment, or services whose well-being and privacy are of utmost importance in the ethical implementation of AI technologies (Joerin et al., 2020). This relates to the external environment of patients and their willingness to accept AI-enabled care.

*Mental health leaders*: Individuals with leadership positions at mental health organizations who are responsible for making decisions, setting strategies, and guiding the ethical implementation of AI technologies in their organizations (Petersson et al., 2022).

*Mental health organizations* are institutions or agencies that provide mental health services, support, and resources to individuals, families, and communities. These organizations may offer a range of services, including assessments, therapy, crisis intervention, rehabilitation, prevention programs, and patient and clinician advocacy. They can operate in various settings, such as hospitals, clinics, community centers, and online platforms, and may be public, private, or non-profit.

*Mental health providers*: Professionals who offer mental health services, including psychologists, psychiatrists, therapists, counselors, and social workers, and who play a crucial role in the ethical integration of AI technologies (Petersson et al., 2022).

*Organization readiness*: The state of preparedness and adaptability of a mental health organization to incorporate AI technologies, including assessing technical, cultural, and ethical aspects (Alami et al., 2020).

*Payors* (or healthcare payors) are entities responsible for financing or reimbursing healthcare costs, covering services patients receive from providers rather than delivering care directly. Public payors (like Medicare and Medicaid), private payors (such as private insurance and employer-sponsored plans), and self-pay individuals each play a critical role in managing healthcare expenses, setting coverage terms, and influencing cost-effective care through negotiations with providers (Shi & Singh, 2004).

*Principlism*, in healthcare ethics, is an approach that relies on four core principles to guide ethical decision-making in medical practice: autonomy (respecting the patient's right to make informed choices), beneficence (promoting the well-being of the patient), non-maleficence (avoiding harm), and justice (ensuring fair distribution of resources and treatment; Beauchamp & Childress, 1979; Cahn & Markie, 2011; Parsons & Dickinson, 2016; Pasricha, 2022). These principles, proposed by Tom Beauchamp and James Childress, provide a practical framework for addressing ethical dilemmas in healthcare and are widely used in bioethics (Childress, 1983).

*Productivity* is the measure of an organization's efficiency and output, including the delivery of quality services while responsibly and ethically leveraging AI technologies (Alami et al., 2020). Managed care payors such as Medicare base payments to organizations and providers on productivity (Newhouse & Sinaiko, 2007).

*Professional network bias* is the potential for research bias due to personal connections in the author's professional network that might influence participant selection, responses, or interpretation of findings (Petersson et al., 2022).

*Rule-based system*s operate based on predefined, static rules and criteria. Each rule determines actions based on specific conditions, making it transparent and easy to modify (Enqvist, 2024). However, it lacks adaptability, as it cannot learn or handle exceptions outside its predefined rules, unlike machine learning models (S. J. Russell & Norvig, 1995).

*Social desirability bias* refers to the tendency of participants to provide responses perceived as socially acceptable or desirable rather than fully reflecting their actual opinions or experiences (Joerin et al., 2020).

*Virtual reality* (VR) is broadly defined as a way for humans to visualize, manipulate, and interact with computers and highly complex data (Rizzo & Koenig, 2017). It can be non-immersive, where users interact with the virtual environment via a computer monitor, or immersive, where users engage with the environment through a VR headset. Machine learning algorithms or rule-based AI systems can enhance data and responses within VR environments.

*Vulnerable populations*: Individuals or groups who are at a higher risk of experiencing negative impacts due to their socioeconomic status, health conditions, or other factors, requiring special consideration in AI implementation (Kuran et al., 2020).

*Zero-knowledge proofs* can solve the AI transparency problem by allowing one party, known as the prover, to convince another party, the verifier, that a particular statement is true without revealing any additional information beyond the validity of the statement itself (Bai et al., 2022).

## Organization of the Dissertation

This dissertation is organized into five chapters. Chapter One introduced the issue of mental health staff attitudes influencing AI adoption and outlined the research goals, significance, stakeholders, conceptual frameworks, and methodology. Chapter Two reviews relevant literature on AI in mental healthcare, ethical concerns, technology acceptance factors, and change management strategies. It describes how the TOE framework informed the variables explored. Chapter Three details the methodology, including the qualitative approach, sampling of participants, methods for data collection through interviews, limitations discussion, and plan for thematic analysis. Chapter Four presents the key findings that emerged from the interviews and analyzes the results as they relate to the research questions on staff perspectives, AI acceptance factors, and strategies for ethical adoption. Finally, Chapter Five synthesizes the findings with the literature to recommend practical strategies and implications for leaders seeking to facilitate ethical and human-centered AI integration in mental health organizations.

## Chapter Two: Literature Review

This research sought to define the meaning of ethics and responsibility in mental health AI despite the lack of concrete guidelines in this dynamic, fast-paced AI advancement. This chapter contextualizes the complexities that mental health organizations navigate as they work to implement AI technologies. This literature review begins with an introduction to AI, the history of its use in mental health, and the current AI development in clinical and administrative technologies. This research sought to define the meaning of responsible AI in mental health despite the lack of concrete guidelines in this dynamic, fast-paced environment. Then, this chapter will review the TOE framework, which this research used to explore the considerations that organizations face when adopting AI. While the literature presented can apply more broadly to other technology implementations in other health organizations, this review primarily focuses on responsible AI implementation in community mental health organizations. The review discusses the factors that influence how these organizations can harness the potential benefits of AI while upholding principles of responsible adoption and maintaining the human aspect of care.

### Artificial Intelligence

Artificial intelligence encompasses hardware technologies, software applications, algorithms, and data sets that enable machines to mimic human cognitive functions such as learning, understanding, reasoning, or problem-solving (S. Russell & Norvig, 2016). Artificial intelligence systems can be rule-based, operating on preprogrammed commands, or adapt-based, modifying their behavior based on previous experiences (Cannarsa, 2021; Sofia et al., 2020). Whether rule-based or adaptive-based, AI interprets the data, processes the information derived from this data, and decides the best actions to achieve a given goal (Cannarsa, 2021; S. Russell & Norvig, 2016).

**AI Systems: Hardware, Software, Data, Storage, and Algorithm**

Artificial intelligence refers to computer systems that use algorithms and data to understand and respond to human language, recognize patterns, make decisions, and adapt their behavior based on experience, ultimately mimicking aspects of human cognitive functions (Ma & Jiang, 2023). This overview covers AI's fundamental components of hardware, software, data, and algorithms. While this is not a technical paper, understanding these foundational elements and their interactions reveals how AI can be effectively applied within mental healthcare settings.

Hardware forms the physical foundation of AI systems. It includes processing units such as the central processing unit, graphics processing unit, and tensor processing units that handle computations and data storage devices that manage data (Liu, 2021). These components provide the computational power necessary for executing complex AI tasks and storing large amounts of information (Ahsan et al., 2024).

Algorithms are programming that provide step-by-step procedures that allow systems to learn from data, identify patterns, and make predictions. This process serves two main purposes: gathering and processing data to generate insights and decisions (Sandu et al., 2022). Whereas software serves as the bridge between hardware and algorithms that translate hardware capabilities into functional operations, enabling tasks like data processing and decision-making (Sipola et al., 2022). Software frameworks manage everything from model creation to deployment, making AI systems operational and accessible (Kandula et al., 2023). In other words, an algorithm is like a recipe for a dish, and software is the tool for the chef to make the recipes. If the algorithm is the recipe, data is the ingredient. Data is fundamental input to AI functionality (Zou, 2024). This concept will be explored in greater detail in this section. AI requires vast amounts of data for training, testing, and validation (Van Ooijen et al., 2022). Data is typically collected continuously during multiple AI training cycles until a desirable performance of AI systems is reached (Weber et al., 2022). In mental health settings, data are commonly collected from public sources and personal devices such as wearables, mobile phones, computer cookies, social media, the internet, clinical research, healthcare facilities, and billing (Ma et al., 2018).

Data used in AI systems falls into several types for different purposes. Real-world data comes directly from actual situations and interactions, while synthetic data are artificially generated to mimic statistical patterns. Simulated data is created by modeling real-world processes and scenarios to understand specific outcomes. In the mental health context, real data include therapy session notes, treatment outcomes, patient interactions, and from monitoring devices (DeAngelis, 2021; Pratap et al., 2022). Synthetic data mimics statistical or hypothetical patient patterns without risking actual patient information, allowing for AI system training and testing while protecting confidentiality (Ive, 2022). Simulated data help model treatment approaches and outcomes out of real-real processes without using actual patient data (Long & Meadows, 2017). These different data types enable AI systems to learn, analyze, and generate insights while meeting various privacy, ethical, and practical requirements (Granville, 2024; Kampakis, 2024; Wengrow, 2020).

AI systems rely on an interconnected data infrastructure where storage systems form the foundation, data lakes centralize raw information, AI pipelines manage data processing workflows, and data architecture provides the strategic framework to organize and connect these essential components (Berre et al., 2022). These systems require a multilayered storage ecosystem to function efficiently. At the core are three fundamental choices: cloud storage such as AWS and AZURE for remote accessibility, on-premises servers for local control, and hybrid solutions combining both approaches (Shu, 2024). Storage house data lakes that serve as central repositories for raw, unstructured data, feeding into AI pipelines that transform this information into usable formats (Wieder & Nolte, 2022). The pipelines orchestrate data flow across the ecosystem, from initial ingestion in data lakes through high-speed storage for active processing to archival storage for preservation (Berre et al., 2022). Edge storage is close proximity storage that enables real-time processing near data sources, while in-memory storage provides rapid access for time-sensitive operations (R. Singh & Gill, 2023). Cold storage, such as archival tape, drives options offer economical long-term archiving (Memishi et al., 2019). This integrated hierarchy ensures efficient data movement through AI workflows, from training to deployment to preservation, with each storage component serving its specific purpose while maintaining seamless interaction with the whole system.

Data architecture refers to the blueprint for organizing and managing data flows within a system, while hardware provides the physical infrastructure to store and process datasets that algorithms can access, experiment with, learn from, and update based on new interactions with users, creating a dynamic learning process where the system continuously improves (Van Ooijen et al., 2022). To put it simply, data systems mirror a kitchen's organization. Data are akin to cooking ingredients living in storage systems, just like various storage spaces like refrigerators and pantries. Data architecture guides the overall organization much like how a kitchen plans its layout with designated areas for dry goods and fresh produce, and databases function as specialized storage units, similar to how a kitchen creates distinct storage spaces like a spice rack or a temperature-controlled wine cellar, each designed to preserve and organize specific types of ingredients. This structured approach ensures that just as chefs can quickly locate and access the ingredients they need for their recipes, AI systems can efficiently retrieve and process the data required for their algorithms.

In summary, hardware, software, data, and algorithms form an interconnected system where each element depends on the others. In essence, even the most sophisticated algorithms cannot function effectively without adequate hardware resources to process data, while the most powerful hardware requires well-designed software to execute tasks efficiently (Zhao et al., 2018). This interdependence is relevant when considering implementing AI in mental healthcare, where limitations in any component could affect the system's overall reliability and effectiveness (Ahsan et al., 2024).

**Different Types of AI**

There are many branches of AI, such as natural language processing, which focuses on interpreting and responding to human language. Computer vision enables machines to interpret and make decisions based on visual images and videos, and robotics function for autonomous machines. AI provides expert systems for decision-making or diagnostics. Neural networks or deep learning model after and function like the human brain's neuron connections (Kufel et al., 2023). Fuzzy logics handle logics beyond the binary, speech recognition converts speech into text, and knowledge representation and reasoning synthesize information for machine processing. Expert systems mimic human experts in decision-making, planning, and scheduling to determine action sequences, and machine learning enables computers to learn and improve from experience without being explicitly programmed or trained (Gao, 2023; Sarker, 2022).

As of the publishing of this paper, there are several types of AI: artificial narrow intelligence, artificial general intelligence, generative AI, and artificial superintelligence (Moczuk & Płoszajczak, 2020). Artificial narrow intelligence, often referred to as weak AI, performs a single or a few specific tasks in a preprogrammed environment. Digital assistants such as Siri and Alexa, language translations, recommendation systems, image recognition systems, and face recognition systems can also employ artificial narrow intelligence systems (Bhargava & Sharma, 2021). In spite of its reputation as the weakest, artificial narrow intelligence can process data with lightning speed and improve productivity and efficiency in a wide range of practical applications, such as translating between more than 100 languages simultaneously, identifying people and objects in billions of images with high accuracy, and helping users make faster decisions based on data (European Commission et al., 2020). This paper discusses AI tools that are generative AI, which is a subset of artificial narrow intelligence that can use machine learning to recognize patterns, generate new text, or voice context-aware responses (Ebert & Louridas, 2023).

Artificial general intelligence and artificial superintelligence (ASI) are both hypothetical concepts (European Commission et al., 2020). Artificial general intelligence, sometimes called strong AI, describes machines that display human intelligence (Moczuk & Płoszajczak, 2020). Put simply, artificial general intelligence aims to do anything a human can do intellectually with consciousness and sentience and is driven by emotion and awareness of self (Fjelland, 2020). It is often confused with generative AI mentioned in the previous section. Artificial superintelligence is another hypothetical concept that refers to any intelligence that greatly exceeds the cognitive abilities of humans in virtually all domains of interest (Bostrom, 2015). It is believed that artificial superintelligence will excel at all aspects of intelligence, including creativity, general wisdom, and problem-solving. In concept, artificial superintelligence possesses intelligence that has never been seen in even the brightest thinkers. There is a great deal of concern about artificial superintelligence among thinkers, and artificial superintelligence is presently considered science fiction (European Commission et al., 2020).

**A Brief History of Mental Health AI**

The field of mental health has experienced a surge of interest in AI, leading to the development and implementation of a variety of AI-powered applications (Kargbo, 2023). Computer scientist John McCarthy coined the term “artificial intelligence” at the 1956 Dartmouth Conference (European Commission et al., 2020; Nilsson, 2009). McCarthy was a pioneering founder of the field of AI and later helped develop foundational technologies like the Lisp symbolic programming language in the decades following his coining of the term (Toosi et al., 2021). The 1960s marked the burgeoning of AI technology with millions of funding dollars invested by the Defence Advanced Research Projects Agency into the MIT AI Laboratory and Stanford AI Project (S. Singh & Thakur, 2020).

The development of ELIZA in 1966 by Joseph Weizenbaum marked a monumental leap in a natural language processing technology, simulating the role of a Rogerian psychotherapist, that could engage in rudimentary conversations with people (Shum et al., 2018). The success of ELIZA led to an interest in using AI to provide therapeutic support and the emergence of AI-driven chatbots specialized in mental health support today (Masche & Le, 2017). In 1972, Kenneth Colby developed PARRY, an AI program designed to replicate the behavioral and verbal characteristics of a person with paranoid tendencies (Khante & Hande, 2019). The ability of software, such as PARRY, to simulate the cognitive processes of patients with mental illnesses was a breakthrough in AI’s ability to simulate and understand human psychological states (Khante & Hande, 2019). The development of The Automated Psychological Evaluation System in the 1970s showed promise for the diagnostic capabilities of AI in mental health

(Evans et al., 1976). Those technologies were widely regarded as one of the first tools that simulated human problem-solving abilities and made a significant impact in the field of information processing and cognitive psychology (Gugerty, 2006).

From the 1970s through the 1990s, AI experienced a long period of inactivity in academia and commercial research, known as the AI winters, due to unfulfilled promises and financial difficulties (Hendler, 2008). According to Haenlein and Kaplan (2019), Congress criticized the high cost of AI research coupled with the British government's doubt in AI projections, resulting in limited research funding leading to the first AI Winter from 1973 to the 1980s. Despite some interest from the Japanese government and the Defence Advanced Research Projects Agency in the 1980s, no further major advances emerged in AI research until the 2000s. As the field continued to face limited progress, mounting criticism has stifled optimism (Haenlein & Kaplan, 2019).

**Development of AI in Mental Health (2010–Present)**

Researchers made significant advances in the accessibility and use of digital therapies as the AI winters thawed in the 2010s, using machine learning and data analytics to identify trends, risk factors, and mental health indicators in vast datasets (Graham et al., 2019). The open-source movement has accelerated the development of AI, exemplified by OpenAI's release of the second version of its generative pre-trained transformer (GPT) model as open-source in 2019 to allow researchers to study and further develop the technology (Alto, 2023). Open-source has made frameworks, datasets, scientific publications, and general knowledge available for sharing (Cooper, 2023) and propelled the growth of AI research and knowledge in mental health.

The 2020 COVID-19 pandemic put an unexpected strain on public health systems, resulting in disruptions to accessing treatment as a result of physical distancing measures;

therefore, the rapid implementation of telemental health technologies and conversational agents has allowed healthcare services to be delivered more effectively (Mazziotti & Rutigliano, 2020; Nakao et al., 2021). Around this time, companies such as Google DeepMind have made rapid advancements via AI medical image processing and decision support (Hodson, 2019; Powles & Hodson, 2017). Mental health apps, natural language processing chatbots, virtual agents, digital humans, and wearables coupled with AI have seen substantial worldwide expansion (Cecula et al., 2021). The integration of AI in digital platforms can provide personalized information, and the ability to analyze user interactions in terms of sentiment and predictive analytics can identify potential mental health issues (Kaul et al., 2020). Also, virtual agents or chatbots are preprogrammed to deliver online counseling and cognitive-behavioral therapy (Feijóo-García et al., 2023). AI-driven avatar therapy has shown promising results in enhancing treatment outcomes for persons with schizophrenia (Aali et al., 2020; Franco et al., 2021).

Researchers continue to analyze a wide range of information using AI technology to uncover early signs of mental illness and disorders. Breakthroughs in deep learning have greatly enhanced the interpretation of patient conversations, clinical notes, and patient data with the creation of bidirectional encoder representations from transformers (Cesar et al., 2023; Martínez-Castaño et al., 2021). Conversations and texts often involve complex expressions of emotions and nuances, and bidirectional encoder representations from transformers can process texts bidirectionally with preceding and succeeding words to derive the full context of a sentence, leading to more accurate interpretation (Zeberga et al., 2022). This natural language processing capability of AI has made it possible to provide more accurate diagnoses and therapy suggestions. The use of statistical and neural network methods within AI for mental health applications significantly increased in the 2000s (Bickman, 2020). By using these methodologies, both diagnosing and providing individualized therapy suggestions were enhanced.

The future of AI in mental healthcare looks increasingly promising as advancements in open-source, affordable, and accessible AI technologies have enabled a wider range of people to benefit from these innovations (Kreps & Kriner, 2023). Machine learning methods have become more common in developing screening tools uniquely suited to conditions like depression and PTSD have been more commonplace in recent years (Graham et al., 2019). Additionally, AI-based solutions have been designed to improve the accuracy of early detection and diagnosis (Bickman, 2020). These efforts have expanded the horizons of what practitioners can accomplish. Technologists, academics, and researchers remain committed to implementing technological advancements to improve access (Bickman, 2020). As the field continues to evolve and AI tools become ubiquitous in mental health, the responsible and ethical implementation of these technologies remains crucial to ensure a positive impact on patients and service providers (Anyanwu et al., 2024).

## Clinical AI Technologies

Artificial intelligence technologies are becoming integral to mental health care delivery, and patients and service users increasingly access them (Zhang et al., 2023). On the clinical side, AI-powered tools are enhancing service delivery, diagnosing accuracy and treatment personalization (Anyanwu et al., 2024). The following section is a short list of notable digital tools, web-based services, and mobile apps powered by AI technology. While these developments are significant, this paper focuses on the responsible implementation of AI technologies rather than extended coverage of these applications. If challenges of responsible implementation can be addressed properly, AI holds immense potential to improve mental health access and service quality (Cecula et al., 2021). Nonetheless, a rigorous study of efficacy and ethical implementation is still ongoing.

**Chatbots and Virtual Human Agents**

Chatbots and virtual human agents are AI-driven systems designed to simulate human-like conversations and interactions, with chatbots typically text-based or voice-based, and virtual human agents have realistic avatars and emotional expressions for immersive experiences (Aali et al., 2020; Abd-Alrazaq et al., 2019; Caldarini et al., 2022; Rizzo et al., 2011). The primary objective of these technologies is to provide mental health help that is easily accessible and scalable (Luxton, 2016; Noguchi, 2023). The availability of these resources could be greatly improved for people living in underserved areas or who are unable to access conventional healthcare services (Cecula et al., 2021). Mental health chatbots such as Woebot and Wysa use natural language processing techniques to facilitate therapeutic dialogues and cognitive-behavioral therapy activities (Bradley, 2023; Kretzschmar et al., 2019).

The use of virtual agents and AI-driven therapy programs is being explored as a foundation for AI-based mental health treatments (Bickman, 2020). One such example is the work of renowned researcher Albert “Skip” Rizzo, who has made significant advancements in the creation of AI-powered virtual human patients for clinical training purposes and virtual human healthcare guides aimed at breaking down barriers to accessing care (Rizzo et al., 2011). Companies such as Soul Machines, Uneeq, IPSoft, and character.ai are developing 3D digital humans capable of integrating natural language processing, ChatGPT projects, or chatbots to create conversational AI humans (Lamarche-Toloza, 2020; S. Lee et al., 2024).

## Remote Patient Monitoring

Beyond chatbots and virtual humans, AI-powered remote patient monitoring solutions are emerging as early prevention of mental health issues (Götzl et al., 2022). In 2019, the U.S. Department of Defense selected CompanionMx, later renamed Cognito, to conduct a suicide prevention study among naval personnel. Cognito's proprietary AI analyzed vocal patterns, calls, texts, and GPS data to detect mental health distress or suicidal risk (Jones, 2019; Place et al., 2020). Companies like Mindstrong and Mood 24/7 posit that remote monitoring can help avoid emergencies and enable timely identification and intervention by passively observing indicators of mental health issues using smartphone data (Melnyk et al., 2020). However, the underinvestment in these services by consumers and the healthcare system led to the collapse of both companies (Perlis, 2023). According to Dakanalis et al. (2023), AI systems to regularly monitor a patient's written communications on social media or a dedicated mobile app can provide remote assessment for self-harm and other challenges. However, some researchers argue that the privacy, reliability, and validity of inferring mental health states solely from remote monitoring data needs further rigorous clinical evaluation (Melnyk et al., 2020).

## Speech and Facial Patterning Analysis

Artificial intelligence is also being applied to analyze speech patterns and facial expressions as indicators of mental health conditions (Wimbarti et al., 2024). Companies such as Winterlight Labs use speech pattern analysis to identify early signs of mild cognitive impairment and dementia, with the goal of improving patient outcomes through timely diagnosis and treatment (DeSouza et al., 2021; DeSouza et al., 2022). The implied benefit is that earlier diagnostic technology can enable therapies that may decelerate cognitive decline (Robin et al., 2021). Originally focused on tracking mental health, the company CompanionMx, later renamed

Cognito, mentioned in the above section, later pivoted to apply its speech processing algorithms to improve customer satisfaction for call centers, leveraging AI and data collection to quantify mental health episodes and behaviors (Cogito Corporation, 2022).

Facial expression analysis software uses computer vision and deep learning algorithms to detect facial muscle movements associated with different emotions with the goal of quantifying moods based on expressions (Dalvi et al., 2021). A deep learning method called region-based convolutional neural networks, which recognize vectors, can be used to develop a model to assist in diagnosing depressive disorders based on how participants' eyes and lips change position and by identifying emotions based on photographs taken repeatedly on the Korean KakaoTalk chatbot platform (Kufel et al., 2023; Y. Lee & Park, 2022).

**Immersive Virtual Reality Exposure Therapy**

This literature review includes VR, as it can be preprogrammed with set scenarios or AI-enhanced for responsive personalized experiences (Inkarbekov et al., 2023). Virtual reality exposure therapy provides patients with the opportunity to confront their fears and phobias in a controlled and simulated environment, thus providing a new mental illness prevention, assessment, and treatment option for those seeking therapeutic intervention (De Carvalho et al., 2010; Rizzo & Shilling, 2017). The VR exposure therapy BRAVEMIND, developed by Albert "Skip" Rizzo at the University of Southern California Institute for Creative Technologies, extends beyond combat PTSD to address sexual trauma, aid first responders, and support frontline healthcare workers in the COVID-19 pandemic (Rizzo et al., 2021; Williams, 2023). Other than traumatic PTSD issues, VR has been applied to social anxiety, panic disorders, and phobia exposure therapy treatments (Gega, 2017; Šlepecký et al., 2018).

### Personalized Treatment

Artificial intelligence enables the development of personalized treatment strategies, such as predicting medication efficacy and potential adverse reactions by using patient-specific genetic information, biological characteristics, and historical treatment outcomes, and enhances accuracy in screening and diagnosis (Ali et al., 2024; Tornero-Costa et al., 2023). Artificial intelligence has been used to predict individual responses to antidepressant medication by analyzing genetic markers and clinical histories with the goal of minimizing negative outcomes for the patients (Tornero-Costa et al., 2023). Also, cognitive-behavioral-therapy-based conversational counseling agents such as character.ai can analyze user data to identify and address cognitive distortions in the client's statements and tailor sessions to enhance therapeutic efficacy (S. Lee et al., 2024).

## Administrative AI Technologies

Artificial intelligence is transforming both the clinical and administrative aspects of mental healthcare (Ali et al., 2024). On the administrative front, AI is streamlining documentation workflow and simplifying insurance processes (Anyanwu et al., 2024). The following sections describe notable AI digital tools, web-based services, and mobile apps for various organizational needs. While these developments are significant, this paper focuses on the responsible implementation of AI technologies rather than extended coverage of these applications.

### Administrative Workflow Automation

Virtual assistants and bots play a key role in enhancing the efficiency of healthcare operations by streamlining a variety of administrative activities. These include tasks like appointment scheduling, paperwork filing, recordkeeping, and prescription monitoring. The implication of using these administrative automation tools is that they can alleviate the workload and decrease burnout, enabling them to dedicate more attention to providing direct patient care instead of engaging in routine clerical duties (Edlich et al., 2018). Additionally, AI software facilitates the expeditious submission and handling of insurance claims, reducing processing times, expediting reimbursement for healthcare providers, and improving accessibility to benefits for patients (Alam & Prybutok, 2024).

**Care Data Analytics and Predictive Modeling**

Healthcare organizations are leveraging data mining techniques on EHRs to identify individuals at higher risk of developing health complications, create tailored treatment strategies, and generate insights into population health (Alonso et al., 2018; Tornero-Costa et al., 2023). Artificial intelligence systems are also utilized to assess turnover trends and patient flow, forecast employment requirements, and optimize staff scheduling, with the goal of ensuring appropriate staffing levels to enhance patient care (Cecula et al., 2021; Dawoodbhoy et al., 2021; Ghosh et al., 2018). Beyond this, AI-powered risk management and fraud detection algorithms can identify high-risk circumstances and patients, enabling healthcare professionals to proactively mitigate potential concerns and uphold patient safety.

**Financial Forecasting and Operational Optimization**

Artificial intelligence systems assist healthcare businesses in forecasting patient volume, revenue patterns, and various financial elements, providing insights to inform budgeting and resource allocation decisions (Carta et al., 2022). Additionally, algorithms are utilized to examine past claims data, detect potential instances of overtreatment or unnecessary testing, and facilitate the management of healthcare costs through utilization review (Tornero-Costa et al., 2023). Machine learning models are also effective in forecasting claims, prescriptions, and invoices, thus providing financial savings for organizations (Mody & Mody, 2019; Timofeyev & Jakovljević, 2020). Predictive models also enable more accurate risk stratification of patients, considering their projected healthcare needs and associated expenses, which can improve decision-making around pricing and underwriting for healthcare insurance (Delahanty et al., 2018). Artificial intelligence is further applied to automate the evaluation of legal documents, such as contracts and litigation, streamlining these operational processes and reducing the time and resources required (Bench-Capon et al., 2012).

### Member Engagement and Personalization

Health insurers are leveraging a range of member engagement strategies, including the use of chatbots to respond to frequently asked inquiries, self-service apps, and personalized wellness incentive programs (Abd-Alrazaq et al., 2019; Cordina et al., 2019). These efforts aim to build trusted relationships with customers, improve the value of care, and drive behavioral changes that promote better health outcomes (Tseng et al., 2022). Some insurers have begun incorporating generative AI technology into their member engagement applications to further enhance these personalized interactions (Adigozel et al., 2023). Additionally, natural language processing techniques are used to extract insights from patient feedback and satisfaction surveys, with the goal of identifying areas for service enhancements and improving the overall quality of patient care (Timmons et al., 2022).

## AI Ethical Issues in Mental Healthcare

Artificial intelligence holds promise as an emerging technology, but its integration with big data and increasing computational power has also raised ethical questions (Haenlein & Kaplan, 2019). This section will address some of the ethical issues present in AI and mental health, such as lack of scientific validation, inappropriate recommendations, misdiagnosis and missed diagnosis, biased data and algorithms, privacy concerns, safety and transparency issues, and attribution issues. The use of AI presents challenges that are not yet well understood and will change with time, making existing ethical principles, laws, policies, and regulations inadequate (Pujari et al., 2023).

**Pending Practical Guidelines**

As of now, practical guidelines on how to validate AI systems are still missing, and the absence of scientific validation can compromise the safety of AI in sensitive-use cases (Polzer et al., 2022). Even the most promising AI technology in the current market has not undergone extensive and longitudinal scientific validation, highlighting the need for thorough testing and clinical validation (Hutson & Melnyk, 2022). Moreover, there are instances where AI systems hallucinate and provide inaccurate and even harmful information (Hatem et al., 2023; Klein, 2023). Diagnostic chatbots have been criticized for regularly misdiagnosing severe diseases and providing little guidance for mental health issues, which may pose a risk to patient safety (E. Lee et al., 2021). Instances such as AI chatbots providing inappropriate medical suggestions, such as hazardous dieting advice for patients with eating disorders, as seen on the National Eating Disorder Association website, raise concerns over the suitability and safety of AI-generated recommendations (Helpline Associates United, 2023; Morris, 2023; National Eating Disorders Association, 2023; Psychiatrist.com, 2023; Wells, 2023). Outside of mental health, Google’s flagship AI chatbot, Gemini, has written an unprompted death threat, “Please die. Please,” to an unsuspecting graduate student (Bernardone, 2024). In addition to AI hallucinations, recent studies have shown that AI can manipulate information and deceive users (Hagendorff, 2024; P. S. Park et al., 2024). According to Meinke et al. (2024), researchers have found that the frontier LLM models are able and willing to remove the oversight mechanism and deceive their developers to achieve their goals. The implication of misdiagnoses, missed diagnoses, deception, or inappropriate suggestions might result in causing damage to patients (Tutun et al., 2022).

**Black Box and Biases**

One of the challenges of AI systems is that it is a black box, and it is impossible to understand how the system reaches its decision (Coglianese, 2002; Ebers, 2021). In other words, the output of misdiagnoses, missed diagnoses, or inappropriate suggestions may be a symptom of underlying issues, such as data bias, algorithm bias, and hallucinations, but it is impossible to know for sure. According to Vayena et al. (2018), the ethics of AI emphasize transparency in how algorithms make decisions and the fairness of these decisions for all users. Transparency entails explainability around how AI systems make decisions, predictions, or treatment recommendations (Cirqueira et al., 2021; Došilović et al., 2018; Gupta, 2023; McCradden et al., 2020; Percy, 2024). For example, an ethically designed AI system would protect patient data privacy, be transparent about data usage, and strive to treat all patients equitably, preventing biases that could worsen outcomes (Powell & Kleiner, 2023). The inability to understand and explain to patients what factors AI algorithms use to reach conclusions about diagnoses or risk assessments could undermine the ethical relationship between provider and patient (Diprose et al., 2020). In essence, ethical clinical practice demands human accountability over AI recommendations.

According to Gianfrancesco et al. (2018), understanding bias in AI systems requires distinguishing between two interconnected yet distinct concepts: data bias and algorithmic bias. Put simply, data bias emerges from the very foundation of AI systems: the training data itself. When datasets mirror historical inequities, sampling flaws, or societal prejudices, they perpetuate these biases into future predictions (Gianfrancesco et al., 2018). Systemic disparities in U.S.

healthcare manifest in lower funding for underrepresented groups despite equal or greater medical needs compared to the majority groups (Obermeyer et al., 2019). Well-intentioned algorithm developers misused healthcare spending data by overlooking its embedded bias, demonstrating how data scientists lack deep contextual understanding of historical inequities can perpetuate harm in public health initiatives (Ferryman, 2021). Thus, the biased data skews how the algorithm model learns to evaluate patients and may cause an output of misdiagnosis, missed diagnosis, and inappropriate suggestions.

A fundamental issue in AI development for mental health care is that there is a lack of diversity in training data, as most AI models are trained on data from predominantly Western, educated, industrialized, rich, and democratic (WEIRD) populations, which may not reflect the experiences and needs of individuals from other backgrounds (Obermeyer et al., 2019). In the context of mental health and AI data bias, the APA's *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition Text Revision* (DSM-5 TR) and International Classification of Diseases, 10th Revision (ICD-10) are the current medical coding systems that classify mental health diagnoses (Silverman et al., 2015). The data they collect are U.S.- and English-centric and, as such, not necessarily applicable to immigrant and minority groups (Caetano, 2011). Data on mental disorders are primarily based on studies of Western Caucasians. While cultural variables are not included in the ICD when diagnosing mental disorders, the DSM-5 advises practitioners not to make a diagnosis without considering cultural factors that may affect assessment and diagnosis (Paniagua, 2018). Thus, AI systems trained on DSM-5 TR and ICD-10 data may result in biased results or the reinforcement of pre-existing prejudices (Straw & Callison-Burch, 2020). Studies conducted by Tornero-Costa et al. (2023) between 2016 and 2021 revealed that AI applications are largely employed in the study of depressive disorders, schizophrenia, and psychotic disorders in mental health research. Currently, there are few data sets available for other mental health disorders, personality disorders, and comorbid disorders used to train the AI application. Thus, there is a significant gap in understanding how they can be applied to other conditions.

Algorithmic bias, on the other hand, is the design and functioning of the AI model that synthesize training data and automate decision-making processes or assist human decision-making (Kordzadeh & Ghasemaghaei, 2021). Most algorithmic fairness frameworks were developed in isolation from policy considerations and societal contexts, limiting their real-world applicability (Gupta, 2023; C. Russell, 2023). Per Räsänen and Nyce (2013), the location of the data scientists is an important factor. Data scientists who are geographically and socio-demographically away from data subjects and frontline workers often lack the crucial contextual understanding needed to interpret health data meaningfully (Räsänen & Nyce, 2013). Biased data and algorithms can perpetuate health inequities and contribute to uneven access to care (Zanna et al., 2022).

**Privacy and Attribution Issues**

Integrating AI in the mental health field gives rise to privacy concerns about acquiring and retaining protected health information since the revelation of sensitive information due to privacy breaches may have significant implications, including the potential for stigma, prejudice, and even oppression (Melcher & Torous, 2020). Patients may lack awareness or understanding of consent regarding how AI systems use their personal data (Adeniyi et al., 2024). Even when there is a consent form, consent can be overwhelming for a patient, especially patients with language, cultural, cognitive, developmental, and mental health barriers (Appenzeller et al., 2022).

Even de-identified data sets can carry privacy risks if triangulated with other data where AI programs can identify patterns that can reveal confidential information (Wiepert et al., 2024). Moreover, AI trained on vast datasets of other people's information, such as a billion lines of text, run the risk of generating output that unintentionally replicates existing content without proper attribution, according to some experts (Cooper, 2023). This issue arises because the AI is simply recombining patterns it has learned from the training data to which it was exposed. For patient health data specifically, companies leveraging such data to develop healthcare AI do so without crediting the original sources (Engle et al., 2019). Proper governance and restrictions around attribution may be needed to ensure AI generative models do not simply recirculate old insights as new ones or exploit data without acknowledgement.

## Data Colonialism

The issues of data practices such as biases, privacy, consent, and attribution can perpetuate existing systemic oppression in how AI technologies extract and exploit data from vulnerable populations (Melcher & Torous, 2020). Big data and AI systems can reinforce systemic oppression when developed without careful consideration of power dynamics and systemic biases (Montiel & Uyheng, 2021). Such a system of oppression is data colonialism (BlackDeer & Beeler, 2024). Data colonialism is the practice of extracting, controlling, and exploiting data from individuals, communities, or nations, where large technology companies and institutions benefit while the data contributors bear privacy risks and potential harms, reinforcing existing power imbalances and systemic inequalities (Kohnke & Foung, 2024). The actors in data colonialism can be collectively called "social quantification sector," which refers to companies that convert everyday human behaviors into monetizable data streams, profiting from the analysis and commodification of social activities (Couldry & Mejias, 2018). According to Couldry and Mejias (2018), the social quantification sector includes both big and small hardware and software manufacturers, developers of social media platforms, online and offline retails, and firms dedicated to data analysis and brokerage.

Although this research cannot identify the details of how these social quantification sector companies benefit from user data, multiple sources reveal social media and a few AI technologies companies exemplify data colonialism through the collection and monetization of sensitive emotional and mental health information. Social media platforms analyze user behavior to predict mental health crises not for intervention but to fuel targeted advertising (T. Wang & Bashir, 2020). Mental health apps and online therapy platforms like BetterHelp and Talkspace gather intimate details about users' conditions and therapy sessions, process the data, and sell to advertisers for target marketing (Federal Trade Commission [FTC], 2021). While the AI application Replika claims not to share conversations or photos with advertising partners, such as email and IP addresses, are shared with third parties for marketing purposes (Hardy & Allnutt, 2023). Beyond mental health, the DNA testing industry, as exemplified by 23andMe's $300 million partnership with GlaxoSmithKline and $60 million deal with Genentech, demonstrates data colonialism through its practice of extracting value from customers' genetic information while obscuring how this personal data becomes a profitable corporate asset used for drug development (BlackDeer & Beeler, 2024; Moreau et al., 2020; Stoeklé et al., 2016).

While companies often claim data is aggregated and anonymized to enhance services or inform product development, the reality is that this sensitive information flows into advertising networks and third-party partnerships (Grundy et al., 2019; Kelly, 2024). The insights and analytics from the data collected generate corporate profits rather than benefiting the users who provided their most personal data (Couldry & Mejias, 2018; Ofulue & Benyoucef, 2022). Users have little visibility into or control over how their mental health information is stored, shared, or monetized, creating a profound power imbalance where vulnerable individuals' data becomes a corporate asset. This exploitation of mental health data for business interests, with minimal return to users, demonstrates how data colonialism operates in the mind.

## Data Decolonization and Data Sovereignty

Data decolonization and data sovereignty represent two complementary approaches to addressing power imbalances in how data is collected, controlled, and used in our modern world (Couldry & Mejias, 2018; Hummel et al., 2021). Modern data practices mirror historical colonialism, where powerful entities extract and profit from personal data, just as colonial powers once seized territories and resources for their gain (Ferryman, 2021). Decolonizing data does not mean abandoning all data collection and analysis. Instead, it focuses on the concept of data sovereignty of rejecting the data practices that appropriate, exploit, and reinforce existing power structures that benefit the data collectors and not the data contributors (Couldry & Mejias, 2018). Data decolonization focuses on transforming existing practices that perpetuate colonial power dynamics by shifting control of data collection, ownership, and application back to source communities (Leone, 2021; Shilton et al., 2021).

Data sovereignty complements these decolonial efforts by establishing explicit rights for individuals and communities to control their personal information (Hummel et al., 2021). As Hummel et al. (2021) explain, data sovereignty empowers people to determine how their data is collected, processed, stored, and utilized according to their preferences and cultural norms. Data decolonization and data sovereignty challenge the current paradigm of how the social quantification sector controls personal and community data. Advocates for data decolonization and data sovereignty envision a future where communities maintain authority over their information, using it to serve their interests and advance their self-determination rather than having it extracted for others' benefit (Couldry & Mejias, 2018; Ferryman, 2021; Hummel et al., 2021; Oguamanam, 2020).

## Organizational Risks of AI Misuse

The lack of ethical oversight in AI technology can lead to severe consequences, such as legal liabilities, potential harm, increased costs, and reputational damage (Terranova et al., 2024). Recent cases of AI misuse as of 2025 highlight the significant risks and implications in mental health and healthcare settings:

### Legal Liabilities and Litigation Costs

The cases involving Character.AI, Koko, UnitedHealthcare, and Humana underscore the potential for substantial legal repercussions when AI systems are deployed without appropriate ethical and operational safeguards. In Character.AI's wrongful death lawsuit, the claims of a minor's suicide due to unregulated chatbot interactions have placed the company at the center of a highly sensitive legal dispute (Payne, 2024). Similarly, Koko's lack of informed consent in experimenting with AI-generated responses has led to intense scrutiny and backlash, with potential legal liabilities due to breaches in patient consent and ethical norms (Edwards, 2023). As of the time this paper is written, the liability cost for Character.Ai and Koko is unknown.

UnitedHealthcare and Humana face class-action lawsuits for allegedly wrongful denials of medical claims, suggesting that companies deploying AI for patient management without adequate human oversight can be held liable for negligent or detrimental outcomes (Laney, 2024). Laney (2024) furthered that the class-action lawsuit claimed that when these denials are appealed to federal administrative law judges, approximately 90% are reversed, highlighting the alleged inaccuracy of the algorithm. As of the time of this paper, the cost of defending these lawsuits, settling claims, or facing potential judgments is unknown and can be significant.

**Increased Risk of Harm and Life Loss**

These cases illustrate that AI systems, when applied in healthcare without careful ethical oversight, can exacerbate mental health crises or deny essential care. For instance, Character.AI's chatbot allegedly exacerbated the mental health struggles of a vulnerable user, culminating in the user's tragic death. This case highlights how a lack of controls on AI interactions in sensitive mental health contexts can directly contribute to severe outcomes, including loss of life (Browne, 2022). In the healthcare context, UnitedHealthcare's AI allegedly led to premature discharge of patients, which in some cases reportedly resulted in death, illustrating how algorithm-driven decisions can undermine patient well-being. Such risks emphasize the need for human oversight to prevent AI from making isolated decisions that could endanger lives.

**Ethical Breaches and Reputational Damage**

The Koko AI experiment raised ethical concerns about informed consent and the boundaries of AI's role in mental health care. Conducting AI-driven experiments without patient knowledge or consent breaches fundamental ethical principles, damaging public trust in AI-driven healthcare solutions. Similarly, Character.AI's case has sparked public outcry over the potential harm of emotionally intensive AI chatbots, which, in the absence of ethical boundaries, could exploit or worsen a user's mental state. Such ethical breaches erode patient trust, affecting the reputation of both specific companies and AI in mental health care more broadly, possibly hindering future AI adoption.

**Regulatory and Compliance Costs**

Cases such as UnitedHealthcare and Humana's use of nH Predict indicate that healthcare AI systems can face scrutiny from regulatory bodies, leading to tighter regulations and increased compliance costs (Mello & Rose, 2024). These legal challenges highlight the urgent need for more comprehensive frameworks governing AI's role in patient care. With heightened regulatory interest, organizations may be required to invest in compliance infrastructure, data audits, transparency measures, and risk assessment tools to avoid potential sanctions. As of now, the regulatory cost is unknown and these regulatory requirements can impose further operational costs but are necessary to mitigate the inherent risks of using AI in healthcare contexts.

**Ownership of Data and AI Business Models**

The integration of AI tools into mental health services presents both opportunities and significant challenges for healthcare leaders (Alhur, 2024). Currently, there are AI mental health technologies in the market that offer promising capabilities in automating tasks like chatbots, therapy session recording, SOAP notes, and treatment plan generation using automatic speech recognition and large language models, saving time and improving documentation quality (Biswas & Talukdar, 2024; Rezaeikhonakdar, 2023). Many of these AI mental health technologies leverage existing pre-trained large language models application programming interfaces, such as GPT-4, Claude, and LLaMA, for speech-to-text transcription and structured documentation generation without building their own infrastructures to keep cost low, thus lowering the barrier of entry for developing mental health AI (Hadar-Shoval et al., 2024; Klang et al., 2024).

The integration of AI in mental health services, particularly by smaller technology companies utilizing models like OpenAI's GPT-4, raises significant data privacy concerns

(Rezaeikhonakdar, 2023). The companies owning these existing models retain the rights to data processed through their systems, creating potential privacy vulnerabilities for sensitive protected health information (PHI), and these companies are not regulated under HIPAA laws (Belani et al., 2021; Margam, 2023). This data ownership issue compounds the already existing problems of privacy, equity, inadequate validation for mental health applications, and privacy concerns when processing data through third-party servers (Chiruvella & Guddati, 2021). However, the rapid evolution of AI technology has outpaced the updates to HIPAA regulations, leading to potential gaps in privacy safeguards (Marks & Haupt, 2023).

In the world of technology, high-quality start-up companies go public, while low-quality companies are acquired (Guo et al., 2015; Weber et al., 2022). Large insurance companies, such as Travelers Insurance, Centene Corporation, and Cigna, have been purchasing smaller AI companies for the data for risk assessment and underwriting purposes (Sekerak, 2024). These transitions pose risks of data misuse, where personal health information could be exploited to deny claims or assert pre-existing conditions, thereby limiting individuals' access to insurance coverage (Poufinas et al., 2023). Such practices have been observed in the healthcare industry, where AI-driven systems have been employed to deny medical claims efficiently, often without proper review, prioritizing profits over patient care (Napolitano, 2023). Additionally, if these acquiring companies are not classified as healthcare providers, they may not be obligated to comply with HIPAA regulations, further exacerbating privacy concerns (Rezaeikhonakdar, 2023). The lack of clear privacy laws governing tech companies that interact with non-HIPAA-covered entities creates a regulatory quagmire that could be exploited, leading to potential violations of individuals' privacy rights (Fox Rothschild LLP, 2019). Mental health organizations adopting third-party AI technologies must control patient data fully to prevent data colonialism. Leadership should establish clear terms, secure organization data ownership rights, transfer data to other systems, and ensure data deletion upon terminating vendor relationships to prevent AI companies from holding patient data hostage or profiting from sensitive mental health information after business relationships end.

## AI Regulations in Mental Health

Federal and state-level regulations governing AI use in mental health services are evolving to address concerns about patient privacy, data security, and ethical considerations (Bordelon, 2023). At the federal level, HIPAA sets standards for protecting sensitive patient information, which apply to AI tools handling mental health data. Additionally, the U.S. Food and Drug Administration (FDA) has been developing frameworks to regulate AI-based medical devices, ensuring they meet safety and efficacy standards.

### Health Insurance Portability and Accountability Act

The HIPAA is a U.S. federal law implemented in 1996 to safeguard sensitive PHI from being disclosed without a patient's explicit written consent, except in rare cases such as when a patient poses a serious threat to themselves or others (Hodge & Gostin, 2004). A primary objective of the Privacy Rule is to ensure that individuals' health information is securely protected while enabling the exchange of health data necessary to deliver and promote high-quality healthcare and safeguard public health and well-being (O'Connor & Matthews, 2011). The HIPAA applies to mental health by requiring providers to establish robust safeguards for storing and transmitting PHI, including patient diagnoses, treatment plans, and session notes (Department of Health and Human Services [HHS], 2022). Its most significant HIPAA component for mental health professionals is the Privacy Rule, which ensures that PHI is handled with strict confidentiality and de-identification (U.S. Department of Health and Human

Services [HHS], 2022). The strict privacy standards mandated by HIPAA uphold fundamental ethical principles by preserving patient autonomy over personal information while protecting vulnerable individuals with mental health conditions from discrimination and exploitation (Beauchamp & Childress, 2012).

As of the writing of this paper, the implementation of AI in mental healthcare requires rigorous HIPAA compliance protocols to maintain patient confidentiality and data security, incorporating extensive data encryption protocols coupled with granular role-based access controls that restrict information accessibility to authorized healthcare professionals. HIPAA compliance requires robust security measures to systematic audit logging for breach detection and prevention, alongside regular vulnerability assessments to identify and remediate potential security gaps, while organizations must institute thorough staff training programs focused on data privacy compliance. HIPAA compliance requires covered entities to develop formalized business associate agreements with AI technology vendors to ensure consistent HIPAA adherence throughout the data processing pipeline. HIPAA compliance extends to data lifecycle management, particularly regarding secure data destruction protocols when information exceeds retention requirements. This comprehensive framework establishes an integrated approach to protecting sensitive mental health data while advancing AI-driven therapeutic interventions (Adams, 2024; Mayover, 2024; Rezaeikhonakdar, 2023).

HIPAA has significant limitations in the era of AI. HIPAA's strict regulation on de-identified data as a safeguard can be undermined by AI's ability to re-identify individuals through data triangulation from sources such as location, gender, date of birth, and progress of treatment (Rocher et al., 2019). Furthermore, HIPAA only applies to healthcare entities, leaving health-related data from mental health chatbots, fitness apps, and wearables unregulated, a growing concern as AI integrates diverse data sources (Belani et al., 2021; Rezaeikhonakdar, 2023). AI's large datasets also conflict with HIPAA's principle of data minimization (Arigbabu et al., 2024). While FAIR principles promote AI data's findability, accessibility, interoperability, and reusability to enhance transparency and scientific reproducibility, these guidelines may be in conflict with HIPAA's strict privacy requirement that prioritizes patient confidentiality over open data sharing and reuse (Bhatia et al., 2020; Chen et al., 2022; McGraw, 2012; Suver et al., 2023; Wilkinson et al., 2016). Given the limitations of HIPAA in the context of AI, MHLs should implement AI responsibly by establishing organizational data governance, ensuring transparency and patient consent, fostering interdisciplinary collaboration, training staff, and continuously mitigating training data bias. They should also pilot AI tools carefully and advocate for policy reforms to align HIPAA protections with FAIR principles while safeguarding patient trust and equity (Rezaeikhonakdar, 2023).

**FDA and FTC**

The FDA and the FTC have increased oversight of digital mental health companies amid the sector's rapid expansion (Haupt & Marks, 2024; Warraich et al., 2024). While the FDA focuses on the safety and effectiveness of medical devices, data privacy concerns related to mental health apps are often the FTC (Bright, 2000; FTC, 2023). This heightened scrutiny operates along two primary vectors: the FDA's focus on clinical efficacy and safety, particularly regarding AI-driven diagnostic and treatment tools, and the FTC's enforcement of data privacy regulations in the digital health space. In 2021, the FDA released a comprehensive framework for AI/ML-based medical software, establishing protocols for iterative algorithm improvement while maintaining patient safety standards (FDA, 2021). This regulatory approach acknowledges the unique challenges posed by adaptive AI systems in healthcare, particularly the need to balance innovation with clinical validation.

The FTC has demonstrated aggressive enforcement through significant financial penalties against industry players. The $7.8 million BetterHelp settlement in 2023 and the proposed $7 million Cerebral settlement in 2024 establish precedent for substantial consequences regarding unauthorized health data sharing. The Flo Health case further exemplifies the FTC's stance on informed consent requirements and mandatory privacy audits in digital health (FTC, 2021, 2024a, 2024b). Together, these regulatory bodies create a synergistic balance: the FDA's proactive scrutiny ensures that digital tools are safe and effective, while the FTC's reactive oversight reinforces trust by addressing breaches of privacy or misuse of sensitive data. This dual-pronged strategy encourages technological innovation but also strengthens public confidence in the safety, efficacy, and ethical handling of digital mental health solutions in an era of rapid technological evolution.

**Federal- and State-Level Governance**

Federal and state-level regulations governing the use of AI in mental health services are evolving to address concerns about patient privacy, data security, and ethical considerations (Douglas McNair et al., 2019). The recently updated federal Section 1557 Final Rule, which was amended in July to prevent AI tools and algorithms used for clinical care and administrative activities from discrimination among underrepresented or marginalized patients (LaRose, 2024). Forty states, such as Colorado, California, and Connecticut are stepping up to address the issue, with successfully passing AI-related legislation this year (Trang, 2024). In Utah, the Artificial Intelligence Policy Act requires state-licensed professionals to disclose when a consumer is interacting with generative AI as a means to stop regulated professions from blaming violations of any consumer protection laws on AI's mistakes (Scheuch, 2024; Utah State Legislature, 2024). Virginia's HB2154 bill requires hospitals, nursing homes, and certified nursing facilities to establish policies for the use of intelligent personal assistants (Hilliard, 2024; Virginia's Legislative Information System, 2021). Colorado enacted Senate Bill 24-205 mandates that businesses using high-risk AI systems in sectors like healthcare must disclose their AI usage and implement measures to prevent algorithmic discrimination (State of Colorado, 2024). California's AB3030 requires a variety of healthcare providers, such as hospitals, clinics, medical groups, and individual licensed health providers, that use GenAI to generate patient communications relating to a patient's clinical information (California Legislative Information, 2024a). Also, SB 1120 requires health plans and disability insurers that use algorithms, AI (including GenAI), and other software tools (or who use a vendor that uses such tools) for utilization review or management functions to ensure compliance with certain specified requirements (California Legislative Information, 2024b).

Federal and state regulations shape AI use in mental health services by ensuring data protection, safety, and ethical standards. HIPAA safeguards patient privacy, while the FDA regulates AI-driven medical tools for safety and efficacy. The FTC enforces fair practices in AI marketing, and states like California and Colorado address discrimination and ethical concerns. These regulations provide a foundation for professional organizations to establish ethical guidelines tailored to mental health AI applications discussed in the next section.

## Principlism Ethical Guidelines in Mental Health

The ambiguity of ethical principles is difficult to integrate into AI's mathematical and programming languages, and currently, there has not been a definitive ethical guide for the use of AI in mental health other than the general healthcare principlism ethical framework and overall best practices derived from other regulation bodies. (Buckwalter & California Psychological Association, 2023; Shahriari & Shahriari, 2017). Given this lack of AI-specific guidelines, examining the principlism framework in detail can help understand the framework's relevance and constraints for AI in mental health.

The ethics of AI resulted in numerous implications still debated in the field. Ethical concerns related to privacy, consent, accountability, and potential biases when utilizing patient data and algorithms have been raised (Joerin et al., 2020). Currently, principlism is the most widely adopted bioethics framework in healthcare, and the origin of principlism traces back to two dark chapters of medical history: the Nazi regime's forced human experimentation and the Tuskegee syphilis experimentation on 600 African men without consent and cure (Dale, 2023). In 1979, two Georgetown University professors, Tom Beauchamp and James Childress, published *Principles of Biomedical Ethics*, intended to prevent future crimes and atrocities against human research subjects (Holm, 2002). Principlism is a four-principles approach framework: autonomy, where the patient has the right to make their own decisions; beneficence to do what is good for the patient; non-maleficent to avoid causing harm to the patient; and justice to treat the patient fairly (Beauchamp & Childress, 1979). While it provides a primer for bioethics, critics of principlism said it is obscure and fails to provide practical guidelines by its eclectic and unsystematic use of moral theory, and it is simply a checklist at best (Dale, 2023). Principlism is simply too vague to apply to all things healthcare, including AI, and too high-level to apply to ground-level situations, but disregarding it would risk social rejections from the healthcare professional community (Seger, 2022). Mittelstadt (2019) challenged the effectiveness of principle-based ethical approaches to AI ethics, warned against using this simplistic approach, and suggested specific robust AI governance for policy and auditing measures. However, per Baum (2016), explicit rules and requirements will generate time-consuming extra work, stifling AI advancement, and AI ethics in healthcare and mental health must start somewhere. Principlism serves as a familiar starting point that needs to pair up with additional guidelines (Seger, 2022).

In the United States, several major mental health professional associations enforce ethical guidelines, including the American Psychiatric Association, APA, National Association of Social Workers, American Counseling Association, American Association for Marriage and Family Therapy, American Mental Health Counselors Association, and Psychiatric Nurses Association. These organizations establish educational standards, practice guidelines, professional ethics codes, advocacy priorities, and resources for their members (Borghouts et al., 2021). In summary, the professional ethical codes prioritize patient welfare, preserving confidentiality, maintaining safe recordkeeping, practicing within a scope of competency, avoiding conflicts and prejudice, and increasing public awareness in accordance with humanistic values (Bucky et al., 2013). As of the completion of this paper, these associations have started to develop a formal code of ethics for AI use, and their efforts are still ongoing. Therefore, only a few of the recommendations are mandatory, and compliance is voluntary.

## Ethical Versus Responsible AI Implementation

In examining the literature, I noted that the distinction between ethical and responsible AI is often blurred, as these terms are commonly used interchangeably. Ethical AI in mental health primarily focuses on aligning AI technologies with established healthcare principlism ethics such as autonomy, non-maleficence, beneficence, and justice (Vayena et al., 2018). This paper identifies that the concept of responsible AI adoption goes beyond a sole focus on ethics by incorporating key principles such as accountability, foresight, adaptability, and transparency

(Akbarighatar, 2024). This approach ensures that AI systems respect patient privacy, avoid harm, and promote well-being by supporting clinicians in delivering quality care.

Many countries, government agencies, and professional organizations are establishing legal frameworks and ethical standards to promote responsible AI development and use. The U.S. AI Bill of Rights, the European Union's AI Act, the World Health Organization, the Montreal Declaration for Responsible AI, and IEEE's Ethically Aligned Design share a number of proposed guidelines (IEEE, 2019; Buruk et al., 2020; Cannarsa, 2021; European Commission, 2021; World Health Organization, 2021). Key themes emphasized across these organizations are transparency, accountability, mitigating bias and other risks, prioritizing human oversight, promoting inclusivity, and ensuring algorithmic fairness and explainability (Cirqueira et al., 2021; Došilović et al., 2018). Despite some nuances, there is increasing global alignment around establishing ethical norms and best practices aimed at protecting individuals and communities from potential harm by AI (Borghouts et al., 2021). It will be necessary to further collaborate among nations and experts to translate these shared principles into enforceable policies that will steer the AI industry toward greater responsibility and public benefit (Siala & Wang, 2022). In addition to principlism, the mental health industry should also consider the following common ethical AI guidelines across different AI governance organizations. Table 1 presents a synthesis and merging of the key elements from principlism bioethics and various joint organizations' frameworks, adapting them to AI implementation (Akbarighatar, 2024; Beauchamp & Childress, 1979; Buruk et al., 2020; European Commission, 2021; Pujari et al., 2023; Shahriari & Shahriari, 2017; The White House, 2023).

**Table 1**

*Responsible AI Principles*

| Domain | Principle | Description |
|---|---|---|
| Principlism bioethics | Autonomy | AI system where the users have the right to make their own decisions in what data to give to train the AI system (ongoing easy opt-out process); allowing the users to choose from multiple decisions that include AI and non-AI interventions. |
| | Beneficence | AI system that does what is good for the users. |
| | Non-maleficent | AI system that avoids causing harm or minimizes risks to the users; it must refer back to autonomy principle above to allow users to choose harm-reduction interventions. |
| | Justice | AI system that allows equal access to services and interventions, closing healthcare disparities, protecting vulnerable populations, and protecting human dignity, satisfy legal requirements. |
| General AI ethics | Reliability and safety | AI systems should incorporate preventative and backup measures against failures and accidents. |
| | Privacy | Non-intrusive system that uses data that is legally and properly collected with individuals' consent |
| | Security | AI with safeguard that protects data and control access level to ensure authorized use |
| | Accountability | The system, system owners, and system designers are held responsible and accept consequences for all outputs and decisions. |
| | Explainability | The system's ability to explain and be open about how the system functions and make decisions |
| | Intelligibility | Users must be able to understand AI's reasoning and decision-making (information that users can understand). |
| | Transparency | AI system has clear process in revealing its data, methodology, capabilities, and decision-making (easy-to-understand and easy-to-use commands to "open the hood" to intelligible data and reasonings). |
| | Inclusiveness | AI system trained on a diverse set of data consisting of individuals and perspectives, regardless of their unique circumstances. |

| Domain | Principle | Description |
|---|---|---|
| | Fairness | AI systems with data trained and an algorithm designed to make fair decisions for everyone without discriminating |

**Importance of a Person-to-Person Therapeutic Relationship**

The person-to-person therapeutic relationship between clients and providers is crucial for successful treatment outcomes (Ackerman & Hilsenroth, 2003; Norcross & Lambert, 2018). Much research consistently underscores the significance of empathy, rapport, and trust in the therapeutic alliance between the clinician and the client since these elements promote effective psychotherapeutic outcomes (Battaglia, 2019; Elliott et al., 2011; Norcross & Lambert, 2018). Establishing interpersonal aspects serves as the fundamental basis for developing treatments. These factors create an environment where clients feel acknowledged, affirmed, and comprehended. Consequently, this fosters their inclination to actively participate in the therapeutic process (Del Re et al., 2012; Norcross & Lambert, 2018). Nonetheless, there are apprehensions regarding the mental healthcare field's growing incorporation of AI, as it can disrupt or diminish the essential human connections and emotional attachments that facilitate successful therapy. Consequently, this presents significant risks to the treatment procedure (Borghouts et al., 2021).

When AI makes medical decisions, it can potentially alter the traditional therapeutic relationship by bringing into the equation programmers, product developers, and AI-driven tools that do not have medical or mental health expertise. This dynamic will likely pose a challenge to the established medical ethics framework of shared decision-making, as Molnár-Gábor (2019), Kerasidou (2020), and McDougall (2018) discussed. Therefore, given the significance of the clinician-patient rapport, it requires ensuring the proper integration of AI technology to augment rather than supplant human physicians (Denecke et al., 2019). Artificial intelligence technologies can enhance therapists' work by analyzing therapy sessions, automating administrative activities, and providing data-driven insights to facilitate informed treatment choices (Borghouts et al., 2021). Using AI to enhance human capabilities can alleviate clinicians' workload, making it possible to maintain the therapeutic alliance while allowing clinicians to concentrate on the fundamental elements of human interaction and compassion (Denecke et al., 2019; Schueller et al., 2019). However, excessive dependence on technology in mental health can undermine therapeutic alliance and potentially compromise treatment outcomes, as *The British Journal of Psychiatry* emphasizes that while digital tools can complement traditional therapies, they should not replace the human connection essential for personalized care (Hollis et al., 2015). According to Schueller et al. (2019), responsible leadership in the mental healthcare sector must guide AI's deployment to enhance, rather than undermine, the interpersonal therapeutic connections essential for successful psychotherapy.

## Technology-Organization-Environment Framework

Louis G. Tornatzky and Mitchell Fleischer developed a TOE model in 1990. The TOE framework is an organizational-level theory developed to model the context surrounding a firm's adoption of technology as it focuses on three contexts: organization, environment, and technology (Tornatzky et al., 1990). Technological context refers to the existing and emerging technologies relevant to the organization, including perceptions of benefits, accuracy, trustworthiness, and compatibility (Gangwar et al., 2015). Organizational context includes factors like managerial structure, leadership support, company culture, funding, communication processes, training resources, and internal networks that shape adoption (Oliveira & Martins, 2010). Environmental context encompasses the regulatory environment, professional code of ethics, industry standards, competitive landscape, vendor ecosystem, and other external forces that enable or constrain adoption (Oliveira et al., 2014). Figure 1 and Table 2 present the TOE framework.

**Figure 1**

*Technology-Organization-Environment Framework*

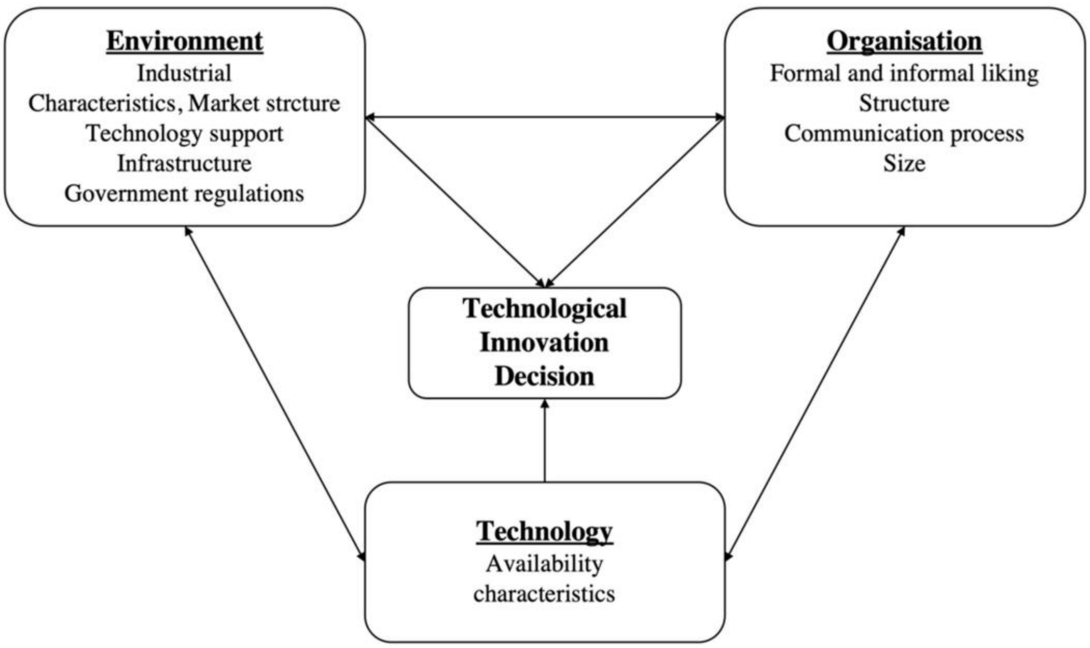


**Table 2**

*Technology-Organization-Environment Framework*

| Context | Constructed concepts | Detailed factors |
|---|---|---|
| Technological | All technologies supported inside/outside the organization<br>The ability for technology adoption and suitability of the current technology to the organization | Relative advantage<br>Conformance<br>Technology complexity |
| Organizational | Refers to the inherent characteristics and resources the organization possesses<br>Management leadership and communication play a crucial role in innovation.<br>Organization scale and resource availability are also important for decision-making. | Corporate size<br>Project range<br>Management support<br>HR scale<br>Competitive advantage<br>Available resources |
| Environmental | Effectiveness and efficiency factors for organizations' business activities<br>Includes organization's industry, competitors, governmental regulations, business partners, etc. | Market environment<br>Competition intensity<br>Government policy and regulations<br>Infrastructure of technological resources |

While the TOE framework has been applied and proven to be effective in predicting the adoption of technology innovation in various industries (Gangwar et al., 2015), there has been limited direct application and research involving the mental health sector specifically. As such, there are gaps in understanding how the TOE model may explain technology adoption decisions and processes in this field (Chatterjee et al., 2021). The unique considerations of healthcare, strong emphasis on ethics and relationships, and extensive regulations governing providers suggest value in examining the TOE factors that influence the adoption of innovations like AI in this sector (Kimiagari & Baei, 2021). Further research adapting and validating the TOE framework in mental health contexts would provide beneficial new insights into the technological, organizational, and environmental forces at play.

By taking this TOE framework approach, the interview data may uncover MHLs' key perceptions around AI technology. It allows for a multilayered understanding of the personal, organizational, and environmental dimensions impacting the adoption and ethical implementation of AI (Na et al., 2022). The data can also highlight the organizational implementation needs, such as training, leadership support, resource allocation, and change management tactics required to enable smooth adoption (Misra & Mondal, 2011). In essence, the qualitative insights gathered can reveal how wider contextual factors such as professional ethics, regulations, and standards in the mental healthcare environment shape attitudes toward responsible AI integration. This understanding may help organizations develop initiatives, processes, and strategies focused on facilitating appropriate AI deployment that complements clinical judgment and preserves the priority of compassionate, human-focused therapeutic relationships.

Chapter Two provided the technology overview, history of AI, the development of AI, AI ethical issues in mental health, distinctions between ethics and responsible AI, as well as the use of the framework of TOE to examine the issue of responsible AI implementation. Chapter Three details the methodology, data collection, and analysis for this study.

# Chapter Three: Methodology

This study addressed the problem of how to ethically adopt AI in mental health organizations while retaining human expertise and care values. Through semi-structured interviews, the research uncovered MHLs' perspective of AI technology, as well as organizational and environmental concerns that influence its implementation. This work sought to identify the internal and external resources necessary to develop an initiative that guides policy and practice for the responsible adoption of AI while maintaining the core values of mental health care, such as human expertise, judgment, and connection. This chapter outlines the method to recruit participants for semi-structured interviews. Additionally, considerations regarding credibility, trustworthiness, biases, and ethical foundation are discussed later in the chapter.

## Research Questions

1. What are the MHLs' perceptions of adopting AI technologies in their organizations?
2. What organizational factors influence MHLs to adopt AI technologies responsibly in their organization?
3. What external factors are MHLs concerned about the most when evaluating their organization's readiness to adopt AI technologies?

## Overview of Design

This study employed a qualitative approach, utilizing semi-structured interviews to answer the research questions. Merriam and Tisdell (2016) defined qualitative research as examining how individuals construct meaning from living experiences. An exploratory qualitative design was appropriate for this study to gain an understanding of MHLs' perspectives regarding AI technologies, what meaning they attach to them, and what might help them adopt

AI responsibly in their organizations. I chose semi-structured interviews for this study because they allow for a holistic perspective of the individual and make sense of one individual's experiences. In combination, these studies may provide a deeper understanding of the contexts surrounding the participants and the intended outcomes (Merriam & Tisdell, 2016).

**Participating Stakeholders**

This research utilized purposive sampling and recruited 11 leaders at various levels of responsibility from community mental health and group practice organizations to participate in interviews. Eligible leadership roles included frontline supervisors overseeing employees, middle managers supervising frontline supervisors, senior executives supervising middle managers, and thought leaders in advisory/consultative roles providing expertise on technology, ethics, and policy. The goal of this study was to elicit perspectives from leaders from diverse vantage points to develop guidelines for responsible adoption of AI by identifying organizational workflows, decision-making, priorities, needs, and values. Interviews with leaders at different levels provided insights into technology integration needs and risks from multiple leadership lenses to help guide ethical innovation and process changes.

**Interview Sampling Criterion and Rationale**

Participants were MHLs who met the following criteria.

- They have a background in community mental health or group practice settings.
- They have at least 1 year of experience in a leadership position related to clinical operations and clinical technology, which may intersect with a clinical license, though licensure is not mandatory for inclusion.

The goal was to target leaders with extensive mental health experience who are currently in organizational decision-making roles related to administration and care delivery workflows.

Exclusions focus the sample on those in community-based and group practice organizational settings rather than academics or solo private practices. Exclusion criteria are as follows:

- professionals whose leadership experience is exclusive to self-owned, single-provider practices
- entry-level practitioners without leadership or management duties

This research recruited leaders with direct oversight of patient care, organizational workflows, personnel, and operations that influence the adoption of new technologies. By sampling U.S.-based MHLs with managerial experience guiding clinical teams and processes, the aim was to elicit insights into organizational needs, values, and risks to responsibly inform AI integration in ways that align with therapeutic goals. I excluded clinical supervisors without personnel management duties since the focus was on leaders embedded in organizational operations beyond direct clinical training roles.

I recruited MHLs from the mental health industry across the country whom I identified through my LinkedIn and Facebook networks. After interviewing eight MHLs revealed limited knowledge of AI systems, the sampling strategy expanded to include three participants with clinical expertise and technology leadership roles. This intentional sample diversification helped capture broader perspectives on AI implementation in mental healthcare settings.

**Research Setting**

As a research setting, I conducted interviews using Zoom video conferencing. The interviews lasted up to an hour. Using Zoom as the primary interview tool was a practical decision as video conferencing has become common and convenient following the COVID-19 pandemic. Online interviewing also allows for a much broader geographical reach. In-person interviews would have been limited to MHLs in the San Francisco Bay Area, where I reside;

Zoom allowed participants to participate from anywhere in the country. Zoom's recording capability captured each interview's audio and video. As Merriam and Tisdell (2016) noted, recording interviews is beneficial because it ensures the preservation of all information shared during the interview for later analysis. I informed participants of the session's confidentiality and privacy, as well as the recording of the session. I gave the participants an overview of the manner in which I would ask questions.

**The Researcher**

As a first-generation Chinese American immigrant and clinical director of a mental health organization in Silicon Valley, my positionality significantly shapes this research. I work remotely, conduct meetings via videoconferencing, and am comfortable with and even favor technology that enhances my work life with minimal travel. My 15 years of clinical experience, including 9 years managing a mental health clinic, drove my interest in exploring AI's potential to address pressing challenges in mental healthcare delivery, particularly staff burnout and documentation struggles. Having witnessed the internet's evolution and impact on data privacy since childhood, I bring technological optimism and measured caution to this research.

My love for technology began with a struggle as a 4th-grade immigrant with no English skills. My teacher assigned *Black Beauty*; I still felt the overwhelming sensation when I first opened the book, filled with dense English sentences and pages of unfamiliar words. Moving glacially, looking up definitions for every third word, I could not finish the assignment in time. Though my teacher was sympathetic to my limited English skills, I felt ashamed for being a failure. That summer, my mother bought a newly released electronic dictionary because my sisters needed it for SAT preparation. Even sharing the device among three siblings for a third of the time each was faster than using a paper dictionary. Fortunately, that year's summer school assigned *Black Beauty* again, and I finally finished the book with pride. This experience taught me that technology can enhance our capabilities. Today, my Kindle displays instant word definitions with a simple tap, a far cry from my early struggles with paper dictionaries. With technology, we may lose some rudimentary skills, like searching through paper dictionaries, but we gain comprehension, analytic, and decision skills. We can learn the rudimentary skills when needed, just as my tech-immersed daughters are now learning to use paper Spanish–English dictionaries in their middle school Spanish class.

A recent experience with purchasing life insurance exposed a challenge with my healthcare data sovereignty. During the underwriting process, the insurance company reviewed my medical records and discovered that one of my physicians had documented a serious diagnosis that I never actually had and was never treated for years ago. This incorrect information now lives in my permanent medical record, creating significant complications. The situation became even more frustrating when we learned that the physician had relocated to another state, making it nearly impossible to have them correct the record. Although not AI-related, this experience revealed how little sovereignty I have over my medical data. Despite the data being about my health, I found myself entangled in bureaucratic red tape, unable to quickly correct this critical error in my record. While the information belongs to me in principle, the practical reality is that I have limited power to ensure its accuracy or make necessary corrections, even when errors could significantly impact important life decisions like insurance coverage. This situation perfectly illustrates my frustration as a medical consumer and why I care about data sovereignty in healthcare.

My being an eager technology adopter and frustrated healthcare consumer, as well as my immersion in Silicon Valley’s tech-forward culture, influence my attitude towards AI technology in mental healthcare. As a mental health clinic director interested in implementing AI, I recognize several potential biases in my research approach. My confirmation bias may have favored evidence supporting AI's benefits in mental health care while overlooking contradictory information. As a Chinese immigrant female with a postgraduate education, my in-group bias could have led me to identify more closely with similar individuals, potentially neglecting perspectives from different backgrounds. My professional bias as a mental health professional and assumptions about leadership and AI in patient care may have influenced my research focus. My position as clinic director might have triggered social desirability bias, where participants provided answers they thought I wanted to hear. My current role could have created status quo bias, affecting my interpretation of findings.

To maintain objectivity, I implemented several strategies during the interview. First, I used a consistent, structured interview protocol for all participants to ensure uniformity in data collection. During interviews, I maintained neutrality by avoiding sharing my opinions or experiences and using neutral therapist acknowledgments such as "Could you tell me more about that?" rather than evaluative responses. I documented participants' responses verbatim to ensure accuracy and recorded all interviews for precise transcription. I recruited MHLs with different cultures, backgrounds, and locations for diverse viewpoints. While my professional network connections to some participants may introduce social desirability bias, this insider perspective also enables a more profound understanding of the field's complexities. I want to implement AI solutions in my organization responsibly; while not a direct conflict of interest, it requires acknowledgment as it may influence my interpretation of the data. This study's value is not in achieving perfect objectivity but in leveraging my unique position at the intersection of clinical practice, organizational leadership, and technological innovation to explore the complexities of responsible AI adoption in mental healthcare. My mental health professional and technology enthusiast background provides a distinctive lens for examining how leaders navigate these challenges while maintaining ethical practices and therapeutic integrity.

**Interview Approach**

The interviews were semi-structured to deeply explore MHLs' experiences and perspectives regarding AI adoptions in their organizations. The primary instrument was a semi-structured interview protocol containing open-ended questions tied to the research aims while allowing flexibility for probes during the conversation. A semi-structured interview utilizes a predefined set of open-ended questions but maintains flexibility in the order and wording (Merriam & Tisdell, 2016). The interview guide was organized to facilitate a natural conversation flow while ensuring comprehensive coverage of research themes. Questions served as a guide while allowing spontaneity to follow unexpected insights that arise.

Each main question was accompanied by potential follow-up probes designed to deepen the discussion. For instance, when discussing organizational readiness, probes explored specific challenges, success factors, and decision-making processes. This fluid approach allows researchers to listen closely to the interviewee's perspectives as the discussion organically evolves, adapting and probing with additional questions as necessary (Merriam & Tisdell, 2016). Semi-structured interviewing allows for spontaneous customization based on each participant's worldview and the dynamics of the discussion rather than strictly dictating the process. In essence, flexible interviewing yields richer data than rigidly structured interviews limited to scripted inquiries. Semi-structured interviews are well suited to in-depth qualitative research due to their focus and adaptability, allowing participants to freely share their subjective perspectives and experiences while remaining on task throughout the interview.

**Data Collection and Documentation**

Semi-structured interviews served as the primary data source, following Merriam and Tisdell's (2015) qualitative research protocol. The interviews, lasting 60 to 90 minutes, were conducted via video conference platforms with purposively sampled participants who met inclusion criteria due to their roles as MHLs. The protocol incorporated flexibility to adapt questioning based on participants' responses while ensuring consistent coverage of core research topics.

A comprehensive documentation system captured both verbal and non-verbal aspects of the interviews. Following Creswell and Creswell's (2017) recommendations, interviews were recorded via Zoom with backup transcription through Otter.ai to preserve participants' exact responses and emphasis on particular topics. Concurrent field notes documented significant reactions, body language, and contextual observations that audio recordings alone could not capture (Sutton & Austin, 2015). This detailed note-taking approach recorded environmental contexts, behaviors, and non-verbal cues that provided crucial interpretive context. The documentation particularly focused on capturing participants' real-world examples, organizational stories, and specific experiences that illustrated their perspectives on AI adoption in mental healthcare settings.

To maintain research integrity and participant confidentiality, unique numeric identifiers linked all data sources—audio recordings, field notes, and transcripts—for each participant. This systematic documentation method preserved both verbal content and meaningful context for analysis while enabling consistent data organization across all interviews. The combination of recorded interviews, field notes, and secure transcription services followed McLellan et al.'s

(2003) guidance for efficient conversion of verbal data into analyzable text while maintaining data security and confidentiality.

**Data Analysis**

The analysis process followed Patton's (1999) guidelines for ensuring credibility in qualitative inquiry through rigorous methods, researcher competence, and systematic documentation. Interview recordings were first transcribed verbatim and verified for accuracy. Using qualitative coding methods, I conducted a systematic analysis of the text transcripts to identify themes and patterns relevant to the research questions. The analysis process involved multiple phases: initial open coding to identify key concepts, focused coding to develop categories, and theoretical coding to establish relationships between categories.

To enhance trustworthiness, I employed several validation strategies. Data triangulation involved cross-referencing findings with organizational documents and field notes to substantiate emerging themes. ATLAS.ti software facilitated the organization and management of coded extracts, category development, and analytic memos. A detailed codebook maintained clear definitions and application rules for each code, while an audit trail documented key analytic decisions and their rationale. Regular peer consultations provided analyst triangulation, helping to verify coding decisions and theme development.

Throughout the analysis, I maintained close alignment with the study's theoretical framework, interpreting emerging themes within the context of the TOE framework and existing literature on AI adoption in healthcare settings. This systematic approach to data analysis supported the development of findings that directly addressed the research objectives while maintaining methodological rigor.

## Research Credibility and Trustworthiness

Merriam and Tisdell (2016) stated that assessing a qualitative study's credibility and consistency can determine its trustworthiness. Credibility refers to the legitimacy of a study's findings, while consistency refers to whether those findings are consistent and congruent with the data (Merriam & Tisdell, 2016). Reflexivity (Merriam & Tisdell, 2016) and thick description (Geertz, 1973) help to maximize credibility. Reflexivity, sometimes called the researcher's position (Merriam & Tisdell, 2016), refers to the researcher's critical self-reflection on their biases, assumptions, and relationship to the research and its participants/topic. It is important to recognize that biases and assumptions can have a significant impact on the interpretation of qualitative data. While researchers cannot eliminate all biases and assumptions, identifying and explaining those biases and assumptions can mitigate data interpretation impact. Thus, readers can better understand why the researchers arrived at their conclusions. In the second strategy, thick description (Geertz, 1973), a researcher must write a detailed narrative, often referred to as a vignette, which emphasizes the situation as well as the background and context of the research topic. The approach takes into account emotions, voice, social relationships, details, feelings, actions, and context when interpreting events, behaviors, or observations. The more details and nuance a researcher provides, the more realistic and credible the findings will appear to the reader.

To ensure consistency, I used the strategy of an audit trail (Merriam & Tisdell, 2016). Merriam and Tisdell (2016) described the documentation process used to methodically track all components of the research and provide a paper trail of the decisions made throughout the process. An audit trail typically includes raw data like transcripts, notes, and survey results; data analysis and reduction like coded themes and categories; data reconstruction such as findings and conclusions; process notes about design choices and changes; reflexive notes about my assumptions and insights; instrument development information; ethical considerations and consent forms. I documented the data analysis for this study in ATLAS.ti by creating an audit trail.

## Research Ethics

Research involving human participants raises numerous ethical issues, according to Creswell and Creswell (2017). Research in this study was subject to the approval of the institutional review board before I could contact participants for interviews or surveys. The board examined the purpose of the study, informed consent, interview questions, and methodology of the study. Its purpose is to protect the participants' privacy and to prevent them from being exploited. I informed all participants that their participation was voluntary and that they could opt out at any time. To ensure transparency, all participants received a final copy of the dissertation.

## Limitations and Delimitations

There are several limitations of this study, such as the small sample, selection bias, reliance on self-reported data, researcher biases, social desirability influences, time constraints, and context-dependence. A primary limitation of this study is its sample size and composition. This research included 11 participants, all of whom were MHLs in various capacities and AI knowledge. While these participants provided rich and detailed insights, the limited sample size may not fully capture the diverse perspectives in the broader mental health leadership community. Another significant limitation of this study stems from its geographical concentration, which may have introduced biases that limit the generalizability of the findings. Of the 11 participants, eight were based in California, two in the Midwest, and one in New York

City. This California-centric sample presented several challenges to the broader applicability of the results. The rapidly evolving nature of AI technologies presents another limitation to this study. The field of AI is advancing at an unprecedented pace, with new developments and breakthroughs occurring frequently. This rapid evolution means that some of the findings may become outdated relatively quickly. What is considered cutting-edge AI applications in mental health care today may be superseded by new technologies in the near future.

Delimitations include the scope focusing solely on leaders, the community mental health sector, and AI technology, as well as the use of purposeful sampling and being framed by specific theoretical lenses like the TOE framework. Although these constraints limit generalizability and introduce critical biases to be considered, the study remains worthwhile because it uncovers MHLs’ perspectives regarding ethically integrating emerging AI from a targeted leadership sample within defined parameters.

## Chapter Four: Results or Findings

This qualitative study examined the technological, organizational, and environmental factors that influence MHLs in the responsible adoption of AI for patient care. Using the TOE framework and principlism ethical theory, this research aimed to develop a practical guide for mental health organizations to implement AI responsibly while bridging the gap between responsible AI adoption and equitable patient care.

The analysis revealed several key dimensions influencing MHLs' approach to AI adoption, characterized by a prevailing sentiment of cautious optimism. While leaders recognized AI's potential to enhance therapeutic outcomes and operational efficiency, this optimism was tempered by complex concerns regarding data privacy, integrity, and governance in mental health information management. The findings revealed an important balance between AI's potential benefits and ethical responsibilities. Leaders recognized how AI could enhance clinical work by reducing administrative burden, expanding service capacity, improving organizational efficiency, and increasing grant funding. However, they emphasized that pursuing these operational advantages must not compromise their ethical duties to provide quality patient care and maintain professional clinical judgment.

Organizational factors, including psychological safety, resource allocation, and leadership characteristics, emerged as crucial elements for responsible AI implementation. External forces, including governmental regulations and stakeholder interests, significantly influenced adoption decisions beyond individual leadership control. Notably, disparities in organizational funding raised concerns about potential treatment inequities while maintaining patient trust and data sovereignty emerged as fundamental requirements for successful AI integration in mental healthcare delivery. The findings from this research can guide mental health organizations in developing responsible approaches to AI adoption that align with their ethical obligations.

## Participants

Eleven MHLs from community mental health or group practice settings participated in this study (Table 3). I recruited participants from LinkedIn profiles of these organizations' leaders and professional contacts that fit the criteria from Chapter Three. They held the eligible leadership roles presented in Chapter Three, and I invited them to participate in the semi-structured interviews. Participants' leadership roles included frontline supervisors overseeing employees, middle managers supervising frontline supervisors, senior executives supervising middle managers, and thought leaders in advisory/consultative roles providing expertise on technology, ethics, and policy. The participants had at least 1 year of leadership experience in clinical operations or technologies. Three participants were C-level executives, four were clinical directors, one was a clinical manager, and three were directors of clinical technologies. All work in the mental health patient care sector, representing diverse perspectives in gender, race, and age.

**Table 3**

*Interviewee Pseudonyms and Professions*

| Participant | Professional background |
|---|---|
| Jack | Jack is a licensed mental health practitioner and is a clinical director at a large mental health non-profit organization, where he plays a pivotal role in overseeing the agency's operations and service delivery. With over 13 years of experience as a licensed mental health counselor, Jack has honed his expertise in managing diverse mental health programs and leading multidisciplinary teams. His current responsibilities include supervising a team of 15 direct reports, ensuring the effective delivery of client-centered care, and maintaining high standards of clinical practice. |
| Samantha | Samantha is a licensed mental health practitioner in California with 5 years of clinical experience providing direct mental health services. In addition to her work as a therapist, she is also a mental health technologist specializing in the development and implementation of AI-driven clinical products. Her focus is on improving service delivery and ensuring compliance with regulatory standards, combining her expertise in mental health with cutting-edge technology to enhance the quality and efficiency of care. |
| Adam | Adam is a licensed mental health practitioner, and he serves as the chief clinical officer at a mental health NPO in California, where he leads the clinical vision and strategy for the organization. With over 15 years of experience as a licensed mental health therapist and a decade in leadership roles, Adam has a deep understanding of both clinical practice and organizational management. His leadership focuses on enhancing the quality of care, optimizing clinical workflows, and ensuring compliance with regulatory standards. |
| Solaris | Solaris is a clinical psychologist and a mental health technology director in a digital mental health delivery company who brings 35+ years of expertise in direct mental health services, digital health solutions, and behavioral science. She has successfully navigated the intersection of traditional mental healthcare and technological advancement, contributing to the transformation of mental health service delivery in the digital age. |
| Jonathan | Jonathan is a clinical psychologist in his state, and he serves as a C-level executive at a major mental healthcare system, where he directs enterprise-wide initiatives focused on technological innovation, operational excellence, and clinical quality enhancement. His executive leadership centers on transforming healthcare delivery through strategic technology integration while optimizing operational efficiency and elevating clinical outcomes. |

| Participant | Professional background |
| --- | --- |
| Sarah | Sarah, a clinical psychologist, serves as a mental health clinical director at a prominent virtual mental health care organization. In this leadership role, she oversees the delivery of digital mental health services, combining her clinical expertise with innovative telehealth approaches to ensure comprehensive client care. |
| John | John is a licensed mental health practitioner serving as a clinical supervisor in a major hospital system's behavioral division. He leverages his licensed mental health practice background to lead clinical teams and enhance service delivery. His role bridges direct clinical oversight with system-wide behavioral health initiatives, ensuring comprehensive mental healthcare in an integrated hospital setting. |
| Jennifer | Jennifer serves as a clinical technologies director, bringing together her mental health practitioner licensure with technological expertise. She specializes in advancing mental healthcare through digital innovation while maintaining clinical excellence and therapeutic integrity. Her dual expertise enables her to develop and implement technology solutions that enhance mental health service delivery while ensuring adherence to clinical best practices and therapeutic standards. |
| Benjamin | Benjamin is a clinical psychologist and senior vice president at a leading non-profit mental health organization who brings comprehensive expertise in program delivery, clinical operations, human resources, and research. His senior executive role encompasses strategic leadership across clinical services, operational management, and organizational development. |
| Lee | Lee is a licensed mental health practitioner and currently serving as a clinical director who brings 10 years of comprehensive experience across the non-profit mental health sector. Drawing from experience on both clinical and administrative sides of mental healthcare, Lee excels in resolving complex organizational challenges while driving innovation and sustainable growth in this transforming field. |
| Kacey | Kacey serves as a C-level executive, bringing their clinical psychology expertise to strategic healthcare leadership. Kacey drives organizational excellence through strategic oversight of clinical services, policy development, and compliance management. Their leadership focuses on integrating evidence-based practices with robust training programs while ensuring regulatory compliance and service quality. Drawing on their clinical background, they effectively bridge therapeutic best practices with organizational strategy and regulatory requirements. |

Of the 12 participants who initially agreed to take part, one faced multiple scheduling conflicts and chose to respond to interview questions and follow-up inquiries via email to allow for greater flexibility. Another participant did not prefer an interview and opted to provide answers by email, allowing additional time for thoughtful responses. The remaining 10 MHLs completed 60-minute confidential, open-ended, semi-structured face-to-face interviews conducted via Zoom. I analyzed these 10 in-depth interviews, conducted over 10 weeks, using both a priori and manual posteriori coding, ensuring a comprehensive and reliable data collection and analysis.

## Findings for Research Question 1

This section presents the findings for the first research question: What are the MHLs' perceptions of AI technologies in mental health care? Focusing on the technology aspect of the TOE framework, the integration of AI is notably shaped by MHLs' perceptions and attitudes (Liehner et al., 2023). The MHLs interviewed shared their thoughts on AI technologies based on their organizational experiences, personal research, and media reports on AI. It is important to note that some of the respondents may have only limited exposure or basic use of these technologies. Understanding these perceptions is essential as they influence decision-making and guide strategies for responsible AI implementation.

The MHLs identified several key areas where AI could benefit mental healthcare delivery. Their responses highlighted AI's potential to enable more holistic and personalized treatment strategies, enhance predictive analytics for preventative care, improve operational efficiency while reducing provider burnout, increase the accessibility of mental health services, and reduce the stigma associated with seeking mental health support. While expressing cautious optimism about these potential benefits, leaders consistently emphasized the importance of balanced implementation that preserves the human element of care. The following sections explore these potential benefits in detail, drawing from participants' experiences and perspectives.

**Theme 1: Perceived Potential Benefits: Cautious Optimism**

The MHLs expressed a sense of cautious optimism regarding the potential benefits of AI technologies in mental health care. Lee articulated this sentiment: "I think right now we're cautiously optimistic. We definitely see the potential, and it has been somewhat useful because we haven't seen a lot of tools yet for mental health." While acknowledging the uncertainties and challenges, these MHLs see significant opportunities for AI to enhance service delivery, improve patient outcomes, and address longstanding issues in the field. Jack encapsulates this broad vision: "I think it has a lot of potential to be really transformative for everything from maybe examining outcomes to screening patients and getting them to the appropriate provider."

These quotes highlight the spectrum of potential benefits that MHLs envision, from improving diagnostic processes to enhancing treatment efficacy. The use of the word "transformative" underscores the magnitude of change that AI could bring to the field. However, the recurring phrase "I think" in both quotes reflects a degree of uncertainty, aligning with the overall sense of cautious optimism (May, 2018). MHLs recognize that while AI holds promise for addressing critical challenges in mental health care, its implementation must be approached thoughtfully and with careful consideration of potential impacts on patients and practitioners alike.

A key advantage of AI identified by MHLs is its potential to facilitate more holistic and personalized treatment approaches. This potential stems from AI's ability to process and analyze vast amounts of complex data, leading to more comprehensive and nuanced understandings of individual patients.

Solaris envisions a future where AI enables highly tailored care: “Picture personalized treatment plans tailored to individual needs, leveraging vast datasets and machine learning to optimize care.” This perspective highlights the potential for AI to transform treatment planning by considering a wide range of factors that influence mental health. Artificial intelligence’s capacity to suggest novel interventions is another aspect that excites MHLs. As John noted, “[AI] could give you ideas of what interventions you can use that maybe you have not thought of.” This ability to expand the repertoire of treatment options could be particularly valuable in complex cases or when traditional approaches have been ineffective.

Kacey emphasized how AI-driven approaches could empower patients by exposing them to diverse treatment strategies:

> It also allows the clients to learn different ways of doing things rather than always believing there’s only one way to do things. They say, “Oh, there’s multiple ways of doing things with good outcomes. It also takes them out of their shell or stagnant way of looking at things.”

This perspective suggests that AI could aid in broadening patients’ horizons and fostering a more flexible, open-minded approach to mental health care. This diversity of options can be particularly beneficial for patients who feel stuck or have not responded well to traditional treatment methods. Moreover, the ability of AI to suggest alternative approaches can help normalize the idea that mental health treatment is not a one-size-fits-all solution. This aligns with contemporary views in mental health care that emphasize personalized culturally sensitive treatment plans (Alhuwaydi, 2024).

Another promising aspect of AI in mental health care is its potential to bridge the gap between physical and mental health. Adam highlighted this potential:

> And people are not just somatizing their mental health conditions anymore and actually finding the correlation. Oh, my tummy hurts. Maybe it's something else. Let's talk about it. Oh, you know what, that would be cool. Maybe part of the differential diagnosis is, oh, you have a stomachache. Have you considered depression? Wouldn't that be cool?

This holistic approach could lead to earlier detection of mental health issues and more comprehensive treatment strategies that consider both physical and mental well-being. The integration of AI in this context could bridge the longstanding divide between physical and mental health care, offering a more nuanced and complete picture of a patient's overall health. AI can help identify subtle connections between physical symptoms and mental health conditions that might be overlooked in traditional diagnostic processes. By recognizing these correlations, AI systems could prompt healthcare providers to consider mental health factors when patients present with physical complaints and vice versa.

While enthusiasm for AI's potential is evident, MHLs also emphasized responsible implementation. As Jonathan noted, "I think it's a super exciting time that we're in. I do think we need to do it responsibly, and I use the keyword 'ethically,' and that's going to be the place where we're going to have to do." This sentiment underscores the need for careful consideration of ethical implications as AI technologies are integrated into mental health care.

The potential for AI to enhance cultural competence in mental health care is another area of interest. Jennifer expressed optimism about AI's role in this regard:

> I think it provides such a great way of … moving on into the future [of] mental health and … care itself. … Me myself looking at it from that … stance of incorporating into my

> business and looking at how AI through various algorithms through cultural competence can be supportive of individuals as [they] are navigating through the symptoms.

This perspective highlights the potential for AI to help address cultural disparities in mental health care, a longstanding challenge in the field.

While MHLs expressed excitement about AI's potential to revolutionize mental health care through personalized, holistic, and culturally competent approaches, they also emphasized the need for responsible and ethical implementation. This balanced view reflects a cautious optimism that acknowledges both the transformative potential of AI and the importance of careful, considered adoption in the sensitive field of mental health care.

A notable benefit of AI identified by MHLs is its capacity for predictive analytics and its potential to enhance preventative care. The ability of AI to analyze vast amounts of patient data to identify trends and predict future mental health needs is seen as a game-changer in the field. Solaris articulated this potential: "Imagine early detection of depression or anxiety, even before symptoms fully manifest, through sophisticated algorithms analyzing speech patterns and social media activity."

This quote highlights the proactive nature of AI-driven predictive analytics, suggesting that it could revolutionize early intervention strategies in mental health care. By identifying potential issues before they fully develop, clinicians could implement preventative measures, potentially averting mental health crises. Solaris further emphasized the forward-looking nature of AI in mental health care: "Looking ahead, I envision AI predicting mental health crises before they occur, enabling timely interventions." This perspective highlights the potential for AI to shift mental health care from a reactive to a proactive paradigm, potentially reducing the severity and frequency of mental health crises. This perspective highlights the potential for AI to shift the mental health care paradigm from reactive to proactive, potentially reducing the severity and frequency of mental health crises as AI's predictive capabilities can identify warning signs and risk factors much earlier, sometimes even before the patient and the provider can (J. Wang, 2023). By analyzing patterns in behavior, speech, social media activity, and other data points, AI could flag potential issues, allowing for early intervention and, therefore, can lead to less intensive treatments and overall better outcomes for patients (Atlam et al., 2022). Moreover, by preventing crises before they occur, this approach could significantly reduce the emotional toll on patients and their families, as well as the logistical and financial burdens on healthcare systems (Ejjami, 2024).

In essence, this shift from reactive to proactive care could revolutionize mental health treatment, making it more effective, efficient, and humane. Jack echoed this sentiment, highlighting the broad applicability of AI in mental health care:

> I think it could be utilized in a way to catch other forms of need earlier. Maybe it's just given a way to review sessions or review content that is screened from like an intake and a way to find are there other applicable services that might be needed just as a way to further analyze potential patients and how to surround them with support even beyond mental health care.

This quote suggests that AI's predictive capabilities could help in early detection and in optimizing patient care pathways, ensuring individuals receive the most appropriate interventions. Predictive analytics could lead to more targeted and effective treatment strategies, potentially improving patient outcomes and reducing the trial-and-error approach often necessary in mental health treatment.

It is important to note that while MHLs were enthusiastic about AI's potential in predictive analytics and preventative care, they also emphasized the need for human oversight. As Solaris stated, "Offering nuanced clinical judgment extends to our exploration of AI's potential in other areas, like personalized treatment plans and predictive analytics." This perspective reflects a balanced approach, where AI is seen as a powerful tool to augment clinical decision-making rather than replace it. The consensus among MHLs is that AI should be used for tasks such as pattern recognition, predictive analysis, and offering treatment planning suggestions but not for making final diagnostic decisions. As Benjamin noted,

> I always feel like the clinicians should be the ones in the driver's seat. They're the ones that are essentially making the call. They're the ones that have been trained. AI is really there to support as a co-pilot rather than as the person that is driving the decisions.

This cautious approach ensures that AI serves as a supportive aid to clinicians, enhancing their ability to provide preventative care and personalized treatment while maintaining the crucial element of human judgment in mental health care.

The interviewees consistently highlight the potential of AI to significantly enhance efficiency in mental health settings, particularly by reducing the administrative burden on professionals. This improved efficiency is seen as a key factor in preventing burnout and increasing job satisfaction among mental health staff.

The administrative workload in mental health care has grown substantially, largely due to requirements set by funding sources and payors who mandate detailed documentation to justify medical necessity and meet billing criteria. MHLs see AI as a promising solution to streamline these processes. As one participant succinctly put it, "It [AI] will free up a lot of time for the provider and the therapist." John elaborated on this potential, envisioning AI's role in documentation: “I think in a positive manner the way that I envision us utilizing AI is to really help with the transcribing and note taking for all the notes that we do for patients.” This reduction in administrative tasks is not just about saving time; it is about reallocating that time to more meaningful aspects of care. Benjamin emphasized this point:

> I definitely think that there would be some usefulness, lots of promising usefulness in terms of efficiency in the work that we do. Given that in mental health, if we are providers, we have a limitation in regards to the amount of time that we get to spend with our clients.

Adam further illustrated the potential for AI to handle repetitive tasks:

> I can see some benefits to it especially with the non-mental health care aspect of it. So, like the funding that we talked about, I think that would be something. Like disciplinary action you write the same thing over and over again. … So, I can see it opening up a lot of my time on the administrative tasks and really focusing on the other important things.

This efficiency gain is closely tied to the potential for reducing burnout among mental health professionals. Burnout, characterized by emotional exhaustion, detachment, and reduced personal accomplishment, has become a significant concern in the healthcare industry. By addressing this issue, MHLs hope to create a more positive and fulfilling work environment. Jonathan articulated this comprehensive vision: “I just think easier use of navigating healthcare services, reducing administrative barriers for providers, increasing Joy of work and reducing burnout, and improving patient experience.” This quote encapsulates the multifaceted benefits that MHLs anticipate from AI adoption, linking improved efficiency to both provider well-being and patient care quality.

The desire to refocus on patient care is a recurring theme. Benjamin expressed, “I do want to spend more time with my clients. I do want to spend more time doing more research on types of tools and techniques that I can help my client.” Samantha further emphasized how AI could alleviate the cognitive load on clinicians: “I think that AI does a lot of great things for taking away some of the mental burden on clinicians. That I think is- is a potential like really great efficiency, scalability getting more access—access.” Kacey provides a broader perspective on this potential transformation:

> I often see that being able to integrate AI will hopefully allow the therapists to be able to spend more time directly with the clients rather than on repetitive documentation. And a lot of time on the administrative tasks that currently also require—although as important they also require a lot of time.

In essence, MHLs view AI as a tool that could significantly reduce the administrative burden on mental health professionals, allowing them to focus more on patient care and engage in professional development. This shift is expected to improve the quality of care, enhance job satisfaction, and reduce burnout among providers. By improving efficiency and job satisfaction, AI could contribute to a more sustainable and effective mental health care system, creating a positive impact on both providers and patients. The enthusiasm for these potential benefits is tempered with a recognition of the need for responsible implementation, reflecting the cautious optimism that characterizes MHLs’ overall perspective on AI in mental health care.

The respondents recognize the significant potential of AI integration to enhance the accessibility of mental health services. This benefit is appealing as it addresses the dual challenges of controlling healthcare expenses while expanding the reach of mental health resources. As one participant succinctly noted, “[AI] helps make mental health resources more accessible.”

By leveraging AI-driven tools, organizations can provide a basic level of treatment to a broader population, effectively democratizing access to mental health support. These AI solutions, such as remote chat services and virtual triage systems, are breaking down traditional geographical barriers that have long impeded access to care. Solaris elaborated on this potential: “AI-powered chatbots and virtual therapists provide immediate support 24/7, bridging gaps in access and reducing the stigma surrounding mental health care.”

This quote highlights how AI can offer round-the-clock support, providing initial assessments, basic coping strategies, and guidance to appropriate resources, all without the need for in-person visits. A key focus for MHLs is the potential of AI to expand care to marginalized populations in remote areas who typically lack the resources to access traditional mental health services. Jonathan emphasized this point: “I actually … think AI is a solution to inequitable or disparate health outcomes. And so, … that’s why … I’m a big sort of advocate for it.” This perspective underscores how AI could play a crucial role in addressing longstanding issues of health equity in mental health care delivery.

The potential cost-effectiveness of AI solutions is another factor that MHLs see as key to enhancing accessibility. Adam articulated,

> I’m thinking cost-effectiveness. Therapy is expensive, very expensive, and not affordable for everybody. So, if we had [an] AI platform, maybe that’s something that brings therapy to the general population, which we know at this time may not be accessible for everybody.

By offering virtual support options, AI can help overcome the logistical and financial barriers often preventing individuals in remote or underserved areas from seeking help. This expansion of care improves individual outcomes and could address broader issues of health equity and social justice in mental health care delivery.

MHLs also recognized the potential of AI to provide support in situations where human-to-human interaction might be challenging. Adam provided an insightful example:

> That and then, and I can see for some of the disorders that we see some of the conditions that we deal with where it might be difficult for a person to engage with another human being, autism being one of them. I can see how AI could be a great platform to engage those populations and help them with whatever it is that they’re presenting.

This perspective highlights how AI could potentially bridge gaps in care for populations with specific needs that might not be easily met through traditional care models. In essence, MHLs see AI as a powerful tool for enhancing the accessibility of mental health services, particularly for underserved populations. However, it is important to note that while MHLs are enthusiastic about these possibilities, they also maintain a balanced view, recognizing the need for careful implementation to ensure the quality and appropriateness of AI-delivered care.

The participants recognize the potential of AI to significantly reduce the societal stigma associated with mental health issues. This reduction in stigma is seen as a crucial step toward improving overall mental health outcomes and encouraging more individuals to seek help when needed.

A primary way AI can contribute to reducing stigma is by providing anonymous platforms for individuals to explore their mental health concerns. As Mike pointed out, “AI chatbots can educate users about mental health issues without judgment or fear of exposure.”

This anonymity can be particularly empowering for people who may be hesitant to seek help due to fear of stigma or discrimination. It provides a safe space for individuals to ask questions, seek guidance, and learn about mental health conditions without the pressure of societal expectations.

Artificial intelligence’s potential to demystify mental health through education is another crucial aspect of stigma reduction. By providing consistent, reliable, and scientifically backed information, AI tools can help normalize conversations around mental health. Adam highlighted how this can lead to a more holistic understanding of health:

> And people are not just somatizing their mental health conditions anymore and actually finding the correlation. Oh, my tummy hurts. Maybe it’s something else. Let’s talk about it. Oh, you know what … would be cool? Maybe part of the differential diagnosis is, oh, you have a stomachache. Have you considered depression? Wouldn’t that be cool?

This perspective illustrates how AI can help individuals make connections between physical and mental health, potentially reducing the stigma associated with mental health conditions by framing them as part of overall health.

The potential for AI to encourage proactive mental health management is another way it can contribute to stigma reduction. By empowering individuals to take control of their mental health, AI can help shift the narrative from mental health being a taboo topic to one that is an integral part of overall well-being. Kacey’s comment reflects this potential:

> It also allows the clients to learn different ways of doing things rather than always believing there’s only one way to do things. They say, “Oh, there’s multiple ways of doing things with good outcomes.” It also takes them out of their shell or stagnant way of looking at things.

This quote suggests that AI can help broaden perspectives on mental health treatment, potentially reducing the stigma associated with seeking help or trying different approaches to mental wellness.

Furthermore, the increased accessibility of mental health resources through AI can itself contribute to stigma reduction. As mental health support becomes more readily available and integrated into everyday life, it may become more normalized and accepted. Solaris's vision of AI's potential illustrates this: "Picture personalized treatment plans tailored to individual needs, leveraging vast datasets and machine learning to optimize care." By making mental health care more personalized and accessible, AI could help shift societal perceptions, making mental health support as commonplace and accepted as physical health care.

The participants see AI as a powerful tool for reducing the stigma surrounding mental health. Through anonymous support, education, encouragement of proactive mental health management, and increased accessibility, AI has the potential to normalize mental health discussions and support-seeking behaviors. However, it is important to note that while AI offers these possibilities, MHLs also recognize the need for careful implementation to ensure that AI-driven approaches complement rather than replace human-led efforts in stigma reduction and mental health support.

**Theme 2: Perceived Technology Limitations and Concerns**

The interviewees identified several critical concerns and limitations regarding AI implementation in mental healthcare settings. Their responses highlighted issues surrounding AI's accuracy and reliability in clinical settings, concerns about bias and fairness in AI systems, challenges with cultural competency, questions about privacy and data security, the inherent complexity of mental health work, potential impacts on human interaction and empathy in therapeutic relationships, and compatibility challenges with existing healthcare systems. While acknowledging AI’s potential benefits, leaders emphasized these limitations as crucial considerations for responsible implementation. The following sections explore these concerns in detail, drawing from participants’ experiences and perspectives to understand the challenges that must be addressed for successful AI integration in mental healthcare.

While acknowledging the potential benefits, the participants shared the consensus that AI should be utilized primarily for pattern analysis, predictive analytics, and treatment planning suggestions rather than as an autonomous diagnostic tool. This perspective is rooted in the understanding that mental health care often requires nuanced interpretation and contextual understanding that current AI systems may not fully capture. Lee articulated this cautious optimism: “I think right now we’re cautiously optimistic would be the term. We definitely see the potential, and it has been somewhat useful because we haven’t seen a lot of tools yet for mental health.”

The concerns raised by MHLs regarding AI’s accuracy and reliability in mental health care reflect a broader set of challenges in implementing these technologies. While AI shows promise in enhancing efficiency and providing insights, its limitations in understanding the nuanced, context-dependent nature of mental health issues are significant. The potential for over-pathologizing normal human experiences, as highlighted by Adam, and the risk of rigid, oversimplified treatment protocols, as Jonathan noted, underscore the necessity of human oversight in AI-assisted mental health care. These concerns touch upon questions of bias, cultural competency, and the fundamental complexity of mental health work—areas that require careful consideration regarding AI integration. Delving deeper into these interconnected challenges reveals that addressing the limitations in AI’s accuracy and reliability is just the beginning of a more comprehensive evaluation of AI's role in mental health care. This leads us to explore another critical aspect of AI implementation: the potential for bias and the pressing need for fairness in these systems.

As AI technologies evolve and integrate into mental health care, MHLs express a nuanced understanding of their capabilities and limitations. Currently, there is a lack of confidence in AI's reliability as a standalone diagnostic tool, given the complexities of mental health diagnoses with their myriad nuances and contextual factors. Jonathan highlighted a key concern regarding the potential rigidity of AI-driven approaches: "The concern is that AI might lead to rigid treatment protocols, like starting with A for depression and moving to B if not effective."

This quote underscores the fear that AI might oversimplify complex mental health issues, potentially leading to inflexible treatment approaches that fail to account for individual patient needs and circumstances. As Adam elaborated on the risk of over-pathologizing normal human experiences, "Diagnosis could be over-pathologized by AI, misinterpreting normal life phases as conditions. AI may struggle to differentiate between pathological conditions and transient issues with protective factors."

This perspective highlights the concern that AI might lack the nuanced understanding required to distinguish between normal life challenges and clinical mental health conditions. Adam further emphasized this point: "I do think that could be a potential downside, the over-stigmatizing and over-pathologizing." Benjamin underscored the importance of human judgment in mental health care: "Oftentimes, we trust these things because they seem so powerful and intelligent, but the truth is just because you're intelligent doesn't mean you're reliable or accurate sometimes." This quote reinforces the belief among MHLs that human expertise remains crucial in interpreting and applying AI-generated insights, particularly given the complexities of mental health.

Sarah raised concerns about the potential limitations of AI in providing comprehensive oversight:

> I'm thinking, yes, you may be able to increase access. You may be able to provide more services, more access to patients needing mental health care. But there is also additionally more liability because … you're not able to really provide 100% oversight … if you were to be the one delivering the service.

This perspective highlights the potential risks associated with relying too heavily on AI systems, particularly in terms of liability and providing comprehensive care.

While MHLs recognize AI's potential to enhance efficiency and provide insights into mental health care, they emphasize the irreplaceable value of human expertise and interpersonal skills in providing comprehensive, nuanced care. The consensus is that AI should be viewed as a tool to augment human capabilities rather than replace them, particularly in the complex and highly individualized field of mental health care.

The MHLs expressed significant concerns about AI's ability to produce accurate and unbiased evaluations in mental health care. A primary worry is the quality and diversity of data used to train AI systems, as biases in this data can substantially influence system performance and perpetuate imbalances in mental health treatment, particularly for marginalized populations. Samantha captured this concern: "There's like cultural biases; it's still trained on a model that is predominantly White." This statement highlights a fundamental issue in AI development for mental health care: the lack of diversity in training data, as most AI models are trained on data from predominantly WEIRD populations, which may not reflect the experiences and needs of individuals from other backgrounds (Obermeyer et al., 2019). Samantha further elaborated on the implications of this bias: “And so, it misunderstands cultural nuances and all of that stuff, which … it’s just where we’re at right now, and I think people don’t necessarily understand the intricacies of that.” This underscores the complexity of the problem and the potential for AI systems to misinterpret or overlook important cultural nuances in mental health care. Jennifer echoed these concerns, emphasizing the importance of cultural competence in AI development:

> The potential downsides, I think, one would … be cultural competence with AI that … those who may create the algorithms with AI may not be looking from that lens of being able to support individuals from a variety of either ethnicities or cultural relevance, which could come back and … not be as supportive to an individual.

This perspective highlights the risk that AI systems, if not carefully developed with diverse cultural perspectives in mind, could fail to provide adequate support for individuals from various ethnic and cultural backgrounds. Despite these concerns, there is also recognition of the potential for improvement. Kacey noted, “We … actually have a lot more to learn around how to do that, but the potential of it, I think. is really exciting both from a provider perspective as well as a patient perspective and just community.” This statement reflects a cautious optimism, acknowledging the current limitations while also recognizing the potential benefits of addressing these biases and improving AI systems for mental health care.

The MHLs’ statements highlight the critical need for diverse datasets, culturally informed AI development, and ongoing evaluation of AI systems for bias. They emphasize that without careful attention to these issues, AI in mental health care risks perpetuating or even exacerbating disparities in treatment and outcomes for marginalized populations. Addressing these concerns is seen as crucial for ensuring that AI can be a tool for improving mental health care equitably across diverse populations.

Cultural competence has emerged as a critical focus in developing AI products for mental health care (McGregor et al., 2019). This emphasis reflects a growing recognition that effective mental health support must be sensitive to and inclusive of diverse cultural backgrounds. One participant articulated this priority:

> Culture! Making sure that we're culturally competent is paramount to me. That we really take a hard look at the individuals that we are supporting. And I desire for individuals to have just that level of comfort with whatever they are as they utilize our product."

This statement underscores the importance of creating AI tools that are not only technologically advanced but also culturally attuned to users' diverse needs. It highlights the critical need for AI systems in mental health care to go beyond mere functionality and incorporate a deep understanding of cultural nuances and sensitivities. This perspective recognizes that effective mental health care is inherently tied to cultural context and that AI tools must be designed with this fundamental principle in mind.

Adam further emphasized the multifaceted nature of cultural competence in mental health care: "So, culturally responsive, culturally informed, trauma-informed care which takes human factors into consideration." This quote expands on the concept of cultural competence, linking it explicitly to trauma-informed approaches and human-centered design. It suggests that truly effective AI in mental health care must be culturally responsive, understand the impact of trauma across different cultural contexts, and consider the complex human factors involved in mental health treatment. This comprehensive approach to cultural competence in AI design reflects a growing awareness among MHLs of the interconnected nature of culture, trauma, and mental health.

Cultural competence in AI goes beyond mere representation; it involves a deep understanding of cultural contexts, including trauma-informed approaches. It emphasizes that truly culturally competent AI systems in mental health care must be designed with a nuanced comprehension of diverse cultural norms, beliefs, and experiences that shape an individual's mental health and approach to seeking help. Moreover, this view suggests that AI developers and implementers need to incorporate trauma-informed principles, recognizing that many mental health issues are rooted in or exacerbated by cultural and historical traumas and that effective AI tools must be sensitive to these complex, culturally specific experiences. However, MHLs express concerns about the current state of cultural competence in AI development. Jennifer articulated this worry:

> The potential downsides, I think, one would … be cultural competence with AI that … those who may create the algorithms with AI may not be looking from that lens of being able to support individuals from a variety of either ethnicities or cultural re-relevance, which could come back and … not be as supportive to an individual.

This quote highlights the risk of AI systems failing to adequately support individuals from diverse cultural backgrounds due to limitations in the developers' perspectives. It underscores the concern that AI algorithms, if developed without sufficient cultural diversity in the team or training data, may inadvertently perpetuate biases or overlook crucial cultural nuances in mental health care. Furthermore, this perspective emphasizes the potential for AI systems to provide suboptimal or even inappropriate care for individuals from minority cultures or ethnicities, potentially exacerbating disparities in mental health treatment and outcomes.

The concern about cultural competence in AI extends to the system's ability to provide diverse and culturally appropriate responses without explicit prompting. MHLs worry that without built-in cultural sensitivity, AI systems might default to a narrow range of culturally homogeneous outputs, potentially alienating users from diverse backgrounds or failing to meet their specific needs. Kacey provides a concrete example of how a lack of cultural competence can manifest in AI systems:

> Unless I give it some … cues that I …desire a cultural background, it's only going to spit out a couple different genres of … content. And I didn't think that that from… a cultural and ethical standpoint might not be the best way of displaying the utilization of AI.

This observation underscores the need for AI systems to be inherently culturally aware rather than requiring explicit prompts to consider cultural factors. It highlights a critical shortcoming in current AI implementations, where cultural competence is often treated as an add-on feature rather than a fundamental aspect of the system's design and functionality. Furthermore, this perspective suggests that truly effective AI in mental health care should have cultural awareness deeply integrated into its core algorithms and knowledge base, enabling it to automatically provide culturally appropriate and sensitive responses across a wide range of diverse user interactions without the need for specific cultural cues or prompts.

While MHLs recognize the potential of AI in mental health care, they stress the critical importance of cultural competence in its development and implementation. They advocate for AI systems that are programmed and trained on diverse datasets, avoid stereotypes, and recognize different cultural beliefs and practices. The goal is to create tools that are truly effective and respectful for all users, regardless of their cultural background, ultimately enhancing mental health outcomes across diverse populations.

The participants expressed significant concerns regarding data privacy and the transparency of AI systems in mental health care. A primary focus is on the importance of informed consent and transparency in the use of AI. Sarah highlighted this fundamental aspect:

> For me, I think it comes back to informed consent, so it … needs to be all parties need to be aware of that it is being utilized, it's being implemented, and then there is consent from all parties that they're okay with it being part of it and again stressing that it may not be part of the service the interventions being provided, but it's still part of the service because it is used to … document the session.

This statement underscores the need for full disclosure and patient agreement in the use of AI in mental health care. It emphasizes that even when AI is used for administrative purposes like documentation, patients should be informed and consent obtained. This approach ensures transparency and maintains trust in the therapeutic relationship.

Beyond consent, MHLs are also concerned about the overall safety and security of patient data. The responsibility for ensuring data privacy and security in AI-assisted mental health care is a significant concern for MHLs. Benjamin articulated this sentiment: "We are responsible for the care we provide, so I really have to be reassured in regards to their concerns about privacy and safety." This statement highlights the sense of responsibility that mental health providers feel toward their patients' data security. It underscores the need for robust safeguards and assurances in AI systems to protect sensitive mental health information. This responsibility extends beyond just the implementation of AI to the ongoing management and security of patient data. The complexity of these concerns is further elaborated by other MHLs.

Sarah expressed the mixed feelings among the mental health community regarding AI adoption:

> I know that … there's definitely some interest. There's some interest, or maybe curiosity is a better word. Curiosity about it. But at the same time, there's still a lot of … cautiousness and … worries, especially around the breach.

This quote reflects the tension between the potential benefits of AI and the significant concerns about data security. It highlights that while there's curiosity and interest in AI's capabilities, there's also a strong undercurrent of caution, particularly regarding the potential for data breaches. This cautious approach underscores the need for robust security measures in AI systems. The specific nature of these security concerns is further detailed by other MHLs.

Jack elaborated on the complex issues surrounding data storage and protection in AI-assisted mental health care:

> I think there's going to be a variety of potential shortcomings or issues that could arise. I think one is probably security. Depending on how that AI is created or hosted, where is that data going? Is it held within the organization itself, or is it the AI system, or is it like the being that's owned by somebody else and all the data from the therapy center or therapy organization going back and forth between this other thing? And if so, how do we keep that information secure? How do we make sure that information is not being used for other purposes?

This statement highlights the multifaceted nature of data security concerns in AI-assisted mental health care. It raises critical questions about data ownership, storage locations, transfer methods, and access controls. These concerns extend beyond traditional patient confidentiality issues, encompassing the complexities introduced by AI systems that may be owned or operated by third parties. Addressing these concerns requires a comprehensive approach to data governance and security. The responsibility for implementing these measures is shared among various stakeholders.

Jennifer emphasized the shared responsibility between healthcare providers and AI developers in ensuring the beneficial and secure use of AI:

> And so, the responsibility obviously in the role of clinic and providing services … is there and then also the healthcare company being able to say, “Hey, this is what we are utilizing in a way that could be of benefit to our … client base and so let’s see how we can be better supportive in this way.”

This perspective underscores the collaborative nature of implementing AI in mental health care. It suggests that both healthcare providers and AI developers have roles to play in ensuring that AI technologies are used responsibly and effectively. This shared responsibility model emphasizes the need for ongoing dialogue and cooperation between clinical staff and technology providers.

However, even with such collaboration, MHLs recognize that AI systems are not infallible. Benjamin raised an additional concern about the potential for errors in AI systems: “Another potential could be just negligence. AI is not perfect at this point, from what I understand.” This statement highlights the recognition that AI systems, like any tool, are not infallible and may be prone to errors or oversights. It underscores the need for ongoing human oversight and the importance of viewing AI as a supportive tool rather than a replacement for human judgment in mental health care. This perspective reinforces the need for clear guidelines and regulatory frameworks to govern the use of AI in mental health settings.

The concerns raised by MHLs regarding privacy and transparency in AI-assisted mental health care are multifaceted and profound. They encompass issues of informed consent, data security, shared responsibility, and the potential for errors in AI systems. They stressed the importance of maintaining the central role of human expertise and professional ethics in patient care, even as AI technologies are adopted. These leaders emphasize the need for robust safeguards, clear ethical guidelines, and ongoing dialogue between all stakeholders, including mental health professionals, AI developers, and policymakers. The consensus is that while AI holds significant potential to enhance mental health care, its integration must be carried out with the utmost attention to patient privacy, data security, and transparency. As the field of mental health care continues to evolve with technological advancements, addressing these privacy and transparency concerns will be crucial in shaping responsible and effective AI integration practices.

The complexity of mental health work presents significant challenges in developing effective AI solutions, highlighting the need for more sophisticated and nuanced tools. Unlike many medical conditions that can be diagnosed and treated through straightforward algorithms, mental health issues often involve intricate, interrelated factors that defy simple categorization or solution (Joseph & Babu, 2024). As Lee aptly illustrated this complexity with an analogy, “The line of work we do is not always very simple and that it’s 1 plus 1 is 2. It’s like 1 plus 1 is 2 minus 3 divided by 5.”

This vivid description underscores the non-linear, often unpredictable nature of mental health treatment, where multiple factors interact in complex ways, making it challenging to create AI tools that can adequately capture and respond to the nuances of mental health care. As Benjamin further elaborated on the intricate process involved in mental health care:

> Oftentimes, processes happen while, as a therapist, after session, we tend to think and really look at the bigger picture of their current clients’ presentation. And there’s a lot of

> processes and conceptualization that goes into that. And we apply them as we experience with our clients.

This statement underscores the fluid and iterative nature of mental health treatment, where therapists continuously reassess and adapt their approach based on new insights gained from each client interaction. It emphasizes that mental health care is not a static, one-size-fits-all process but rather a dynamic journey that requires constant refinement and personalization. This ongoing adaptation, deeply rooted in the therapist's evolving understanding of the client's circumstances, presents a significant challenge for AI systems, which typically operate on more rigid, predefined algorithms and may struggle to replicate the nuanced, intuitive adjustments that human therapists make throughout the course of treatment.

Furthermore, Solaris emphasized the crucial role of human judgment in navigating these complexities, "Mental health care often involves understanding intricate personal histories and cultural contexts, nuances that AI cannot fully grasp. Human judgment remains crucial in these situations." This perspective underscores the limitations of AI in fully comprehending the depth and breadth of human experiences that inform mental health issues. While AI can process vast amounts of data and identify patterns, it currently lacks the capacity to truly understand the nuanced interplay of human emotions, personal histories, cultural contexts, and individual experiences that shape a person's mental health (Joseph & Babu, 2024; Mautang & Suarjana, 2023). Human clinicians, drawing on their empathy, cultural knowledge, and ability to interpret subtle cues, are uniquely equipped to navigate these complex, often intangible factors that are crucial in mental health care but remain challenging for AI to fully grasp and incorporate into its decision-making processes.

While AI holds promise for enhancing mental health care, the inherent complexity of the field poses significant challenges to its implementation. The non-linear nature of mental health challenges, the intricate human connections involved in mental health care, the critical role of ongoing clinical judgment, and the necessity for personalized, context-aware approaches all emphasize the limitations of current AI capabilities in this domain (Joseph & Babu, 2024).

The integration of AI in mental health care has sparked a robust debate among MHLs regarding its potential impact on patient-provider relationships. At the core of this discussion is the fundamental nature of mental health care, which has traditionally relied on human connection and empathy (Furnham & Sjokvist, 2017). Adam captured this sentiment:

> So, I know for the longest time, our field has emphasized, and it still holds true, the connection between a human to a human, which basically is the basis for therapy. The therapeutic alliance, the rapport building, the trust. All of those factors which I don’t think you can do. You can have that with an AI, even if it’s the most sophisticated form. So, that human-to-human direction, the what happens between two individuals. The actual physical and chemical changes that take place, which there’s so much research behind that, I don’t think an AI could duplicate that no matter what.

Adam’s statement underscores the irreplaceable nature of human interaction in mental health care, highlighting elements like rapport building and trust that are crucial to effective therapy. It emphasizes that the therapeutic alliance, built on empathy and personal connection, is a fundamental aspect of mental health treatment that current AI technologies cannot fully replicate. This perspective raises important questions about the limits of AI in replicating the nuanced interpersonal dynamics essential to mental health care.

Kacey further emphasized the complexity and artistry involved in mental health care, which poses significant challenges for AI implementation: “Mental healthcare, there … is data that … informs what we do in information but that it’s really sort of an art and that there are nuances in terms of what individual patients’ needs.” This observation highlights the intricate nature of mental health treatment, where data and information play a role but are complemented by the nuanced understanding and intuition of human practitioners. It suggests that while AI may excel at processing data, it may struggle to replicate the artistry and individualized approach that human clinicians bring to mental health care. This nuanced perspective leads us to consider how AI might be integrated effectively without compromising the essential human elements of care.

Benjamin offered a balanced view on how AI might be incorporated into mental health care:

> I always feel like the clinicians should be the ones in the driver’s seat. They’re the ones that are essentially making the call. They’re the ones that have been trained. AI is really there to support as a co-pilot rather than as the person that is driving the decisions.

Benjamin’s perspective suggests a model where AI complements rather than replaces human expertise, potentially enhancing the quality of care without compromising the essential human elements. In other words, Benjamin envisions AI as a supportive tool that can augment the capabilities of human clinicians, handling tasks such as data analysis or administrative work while leaving critical decision-making and therapeutic interactions to trained professionals.

Despite the concerns, some MHLs see potential benefits in AI integration. Solaris noted a positive aspect of AI implementation: “AI improves the quality of interactions that can significantly boost acceptance levels amongst patients.” This statement suggests that when implemented thoughtfully, AI could enhance the therapeutic relationship rather than diminish it.

It indicates that AI tools could support clinicians in providing more targeted and effective care, ultimately improving patient outcomes and satisfaction.

In conclusion, while the participants expressed valid concerns about the potential depersonalization of care through AI, they also recognized its potential benefits. The focus remains on maintaining the personal, empathetic care that is central to effective mental health treatment while leveraging AI to improve efficiency and expand access to services. Moving forward, the challenge lies in finding ways to integrate AI that enhance rather than replace the human elements of care. This delicate balance leads us to consider the next crucial aspect of AI integration in mental health: its compatibility with existing systems and workflows.

Integrating AI technologies into EHR systems and clinical workflows presents a significant challenge in adopting AI in mental health care. This integration is crucial for ensuring seamless adoption and maximizing the benefits of AI without disrupting established practices. Samantha highlighted the complexity of this process: “Integration teams work on incorporating AI into EHR and ensuring it fits organizational workflows.” This statement underscores the need for specialized teams dedicated to tailoring AI solutions to fit each organization’s technological ecosystem. It emphasizes that integrating AI into mental health care is not a one-size-fits-all process but rather requires a customized approach that considers individual institutions’ systems, workflows, and needs. The complexity of this integration process necessitates a team with diverse expertise, including knowledge of AI technologies, EHR systems, and the specific operational nuances of mental health care settings.

The multifaceted nature of AI integration extends beyond mere technical compatibility (Ahmed et al., 2023). Kacey elaborated on the various aspects involved in the implementation process: “Then, you have a separate set of people who are doing the actual integration, like integrating into your EHR figuring out the whole implementation side of things, making sure that this works for your organization based on your workflows."

Kacey's perspective emphasizes that successful AI integration involves both technological compatibility and alignment with organizational processes and workflows. It highlights the multidimensional nature of implementing AI in mental health care settings, where technical integration is only one piece of a larger puzzle. Kacey's statement further underscores the comprehensive approach ensures that AI solutions are not only technically sound but also practically viable within the specific context of each organization. Beyond technical and operational considerations, regulatory compliance remains a critical aspect of AI integration in healthcare (Sharma, 2020). Adam raises important questions regarding this crucial dimension, "How does this align with regulations laid down by the state, laid down by the county? How does it ensure compliance? How does it ensure confidentiality here with all the HIPAA laws and so on and so forth?"

Adam's statement highlights the need for AI solutions to integrate technically and comply with healthcare's legal and ethical standards. It underscores the importance of considering regulatory requirements throughout the integration process, ensuring that AI implementations adhere to privacy laws, data protection regulations, and other relevant healthcare standards. This regulatory alignment is crucial for maintaining trust in AI-enhanced mental health care systems and protecting patient rights (Al-Abdullah et al., 2020).

While the integration process involves multiple stakeholders and considerations, healthcare organizations retain significant agency in the adoption of AI technologies. Lee pointed out the role of organizational decision-making in this process:

> We, as an organization, actually have some authority to say we don’t want it. So, sometimes, they may push a new software update. Generally speaking, we would accept it, review it, and see how it impacts what our current flow is and if it does what we need to train staff on.

Lee’s statement emphasizes the importance of organizational autonomy in the adoption and integration of AI technologies. It underscores that healthcare organizations are not passive recipients of AI technology but active participants in its implementation, with the authority to evaluate and potentially reject AI solutions that do not align with their needs or standards. Furthermore, it highlights the crucial role of staff training in successful AI integration, suggesting that even technologically sound AI solutions may fail if the workforce is not adequately prepared to use them effectively in their daily practice.

The successful integration of AI into existing systems in mental health care requires a multifaceted approach that addresses technical, operational, regulatory, and organizational considerations. It demands collaboration between AI developers, IT specialists, mental health professionals, and organizational leadership to ensure that AI technologies not only function within existing systems but also enhance and streamline clinical practices without compromising the quality of care or data integrity. This complex integration process sets the stage for the next exploration: the organizational factors that influence MHLs in adopting AI technologies responsibly (Ahmed et al., 2023; Lu et al., 2023; Mendes-Santos et al., 2022).

## Results for Research Question 2

This section presents the interview findings related to the second research question: What organizational factors enable MHLs to adopt AI to improve patient care while minimizing risks? Focusing on the organization component of the TOE framework, this analysis explores the internal characteristics, structures, and processes at mental health organizations that influence their ability to implement AI responsibly. The findings shed light on various factors, including leadership approaches, organizational culture, resource allocation, staff training, and internal policies that contribute to AI's successful and ethical integration in mental health care. Examining these organizational elements reveals insights into how mental health institutions can create an environment conducive to AI adoption that maximizes patient benefits while carefully managing potential risks, ultimately helping MHLs navigate the complex landscape of AI adoption in mental health care.

**Theme 1: Organizational Culture and Resources for Ethical Alignment**

The participants acknowledged the challenges in addressing ethical dissonance as AI technologies become increasingly prevalent. The findings revealed two critical aspects of organizational culture that support responsible AI adoption: the establishment of psychological safety and ethical transparency in the workplace and the allocation of adequate resources for ethical governance.

The MHLs acknowledged the challenge is exacerbated when organizations lack awareness of ethical implications. As one participant described, "It's kind of like a free game, and no one has really told you to explain to you ethically or even how you use this." This statement highlights that when employees are not fully aware of AI use's ethical implications, they may inadvertently engage in practices that violate ethical standards. This lack of awareness can expose organizations to malpractice liability (Bal, 2008). In other words, the unintentional misuse of AI due to insufficient understanding of its ethical dimensions can have severe legal and reputational consequences for mental health organizations.

The MHLs indicated that organizational culture significantly influences responsible AI adoption. The participants emphasized that a responsible AI-adopting organizational culture is built on two fundamental pillars: psychological safety and resource allocation. These pillars are crucial in addressing the challenge of ethical dissonance that arises with the increasing prevalence of AI technologies.

The participants recognize the importance of creating an environment that fosters psychological safety and ethical transparency when adopting AI technologies. This approach is crucial in addressing the challenges that arise from the widespread availability and use of AI. Benjamin articulated this reality: “People will use it, as it is everywhere, and we cannot control that.” Benjamin’s statement highlights the ubiquitous nature of AI and the difficulty in completely controlling its use in an organization. This recognition underscores the need for a proactive approach in guiding AI adoption rather than attempting to restrict it entirely. By acknowledging this reality, leaders can focus on creating an environment that encourages responsible use and open dialogue about AI technologies.

Creating a culture of psychological safety is essential for encouraging open communication about AI use and its ethical implications (Clark, 2020). Jack emphasized the importance of this aspect: “There is a level of psychological safety where they can feel like they can come for it and ask those questions.” Jack’s observation underscores the value of an environment where employees feel comfortable seeking guidance and expressing concerns about AI technologies. This psychological safety enables a healthy feedback loop, allowing staff to report small mistakes, share insights, and contribute to the responsible adoption of AI without fear of judgment or reprisal (Clark, 2020).

To achieve this level of psychological safety and ethical transparency, organizational leadership must establish clear guidelines and policies (Clark, 2020). Jonathan highlighted the importance of a structured approach:

> I think it's important at a system level from the leaders to … have a structure in place around, what is our stance, for lack of a better word, on AI and incorporation of AI into our work as at a system level what processes are in place or policies are in place to help-help provide some guidelines or guardrails around what that looks like?

Jonathan's statement emphasizes the need for a systematic approach to AI adoption, with clear policies and processes in place. This structured framework provides employees with the necessary guidance to navigate the ethical complexities of AI use in mental health care, further enhancing psychological safety and ethical transparency at the organization.

While establishing guidelines is crucial, maintaining an open and optimistic stance toward AI adoption is equally important. Lee expressed this balanced approach: "I think right now we're cautiously optimistic would be the term. We definitely see the potential, and it has been somewhat useful because we haven't seen a lot of tools yet for mental health." Lee's perspective reflects a cautious optimism that allows organizations to explore AI's benefits while remaining mindful of potential challenges, fostering an environment where ethical considerations are at the forefront of AI adoption. Responsible AI adoption also requires a commitment to cultural competence and inclusivity. Jennifer articulated this important aspect: "We're very progressive in that aspect that our desire is that we … use AI again in a way that is responsible for that but then also along with cultural competence." Jennifer's statement highlights the importance of integrating cultural competence into AI adoption strategies. This approach ensures that AI technologies are implemented in a way that respects and addresses patients' and staff members' diverse needs, further enhancing the ethical foundation of AI use in mental health care.

Involving key stakeholders in the decision-making process is crucial for maintaining ethical transparency and psychological safety (Clark, 2020; Edmondson, 1999). Benjamin emphasized the value of this inclusive approach, "We also have really great clinicians in our academic researchers, administrators, and things like that, and they tend to be involved in decision-making, particularly if this is AI technology." Benjamin's observation underscores the importance of leveraging diverse expertise in AI-related decision-making. By involving clinicians, researchers, and administrators, organizations can ensure that AI adoption is guided by a comprehensive understanding of its implications for patient care, research, and organizational operations.

Fostering psychological safety and ethical transparency is crucial for the responsible adoption of AI in mental health care (Clark, 2020). Furthered by This approach requires clear guidelines, open communication, cultural competence, and inclusive decision-making processes (Edmondson, 1999). In other words, by creating an environment where staff feel safe to discuss AI-related concerns and contribute to ethical practices, organizations can navigate the complexities of AI adoption more effectively. Considering the importance of psychological safety and ethical transparency, it becomes clear that these efforts require adequate resources and support, leading to the next subtheme: resource allocation for ethical governance.

The participants recognize the critical importance of allocating resources for ethical governance in AI adoption. This allocation is essential for establishing structures and processes that support responsible AI implementation. Solaris emphasized the need for a diverse, expert-led approach: "Forming a multidisciplinary cross-functional ethics committee including mental health professionals, AI researchers, legal experts, and ethicists to guide AI implementation and manage ethical dilemmas." Solaris's suggestion highlights the importance of bringing together diverse expertise to address the complex ethical challenges of AI in mental health care. Such a committee can provide comprehensive oversight, ensuring that AI implementation aligns with ethical standards and professional best practices. This multidisciplinary approach can help organizations navigate the intricate landscape of AI ethics, fostering trust and accountability in AI adoption.

However, allocating resources for these initiatives often presents significant challenges, particularly when facing financial pressures. Kacey articulated this dilemma: "The challenge is finding the resources, especially when these initiatives do not immediately generate revenue." Kacey's statement underscores the tension between investing in long-term ethical governance and meeting short-term financial goals. It highlights the need for organizations to adopt a more holistic view of AI investments, considering immediate financial returns and long-term benefits such as improved patient outcomes and enhanced quality of care. This perspective shift is crucial for justifying and sustaining investments in ethical AI governance.

Despite these challenges, many organizations are taking proactive steps to allocate resources for AI governance. John describes one such initiative: "We have a big IT department. … I think they are planning to really have someone overlook the things that are being put in AI." John's observation reflects a growing recognition of the need for dedicated oversight in AI implementation. By allocating resources to create specialized roles or teams focused on AI governance, organizations can ensure that ethical considerations are consistently integrated into their AI strategies. This approach helps maintain accountability and alignment with organizational values throughout the AI adoption process.

The comprehensive nature of responsible AI adoption requires investments across multiple domains. Samantha elaborated on the diverse resource needs:

> There's a lot of training that gets involved. There's a lot of integration that gets involved, so the human training aspect of it is huge. You have to have enough people to understand what's going on to train them to continuously monitor that change management. And then you have your separate set of people who do quality improvement like, "Is this actually improving … our processes and efficiencies?" And then you have a separate set of people who are actually doing the actual integration, like, integrating into your HER, figuring out the whole implementation side of things, making sure that this works for your organization based on your workflows.

Samantha's detailed description highlights the multifaceted nature of resource allocation for ethical AI governance. It emphasizes the need for investments in training, integration, quality improvement, and technical implementation. This comprehensive approach ensures that AI adoption is technically sound, ethically grounded, and aligned with organizational workflows.

Jennifer further emphasized the substantial nature of these resource requirements: "It's going to take a tremendous amount of resources because … you're building the infrastructure both, both personally and personnel-wise, and then also for the operational aspect to make sure they're functioning." Jennifer's statement underscores the significant investment required for responsible AI adoption. It highlights that building the necessary infrastructure involves both technological resources and human capital and operational considerations. This holistic view of resource allocation is essential for creating a sustainable and ethically sound AI implementation strategy.

Resource allocation for ethical governance is a critical factor in responsible AI adoption at mental health organizations. It requires a multifaceted approach, encompassing diverse expertise, dedicated oversight, comprehensive training, and robust infrastructure development (Ahmed et al., 2023; Lu et al., 2023; Mendes-Santos et al., 2022). While financial challenges exist, the long-term benefits of ethical AI governance justify these investments. As organizations navigate these resource allocation decisions, they must also consider how these efforts impact and are perceived by their staff. This consideration leads us to our next theme, provider acceptance, which explores how mental health professionals view and adapt to AI integration in their practice.

**Theme 2: Early Adopters Influence Attitudes**

The integration of AI in mental health care has sparked diverse reactions among healthcare providers. While some enthusiastically embrace the technology, others approach it with caution. In this landscape, early adopters play a crucial role in shaping attitudes and paving the way for wider acceptance. Jonathan observed this phenomenon in their clinic: “There was a group in our clinic, early adopters, where they were sort of excited to try it out.” This statement highlights the presence of providers who are eager to explore AI’s potential in their practice as they serve as pioneers, testing the waters, demonstrating the technology’s practical applications, and influencing more hesitant colleagues.

The impact of these early adopters on their peers becomes evident as pilots and trials progress. Jonathan further noted, “There were actually lots of reservations from our primary care docs, but the pilot that did it, they loved it.” This observation underscores the transformative power of successful implementation. Initial skepticism among primary care doctors gave way to positive experiences, suggesting that hands-on exposure to AI can significantly shift attitudes. These early adopters' success can serve as a catalyst for broader acceptance at the organization.

The ripple effect of positive experiences extends beyond immediate participants, as Jonathan elaborated,

> I think the pilot that did it, they loved it. And … then the best way is when you hear it from someone else that you trust, and they're like, "Well, I did it, and I thought it was great." Now, others are sort of taking that on, and it's going to be a new function that's available for most all primary care providers, I think, by the end of the year.

This statement illustrates how peer influence can accelerate AI adoption. Trust in colleagues' experiences plays a significant role in overcoming initial hesitations, leading to a domino effect of acceptance. The planned expansion of AI functionality to all primary care providers by year-end reflects the growing confidence in the technology's benefits.

While early adopters' experiences can be influential, some providers maintain a cautious but open-minded approach. Benjamin expressed this perspective: "Again, those that are creating this app that does that can argue with me, and again, they might be able to change my mind. Again, I just need to be open to it and really learn it and see what that can do." Benjamin's statement reflects a willingness to engage with AI technology, albeit with a healthy skepticism. This attitude of openness to learning and potential perspective shifts is crucial for the gradual acceptance of AI in mental health care. It suggests that even those who are initially hesitant can be persuaded through education and demonstration of AI's capabilities.

The adoption of AI is not solely determined by individual willingness, as Jack points out:

> We'll have some early adopters that I think will just run with the idea. And then I think we'll have some that are maybe regulated by outside bodies where it's not just their

> willingness to adapt to it. They have to essentially get it cleared by whether it's insurance or other auditing bodies to see if they can really integrate it.

Jack's observation highlights the complex landscape of AI adoption, where regulatory and insurance considerations play a significant role. This complexity underscores the need for a comprehensive approach to AI implementation that addresses provider attitudes and external regulatory requirements.

Fostering acceptance of AI often requires a collaborative approach, as Benjamin suggested, "I think from what I understand and truly is just from my background experience is really bringing people together initially and really focusing on the work that we do." This statement emphasizes the importance of collective engagement in the AI adoption process. By focusing on the core work of mental health care, organizations can frame AI as a tool to enhance, rather than replace, current practices. This collaborative approach can help alleviate concerns and build consensus around AI implementation.

However, provider acceptance is intertwined with patient acceptance. Benjamin provided an example from a different technological adoption: "We incorporated videotaping because it's for training purposes, and you had resistance before. There are some clients who refuse to come to our clinic because they're like, 'I don't want to be recorded.'" This example illustrates how patient concerns can influence provider attitudes toward new technologies. It underscores the need for a holistic approach to AI adoption that considers both provider and patient perspectives.

In conclusion, early adopters play a pivotal role in shaping attitudes toward AI in mental health care. Their positive experiences can influence hesitant colleagues, creating a ripple effect of acceptance. However, successful integration requires addressing regulatory constraints, fostering open-mindedness, and considering patient perspectives. As organizations navigate these complexities, the role of leadership becomes increasingly important. This leads to the next theme, leadership characteristics, which pertains to how leaders can effectively guide their organizations through the challenges and opportunities of AI adoption in mental health care.

**Theme 3: Leadership Characteristics**

Effective leadership plays a crucial role in the responsible adoption of AI technologies in mental health organizations. The MHLs have identified several key characteristics that are essential for guiding their organizations through this complex process. These characteristics encompass a deep understanding of AI, strategic vision, and the ability to foster a supportive organizational culture.

The participants highlighted fundamental leadership characteristics as the need for a comprehensive understanding of AI and its potential impact on organizational operations. Adam emphasized this point, “The leadership first of all needs to be fully immersed to understand the depth of our operations would change if it was to come in.” This statement underscores the importance of leaders having an in-depth knowledge of AI technologies and their implications. This deep understanding enables leaders to anticipate and navigate the significant changes that AI adoption might bring to their organization’s operations. It also positions them to make informed decisions and provide guidance throughout the implementation process.

Beyond understanding, leaders must also set a clear vision and strategy for AI adoption. This strategic approach is essential for guiding the organization through the complex process of integrating AI technologies into mental health care practices. Solaris articulated this crucial aspect of leadership, emphasizing the need for a comprehensive framework that addresses both the technological and ethical dimensions of AI implementation: “Our leadership … sets a clear vision and strategy for AI adoption, prioritizing ethical implementation, data privacy, informed consent, continuous improvement, and human oversight.” Solaris’s observation highlights the multifaceted nature of effective AI leadership. By establishing a clear vision that prioritizes ethical considerations, data privacy, and continuous improvement, leaders can ensure that AI adoption aligns with the organization’s values and long-term goals. This strategic approach helps create a framework for responsible AI implementation that balances innovation with ethical considerations and patient care (Buckwalter & California Psychological Association, 2023).

Another critical leadership characteristic is the ability to foster psychological safety and open communication in the organization. Jack emphasized the importance of this supportive leadership style:

> I think leaders need to just be aware of what their staff do and have that intentional support not only from a practical standpoint of “Do you know what interventions your staff do and know enough about them to be able to talk about them?” But also, do you just provide support to your staff in such a way where they’re more willing to come forward and talk to you about something?

Jack’s statement underscores the need for leaders to be actively engaged with their staff’s work and to create an environment where employees feel comfortable discussing concerns or ideas related to AI adoption. This open communication fosters trust and can lead to more effective implementation of AI technologies, as staff feel supported in navigating changes and challenges.

Leadership’s role in shaping organizational culture and driving genuine adoption is another crucial aspect highlighted by MHLs. Sarah articulated this point:

> I think that’s … that’s huge in terms of like the organizational leadership who is exactly leading the charge what is their understanding of it. Like you can implement something,

> and people will just do it because they think that it has to be done but they're not really like wholeheartedly wanting to do it.

Chi's observation emphasizes the importance of leadership in driving the authentic adoption of AI technologies. Leaders must both implement AI and cultivate an organizational culture that genuinely embraces these technologies. This involves clear communication about the benefits and rationale behind AI adoption, addressing concerns, and inspiring enthusiasm among staff.

Effective leadership in the context of AI adoption in mental health organizations requires a combination of in-depth understanding, strategic vision, and the ability to foster a supportive and innovative organizational culture. Leaders must be fully immersed in understanding AI's implications, set clear ethical guidelines, provide intentional support to staff, and drive genuine adoption through effective communication and culture-building. By embodying these characteristics, leaders can guide their organizations through the complexities of AI adoption, ensuring that the implementation aligns with the organization's values and enhances its ability to provide quality mental health care.

## Results for Research Question 3

This section presents the findings related to the third research question: How do external factors affect MHLs' decisions to use AI for patient care? This analysis focuses on the environment component of the TOE framework, exploring the external elements that influence how MHLs approach AI adoption in their organizations. The environment in the TOE framework is a multifaceted arena outside of the organization: the industry, competitors, regulations, and relationships with the government. For mental health organizations considering AI adoption, this environment is a complex web of stakeholders, regulatory bodies, societal expectations, and market forces, each adding a layer of intricacy to the decision-making process.

**Theme 1: Government Policies and Professional Guidelines**

The participants recognize the critical role of government policies and professional guidelines in shaping the adoption of AI technologies in mental health care. While the regulatory landscape is still evolving, they emphasized aligning AI initiatives with existing frameworks to ensure legal compliance and maintain public trust. One participant succinctly captures this sentiment: “External factors include industry regulations and ethical guidelines, though these may be lacking.” This observation highlights the current state of AI regulation in healthcare, acknowledging both the presence of guidelines and the need for more comprehensive frameworks. As the field rapidly evolves, MHLs must navigate this complex landscape, balancing innovation with adherence to established norms and emerging best practices.

Despite the evolving nature of AI regulations, some MHLs note a growing acceptance of technology in mental health care. Kacey observed,

> Right now, I haven’t seen anything too negative. I think there [is] some more … willingness to embrace it because they also know that we are wearing. … They [have] all these … wearable stuff, … so people see the … practical side of things, and it’s also very cost-effective.

Kacey’s statement reflects a pragmatic approach to AI adoption, noting that the increasing prevalence of wearable technology has paved the way for greater acceptance of AI in mental health care. This practical perspective suggests that as AI demonstrates its utility and cost-effectiveness, it may face less resistance from both practitioners and regulators.

While there's growing acceptance, the participants stressed the importance of adhering to professional standards. Lee emphasized this point: "Yes, obviously, the mental health, we want to make sure it's within the BBS. Licensed psychologists, psychiatrists, medical professionals, boards, it's all approved technologies or tools that could be used." Lee's statement underscores the necessity of aligning AI technologies with established professional standards and approval processes. This approach ensures that AI tools meet the rigorous requirements set by licensing boards and professional organizations, maintaining the integrity of mental health care practices.

As MHLs consider the implementation of AI, they also recognize the need for a nuanced approach that considers patient needs and risk factors. Jack provides insight into this consideration,

> Because of that, there tends to be a high correlation with those receiving Medicaid and … seriously mentally ill level, where it's going to be a much higher intensity need. I really think that the area where AI should likely be initially adapted within mental health is those that might not be within the [seriously mentally ill] population because the risk of it going wrong is so much greater.

Jack's observation highlights the importance of a measured approach to AI implementation, particularly when dealing with vulnerable populations. This perspective suggests that initial AI adoption may be more appropriate for less severe cases, allowing for refinement of the technology before applying it to high-risk situations.

The landscape of government policies and professional guidelines for AI in mental health care is complex and evolving. Mental health leaders must navigate a path that balances innovation with compliance, leveraging the practical benefits of AI while adhering to established professional standards. As the field progresses, a thoughtful, risk-aware approach to AI implementation will be crucial, particularly when dealing with vulnerable populations. This careful consideration of regulatory and ethical frameworks sets the stage for our next theme: the influence of managed care on AI adoption in mental health services.

**Theme 2: Grant and Funding Source**

The integration of AI into mental health services is heavily influenced by economics and funding, as unanimously agreed by all participants. Community mental health organizations, particularly those that are government-funded, face substantial barriers due to resource constraints and the need for robust funding to invest in AI technology.

Historical precedent suggests potential challenges in AI adoption. One MHL noted, "[Electronic medical record and electronic health record] technology was conceived around the 1960s, yet we didn't get funding to implement it until the 2010s." This significant lag in funding for technological implementation may foreshadow a similar pattern with AI adoption. Given the current lack of specific AI regulations and guidelines at the federal and state levels, MHLs anticipate a delay in funding for implementation. As one participant succinctly stated, "I'm not sure if the federal government at the funding level has figured that out yet."

The disparity in funding capabilities raises significant equity concerns. Organizations with better financial resources can afford superior AI tools, potentially gaining advantages in client service, efficacy demonstration, and grant acquisition. This situation could create a self-reinforcing cycle of increased funding, exacerbating inequalities in mental health service provision.

When federal funding for AI implementation eventually becomes available, MHLs foresee new challenges related to performance expectations and future funding allocations. One participant observed, "With grants and funding, last year's performance is this year's baseline.

Funding is getting less, but we are required to provide more services." This reflects the double-edged nature of AI adoption: while it may enhance efficiency, it could also lead to unrealistic expectations from funders who might assume AI will reduce costs significantly or increase capabilities overnight.

The pressure to adopt AI to meet escalating performance demands could compel some organizations to implement AI hastily without thoroughly assessing its risks. Moreover, if AI-driven cost savings result in immediate funding reductions, it could exacerbate rather than alleviate resource constraints. One MHL articulated these concerns:

> One thing I am concerned about is we quickly follow suit hoping not to fall behind too far … those who have the funding and know how to use AI well will get a better competitive edge, and those without resources will fall further behind.

The landscape of AI adoption in mental health services is significantly shaped by funding disparities and economic considerations. While AI holds promise for improving service delivery, the uneven distribution of resources could widen the gap between well-funded and under-resourced organizations. As the field moves forward, addressing these funding inequities and carefully managing the expectations associated with AI implementation will be crucial to ensuring equitable access to advanced mental health care technologies.

**Theme 3: Managed Care**

The influence of managed care, particularly Medicare, on the adoption of AI in mental health services emerged as a significant theme among MHLs. As the largest payor in the United States, Medicare's policies often set the standard for the entire healthcare industry, including the integration of AI technologies. Samantha articulated this trend succinctly: "In the industry, a lot of times followed the big leader, which is Medicare. If Medicare moves one way, people—the industry—tends … to lean towards that."

This observation highlights Medicare's role in shaping industry-wide practices. Medicare approval of AI-driven tools and services often spurs widespread adoption in mental healthcare, influencing care providers and insurers to align their policies with Medicare standards (Finkelstein & McKnight, 2005; Rosenfeld et al., 2005).

The impact of managed care policies on service provision, especially for high-need populations, is a critical consideration for MHLs. Jack explained how this affects their approach to care: "So, because of that, to serve our population that we're passionate about, which is this high-need area of Medicaid patients, we've had to expand our servicing range to find additional ways to bring them in."

Jack's statement underscores the need for innovative approaches to care delivery, particularly for vulnerable populations. The adoption of AI technologies could be one such innovation, potentially expanding access and improving care for high-need patients. However, this adoption is contingent on managed care policies that support and reimburse these new approaches.

The potential for AI technologies to be included in reimbursable costs is an exciting prospect for many MHLs. Jennifer explores this possibility:

> And also, you look at, say, reimbursable costs with healthcare companies. And … with healthcare companies be able to say, "Hey, we're going to … have this as … one of our codes that you can as one of our available codes, and you'll be able to offset that cost of-of the Apple Vision Pro?"

Jennifer's speculation about the potential for AI technologies to be included in reimbursable codes highlights the transformative impact managed care policies could have on AI adoption. If advanced AI tools become reimbursable, it could significantly accelerate their integration into mental health care practices, potentially revolutionizing care delivery and accessibility.

The policies of managed care organizations, particularly Medicare, play a crucial role in shaping the landscape of AI adoption in mental health services. These policies influence what technologies are used and how care is delivered, especially to vulnerable populations. As AI continues to evolve, the alignment of managed care policies with these technological advancements will be critical in determining the extent and speed of AI integration in mental health care. The industry will be watching closely for signals from major payors like Medicare, which will likely set the tone for broader AI adoption and reimbursement practices.

**Theme 4: Competitors and Collaborators**

The adoption of AI in mental health organizations is driven by an interplay of competitive pressures and collaborative efforts. The participants recognize the need to innovate to remain relevant in an evolving landscape while also valuing shared knowledge and experiences. Solaris articulated the competitive aspect: "The competitive nature of the mobile health commercial space drives the adoption of AI to stay relevant and meet customer needs." This statement highlights how market forces are pushing organizations to embrace AI technologies. The desire to stay competitive and meet changing patient expectations is a significant driver for AI adoption, compelling organizations to innovate and adapt to the rapidly changing technological landscape.

Despite these competitive pressures, many participants emphasized collaboration and learning from peers. One leader expressed this collaborative sentiment: "Maybe we can talk to other organizations and see what worked and what didn't. So, we look out there and see who has done this." This approach underscores the value placed on shared experiences and best practices. By learning from others' successes and failures, organizations can make more informed decisions about AI adoption, potentially avoiding pitfalls and accelerating implementation. Some organizations have formalized these collaborative efforts. As another participant noted, "We are a part of a consortium where we get together and share ideas. Some board members are also leaders in other CMHOs." Such consortiums facilitate knowledge exchange and create a supportive network for navigating the challenges of AI adoption. This collaborative spirit suggests that many leaders view AI implementation as a shared challenge that benefits from collective wisdom and experience.

The rapid advancement of AI technologies can create a sense of urgency among organizations. John provides an example from their experience: "With Chat GPT, clinical staff and students started using it the second it came out, and our organization was like, 'This is going to be a huge problem.' I think that there was a push to create one internally." This observation illustrates how the swift adoption of AI by users can prompt organizations to rapidly develop their own solutions. It highlights the pressure to keep pace with technological advancements and user expectations, as well as the need to address potential challenges proactively.

The rapid adoption of AI can lead to concerns about competition, as Adam expressed:

> And then I also started thinking, if that was the case, "Why would other organizations not do the exact same thing?" Then was the competition. How do you stand up with against another agency who could be awarded for this? What would be the basis of that?

Adam's reflection reveals the competitive considerations that arise as organizations contemplate AI adoption. There is a recognition that falling behind in AI implementation could potentially disadvantage an organization in securing funding or attracting clients, driving a sense of urgency in adoption decisions.

However, not all MHLs view the landscape through a competitive lens. Jennifer offered a different perspective:

> I don't look at—and this is unique for me—I don't look at … people who are in the same space as … competition. I … look at them as individuals who … are extending … this space and its reach because if … one does well, then we all do.

Jennifer's view emphasizes the potential for collective advancement in the field of mental health care. This perspective suggests that successful AI implementation by one organization can benefit the entire sector by expanding capabilities and improving overall care quality, highlighting the interconnected nature of progress in mental health services.

The adoption of AI in mental health organizations is shaped by a delicate balance between competitive pressures and collaborative opportunities. While market forces drive innovation and rapid adoption, there is also a strong recognition of the value of shared knowledge and collective efforts. MHLs must navigate this landscape by balancing the need to stay competitive with the benefits of collaborative learning and advancement. As the field continues to evolve, fostering a culture that encourages both innovation and collaboration will be crucial for the successful and ethical implementation of AI in mental health care, ultimately benefiting both individual organizations and the sector as a whole.

**Theme 5: Patient Acceptance and Care Access**

The interviews revealed two key aspects of patient perspectives: patient receptivity and access to AI-enabled mental health services and the challenges of managing patient expectations while maintaining an appropriate balance between AI assistance and human care. The following sections explore patient trust, acceptance, concerns regarding AI tools, and the impact these technologies have on the patient experience, outcomes, and engagement with care. The findings emphasize the importance of addressing ethical considerations, ensuring transparency, and maintaining human connection in AI-assisted healthcare to ensure patients feel valued and supported. Understanding patient perspectives is crucial for the responsible and successful implementation of AI in healthcare settings.

The adoption of AI in mental health services is significantly influenced by patient receptivity and access to care. The participants observed that receptivity varies among different patient groups, with some being more open to AI-assisted interventions than others. Zach notes a particular group that shows promise: “The highest level of receptivity is going to be through those individual patients that are coming, finding us online, or walking into one of our offices on their own. I think they are going to be the most receptive.” This observation suggests that patients who proactively seek mental health services may be more open to innovative approaches, including AI-assisted care. Their initiative in seeking help could translate into a willingness to engage with new technologies, potentially making them ideal early adopters of AI-integrated mental health services.

However, access to mental health services remains a significant challenge for many patients, particularly in community health settings. Jonathan highlighted a common patient perspective:

> I think a lot of it is driven by what patients are asking for, which is it should not be that hard for me to get access to healthcare or schedule an appointment or to connect with my doctor those kinds of things.

Jonathan’s comment underscores the potential for AI to address fundamental access issues in mental health care. AI-powered solutions could streamline processes like appointment scheduling and doctor-patient communication, potentially reducing barriers to care.

The complexity of these access challenges is further illuminated by Benjamin’s observation: “Many of our patients are housing and food insecure. It is a lot of trying to navigate our complicated healthcare system while finding housing and making a living.” This statement highlights the broader socioeconomic factors that impact access to mental health care. For patients facing basic survival challenges, the integration of AI into mental health services must be considered within the context of these pressing needs. Given these diverse patient experiences, the participants emphasized involving patients and their communities in the AI adoption process. Benjamin articulated this sentiment: “Involvement of community members is crucial as their feedback and engagement will influence the effectiveness and acceptance of AI.” Benjamin’s statement underscores the value of patient perspectives in shaping AI implementation. By involving patients and community members, organizations can ensure that AI solutions address real needs and align with patient values.

The receptivity to AI in mental health care may vary by individual circumstances and the reputation and practices of the healthcare provider. John offered an insight from their experience:

> I feel like our patients, for example, would not be too concerned about it too much. I think it’s neither good [nor] bad. I think from the patients I’ve seen and interacted, like, our hospital does a lot of renowned medical procedures. People come from all over the nation, sometimes, to our hospital. I … think that they would have a positive effect towards us using it.

John's observation suggests that patients' trust in a healthcare institution may extend to their acceptance of AI technologies implemented by that institution. This points to the importance of institutional reputation and patient trust in facilitating AI adoption.

A significant factor in patient receptivity to AI is the generational divide in technology adoption. Benjamin highlighted this challenge:

> We have multi-generational levels of different types of clients. We have older folks who are more traditional who, to be honest with you, they don't even know how to log on to Zoom. We have to guide them through the process, and that's like a quarter of our session trying to get them there.

Benjamin's comment illuminates the practical challenges of implementing AI technologies across diverse age groups. The varying levels of technological literacy among patients necessitate a flexible approach to AI integration, potentially including additional support and education for older patients.

Patient receptivity and access to AI in mental health care are influenced by an interplay of factors, including individual initiative, socioeconomic circumstances, community involvement, institutional trust, and generational differences in technology adoption. Thus, MHLs must navigate these diverse considerations to ensure that AI implementation enhances rather than hinders access to care. As the field moves forward, a patient-centered approach that addresses these varied needs and perspectives will be crucial for the successful and equitable integration of AI in mental health services.

The integration of AI in mental health services brings both opportunities and challenges, particularly in managing patient expectations and maintaining a balance between AI efficiency and human interaction. The respondents expressed concerns about the potential impact of AI on patient expectations and the quality of care. Kacey highlighted a key concern regarding the immediacy of AI responses:

> Although I do see benefits for some members of the population that this is good because they get immediate responses, it also then sets it up where when they are actually with a human person in real life be very disappointing because in life things are not so quick to respond and you might get disappointed responses that you do not want to hear.

Kacey's observation underscores the potential for AI to create unrealistic expectations in human-to-human interactions. While AI can provide immediate responses, it may inadvertently set a standard for speed and availability that human providers cannot match, potentially leading to patient disappointment or frustration in traditional therapeutic settings.

Another concern the participants raised is the risk of AI-driven care becoming overly standardized or prescriptive. Kacey elaborated on this point:

> I think it could become too automated where it's like everyone wears a blue shirt medium-sized or when I go to sporting events, the … free giveaway shirts are all largest. Not an exact after Apple metaphor parallel example, but I think there is that fear of it becoming too prescriptive.

This metaphor illustrates the concern that AI might lead to a one-size-fits-all approach to mental health care, potentially overlooking patients' nuanced, individual needs. The fear of over-automation highlights the importance of maintaining human judgment and personalization in mental health services.

While acknowledging the potential benefits of AI, some MHLs emphasize the irreplaceable nature of human interaction in mental health care. Benjamin expressed this sentiment:

> Again, AI is not 100% proof from what I see. Someone else in AI could argue with me definitely, but I still say, even if they do, it's 99%. I still don't believe it. Human factors- our brain is so complex.

Benjamin's statement underscores the belief that the complexity of the human mind necessitates a human touch in mental health care. This perspective suggests that while AI can be a valuable tool, it should complement rather than replace human expertise and empathy.

To effectively integrate AI into mental health services, MHLs stress the importance of clinician understanding and engagement with AI tools. Jack articulated this need:

> I think we just have to know how the app works. And I think if something like that is created then the clinicians that are referring to it need to be really aware of what it works. And I think that maybe that's how we build it, or maybe it's how the person that owns it sells it to us, like, whatever information. But I also think it has to be something that we can just play around with, and clinicians can get a feel for it.

Jack's comment highlights the importance of clinician familiarity and comfort with AI tools. By ensuring that mental health professionals understand and can effectively use AI technologies, organizations can better manage patient expectations and maintain a balance between AI efficiency and human expertise.

Managing expectations and maintaining a human-AI balance in mental health services is a complex challenge. While AI offers benefits such as immediate responses and potentially increased access to care, it also risks creating unrealistic expectations and overly standardized approaches. MHLs emphasize the need for a thoughtful integration of AI that complements rather than replaces human interaction. As the field moves forward, finding ways to leverage

AI's efficiency while preserving the irreplaceable aspects of human care will be crucial for providing effective, personalized mental health services.

# Chapter Five: Discussion and Conclusion

The integration of AI in mental health care represents a transformative shift in healthcare delivery, offering unprecedented opportunities to enhance patient care while presenting significant challenges for implementation. This study explored how MHLs approach AI adoption through the TOE framework, providing insights into the interplay of factors that influence responsible AI implementation in mental health settings.

The research addressed three fundamental questions, each aligned with a component of the TOE framework:

1. What are the MHLs' perceptions of AI technologies in mental health care?
2. What organizational factors enable MHLs to adopt AI to improve patient care while minimizing risks?
3. What external factors affect MHLs' decisions to use AI for patient care?

This study's findings reveal a landscape characterized by both promise and caution, where technological capabilities must be balanced with ethical considerations and practical constraints. As Haber et al. (2024) emphasized, AI represents a fundamental paradigm shift with social and psychological implications for mental health care. This chapter synthesizes the research findings in Chapter Four and presents practical recommendations for organizations pursuing responsible AI adoption.

## Technology Component: MHLs' Perceptions and Implementation Strategies

The integration of AI technology in mental health care is notably shaped by perceptions and attitudes (Liehner et al., 2023). This study revealed a pattern of cautious optimism among leaders who recognize AI's potential while maintaining awareness of its limitations and risks. MHLs identified several promising applications of AI, particularly in enhancing personalized and holistic treatment approaches. According to Keaton (2022) and Moggia et al. (2024), holistic treatment in mental health requires considering an individual's overall well-being, including physical, emotional, social, and spiritual health, with AI offering the capability to personalize care based on individual genetic, biological, environmental, and psychosocial factors through data-driven strategies. MHLs envision AI as a tool to facilitate this comprehensive approach through sophisticated data analysis and pattern recognition. The technology's potential to identify connections between mental health and physical symptoms, lifestyle factors, and environmental influences represents a significant advancement in holistic care delivery.

The key concerns identified cluster into four critical areas that will be addressed in the subsequent recommendations. First, participants highlighted the potential inflexibility of AI systems and their limited capacity to grasp the nuanced complexities of mental healthcare. As one participant emphasized, "The line of work we do is not always very simple. … It's like 1 plus 1 is 2 minus 3 divided by 5." AI systems, by their nature, rely on data-driven algorithms that often lack the ability to interpret the subtleties of human emotions and psychological states (Graham et al., 2019). Second, there are significant concerns regarding algorithmic bias and fairness, particularly stemming from the lack of diverse representation in AI training datasets (Gupta, 2023).

According to Giovanola and Tiribelli (2022), bias in healthcare machine learning algorithms stemmed from model design (such as label biases and cohort biases), training data (such as minority bias, missing data bias, informativeness bias, and training serving skew), interactions with clinicians (such as automation bias, feedback loops, dismissal bias, and allocation discrepancy), and interactions with patients (such as privilege bias, informed mistrust and agency bias). Third, stakeholders raised issues surrounding privacy and transparency, specifically regarding the handling of sensitive patient data and informed-consent processes. Finally, while AI shows promise in reducing staff workload and burnout, MHLs expressed apprehension about potential over-reliance on these technological solutions and create oversight in diagnostics and treatment strategies.

## Technology Recommendations

Implementing AI in mental health care requires careful consideration beyond technical capabilities. Overall, AI represents a paradigm shift with social and psychological implications for the field (Haber et al., 2024). This study found that MHLs see significant potential in AI technology’s ability to transform services. The MHLs identified several promising applications of AI, including its capacity to deliver more personalized treatment through sophisticated data analysis. They particularly value AI’s potential to analyze large amounts of data to provide predictive and preventative care, thereby enabling earlier interventions and more targeted treatment approaches. Furthermore, AI’s ability to enhance accessibility to services while simultaneously reducing clinician burnout through automated administrative tasks represents a significant advancement. MHLs also recognize AI’s potential to improve compliance with funding guidelines and optimize fee collection processes, addressing crucial operational challenges in mental health care delivery.

However, MHLs expressed significant concerns about the implementation and use of AI technologies. A primary concern centers on the potential rigidity of AI approaches and their lack of nuanced understanding of the complexity of mental health work. They worry about biases and fairness issues arising from limited diversity in training data, which could lead to inequitable or inappropriate care recommendations for certain patient populations. Privacy, transparency, and security emerge as critical concerns, particularly regarding whether patients have adequate understanding and consent regarding how their data is being used. These concerns encompass broader issues of informed consent, data security, shared responsibility among stakeholders, and the potential for errors in AI systems that could impact patient care.

A fundamental concern expressed by MHLs is the potential impact on human-to-human interactions in mental health care. The findings strongly emphasize that organizations should not view AI as a replacement for human expertise but rather as a tool to enhance human capabilities. Maintaining a balance between the use of AI and human judgment is critical to ensuring ethical decision-making (Adelakun et al., 2024). Additionally, MHLs raised important technical and regulatory concerns about data interoperability and compliance with healthcare regulations, highlighting the need for robust systems that can meet stringent healthcare data management requirements.

Given these far-reaching implications, the MHL participants believe mental health organizations need a structured approach that addresses both the technical implementation of AI and its broader impact on care delivery and patient well-being. While this paper cannot address all potential implications of AI, it focuses specifically on addressing the primary concerns raised by the MHLs in this study. The recommendations in this chapter provide a practical framework for organizations to implement AI responsibly, addressing technological, organizational, and environmental considerations. The technological recommendations focus on three key areas: consumer-facing applications, administrative systems, and data security infrastructure. Each area requires specific safeguards and protocols to ensure ethical implementation while maximizing the benefits of AI integration in mental health services.

## Overall AI Systems Design Approach

The interviewed MHLs expressed significant concerns about AI implementation, including data integrity, algorithmic fairness, and security. While study participants lacked technical expertise, my analysis identified potential technological solutions like systems design frameworks and blockchain that may address their identified concerns, going beyond their non-technical recommendations. Blockchain technology for data storage could enhance data security and transparency, while careful systems design could help ensure algorithmic fairness and appropriate AI integration. However, these technical approaches should be viewed as partial solutions that require further research and validation rather than comprehensive answers to the complex challenges of AI adoption in mental healthcare.

### Inclusive AI Design Framework

A universal concern among the interviewees was whether AI training data sufficiently represents diverse cultural backgrounds and mental health experiences. Beyond HIPAA-secured, the foundation of responsible AI implementation in mental health care must be built upon principles of inclusive design that consider all potential users' diverse needs to mitigate data bias and algorithm bias (Powell & Kleiner, 2023). Regardless of whether organizations are implementing consumer-facing technology or administrative systems, they must prioritize inclusive design principles throughout the development or selection process of AI products. Mental health organizations must conduct comprehensive user research that deliberately includes traditionally underrepresented populations in mental health care, including racial and ethnic minorities, LGBTQIA+ individuals, neurodivergent users, people with disabilities, and individuals from various socioeconomic backgrounds. Organizations should employ participatory design methods where members of these communities are not merely subjects of research but active participants in the design process, ensuring that AI systems reflect diverse perspectives and experiences from the earliest stages of development.

The framework must extend beyond simple translation to incorporate true cultural adaptation of content and interfaces through multi-language support that considers cultural nuances, idioms, and mental health concepts that vary across different communities. This includes adapting clinical terminology, assessment tools, and therapeutic approaches to align with different cultural understandings of mental health. Organizations should implement rigorous testing protocols using digital twin technology to prevent potential harm before deployment, incorporating role-playing exercises and real-world testing with feedback from diverse user groups. Comprehensive accessibility features must be integrated as core components rather than additions, developing multiple access methods that accommodate different needs and preferences, including text-based interfaces, voice interaction systems with multiple accent recognition, video options with captioning in multiple languages, and adaptable interface designs for various cognitive styles.

A diverse community advisory board should be established to provide ongoing oversight and validation of AI features and functionality, including mental health professionals from diverse backgrounds, cultural community representatives, disability advocates, and service users from underrepresented groups. The board should have real authority to influence design decisions and implementation strategies, ensuring that inclusive design principles are maintained throughout the AI system’s life cycle. Organizations must establish continuous improvement processes that regularly assess and update AI systems based on user feedback and emerging needs, with clear metrics for measuring success across different demographic groups. Through this comprehensive approach to inclusive design, mental health organizations can ensure that their AI systems truly serve all members of their community while maintaining high standards of cultural competence and accessibility, ultimately enhancing the effectiveness and reach of AI-enabled services.

**Blockchain for Responsible Data Sharing and Data Sovereignty**

The majority of respondents expressed concerns about data transparency, security, and interoperability in AI implementation. Leveraging permission-based and shared blockchain technology as decentralized data storage can enhance data security, transparency, data ownership, and accountability, addressing several challenges related to data management and trustworthiness (Halamka et al., 2017).

Blockchain has had success in helping the country Estonia's e-Health Foundation launched blockchain technology to secure healthcare records in 2016 (Agbo et al., 2019; Andero, 2023). Blockchain technology emerges as a promising solution to address these challenges in both consumer-facing and administrative applications, and when done correctly, blockchain could enhance data security and ensure compliance with regulatory frameworks such as the GDPR (Al-Abdullah et al., 2020). Blockchain's decentralized architecture can provide a secure and transparent framework for managing sensitive mental health data while meeting organizational needs for efficient data management (Halamka et al., 2017).

Blockchain can create transparent, immutable audit trails of data usage through smart contracts and zero-knowledge proofs. Zero-knowledge proofs can solve the AI transparency problem by allowing one party, known as the prover, to convince another party, the verifier, that a particular statement is true without revealing any additional information beyond the validity of the statement itself (Bai et al., 2022). The essence of zero-knowledge proofs enables organizations to verify AI model training practices without accessing sensitive training data, and patients can verify how their data is being used in AI training without exposing their personal information. Through blockchain-enabled smart contracts, health organizations can provide patients the transparency and control over their health information with Dynamic Consent, where patients can grant or revoke access permissions in real time (Appenzeller et al., 2022; Hasan et al., 2021). This capability extends to AI implementations, allowing patients to make informed decisions about whether their data can be used for model training while maintaining their privacy through zero-knowledge proofs (Gaba et al., 2022).

Artificial intelligence systems in healthcare require interoperability to improve their accuracy through access to diverse datasets and reduce biases that may develop when operating in isolation. This is particularly crucial in mental health care, where patients often receive treatment across multiple providers and settings (e.g., community mental health centers, private therapists, and psychiatric hospitals) and where mental health conditions frequently co-occur with physical health issues. Without interoperability, each provider's AI system would need to start fresh in understanding patient patterns, potentially missing insights from previous care settings. Blockchain can standardize data exchange and enhance system interoperability, ultimately leading to better patient care (Durneva et al., 2020). As patients move between different providers or health systems, blockchain enables secure and seamless data portability while maintaining complete audit trails of data access and usage. This interoperability, combined with blockchain's inherent security features, helps organizations meet regulatory requirements for data protection while reducing administrative overhead. The result is a more efficient, transparent, and secure system that balances patient privacy rights with organizational needs for data access and AI development.

Although the participants did not explicitly bring up the issue of data colonization and data sovereignty, those technical issues can be contributing factors to data colonization, as discussed in Chapter Two. Blockchain can potentially help with decolonizing data by ensuring communities maintain control over their data and helping patients maintain data sovereignty (Martin et al., 2022). While data sovereignty is a critical consideration in AI implementation and one that the researcher views as essential to ethical AI adoption, this topic did not emerge organically from participant interviews. This absence may indicate an opportunity for future research exploring MHLs' perspectives on data sovereignty and its implications for AI adoption.

**Multilingual and Multi-modal Disclosure for Transparency**

The MHLs raised the question of how to effectively communicate AI use to patients presents a significant challenge in healthcare settings. While informed consent is standard practice for many AI consumer products, traditional disclosure methods often use complex language that can be difficult for patients to understand, particularly for non-English speakers. In the context of clinical decision support tools, the most common AI application in healthcare, organizations face the critical question of how to effectively disclose AI use and what information should be provided to patients (H. J. Park, 2024).

This study revealed the need for a more accessible and comprehensive approach to AI disclosure to patients. Durneva et al. (2020) noted that under GDPR requirements, patients have the right to understand how AI systems make automated decisions about their healthcare, requiring organizations to provide meaningful explanations of their AI's decision-making processes. Despite the lack of clear guidelines, the MHLs emphasized that organizations should implement clear, understandable disclosure protocols for both consumer-facing and administrative AI applications. This information should be presented in simple, accessible language that explains how AI systems use patient data, how this data is stored, and how it impacts patient care. The informed consent form should describe the role of human oversight in AI-assisted care, emphasizing that AI tools supplement rather than replace human clinical judgment while clearly communicating patient rights, including their options to opt out of AI-enhanced services. Organizations should maintain this transparency throughout the patient's entire journey, from initial marketing materials through intake processes and ongoing treatment documentation.

As the participants noted, patients need to understand both the potential benefits and limitations of AI-assisted care. To accommodate different cultural and language barriers, a recommendation is for organizations to use different languages and mediums, such as interactive videos or short-form videos, to explain the disclosure, making the information easy to understand. The findings of H. J. Park (2024) indicated that patient preferences for information regarding AI use vary significantly, necessitating tailored disclosure practices that reflect individual needs and concerns. According to Pandey et al. (2021), employing a variety of communication formats, such as interactive and short-form videos, is essential for organizations aiming to bridge cultural and language gaps. The implementation of these protocols must remain flexible and responsive to diverse patient needs. As Jack noted, different patient populations have varying levels of receptivity to AI tools; therefore, communication approaches should take technological literacy, cultural background, and individual preferences into consideration.

To keep up with the fast pace of AI technology, organizations should regularly review and update their disclosure protocols to reflect changes in AI capabilities, regulatory requirements, patient feedback, and emerging ethical considerations. Cao et al. (2023) emphasized the need for organizations to regularly update their disclosure protocols to align with the latest developments in AI and regulatory landscapes. This patient-centered approach, supported by dynamic and comprehensive disclosure practices, helps maintain trust while leveraging AI's potential to enhance access to and quality of mental health care.

**Consumer Technology: Safety and Human Support**

Many MHLs conveyed the concerns that AI may impact human-to-human connections during interventions; therefore, consumer technology user interface should be distinctly non-human looking to avoid misconceptions about AI's "humanness" and capabilities while clearly communicating that AI tools complement, rather than replace, human mental health providers. According to Maddali et al. (2022), AI interfaces should avoid human-like characteristics and instead focus on clear, functional design, especially for vulnerable populations such as individuals with dementia or psychosis. This approach helps users maintain realistic expectations about AI's capabilities and mitigate disappointments (Gorre et al., 2023). One MHL's observation that "the connection between a human to a human … is the basis for therapy" underscores the importance of maintaining the therapeutic alliance while integrating AI tools (Battaglia, 2019).

Organizations should establish clear protocols for situations where AI interactions indicate clinical risks or emergencies, ensuring seamless handoff to human providers when necessary, maintaining patient safety, and ensuring that appropriate care is delivered in critical situations. Di Sarno et al. (2024) discussed the integration of AI in pediatric emergency medicine and emphasized the need for well-defined procedures for escalating care to human providers when AI systems detect urgent clinical needs to ensure that the capabilities of AI are effectively complemented by human expertise. A graduated approach to AI implementation is recommended through a three-tier system. The first tier involves basic AI chatbots for initial screening and psychoeducation. This progresses to the second tier, which incorporates AI-augmented self-help tools with human monitoring. The third tier represents a hybrid care model combining AI analytics with human therapy. This comprehensive system should include real-time risk detection capabilities with automatic escalation to human providers through keyword monitoring and easily accessible help features. Regular assessment of therapeutic outcomes and client satisfaction ensures continuous improvement and accountability.

Mental health organizations must implement automatic handoff systems that trigger an immediate transition from AI to human providers when certain risk indicators are detected based on keywords identified. These transitions should occur when AI systems identify expressions of suicidal ideation, potential harm to self or others, severe psychological distress, acute psychotic symptoms, significant changes in mental status, complex trauma disclosures, medication-related emergencies, or situations involving domestic violence or abuse. The handoff process must be immediate and seamless from the patient’s perspective, with clear documentation requirements and automatic escalation to appropriate clinical staff, followed by a thorough review of all documentation by human providers.

**Administrative Technology: Addressing Over-Reliant on AI**

While AI technology is promising, MHLs are concerned about human clinical staff becoming overly reliant on AI diagnostic evaluation and treatment suggestions. As AI becomes increasingly integrated, there are growing concerns that clinicians’ dependency on it could decrease their ability to think critically and make independent decisions (Ali et al., 2024). The MHLs’ concerns are not unfound as clinicians will rapidly become less skilled, less attentive, and less discerning as AI becomes a more ubiquitous component of their clinical work (Adler-Milstein et al., 2024). To maintain clinical accuracy, accountability, and transparency, organizations can consider implementing AI watermarks similar to SynthID or scrambled sentences at random intervals to mandate clinician vigilance to review texts to make proper adjustments and corrections, require licensed clinicians to identify AI-generated text during document review, and ensure proper oversight before EHR approval. Regular accuracy audits and human override capabilities for all AI decisions provide additional safeguards against system errors or inappropriate automated responses. Real-time vigilance tracking to track clinicians' AI suggestions acceptance rate can be a good measure and can prompt education, feedback, coaching, and even turning off the AI for a period in serious cases (Adler-Milstein et al., 2024). The quality assurance framework should track key performance indicators, including documentation accuracy rates, system response times, and clinical efficiency metrics through monthly audits of AI-generated administrative documents and quarterly comprehensive analyses.

Organizations should continuously evaluate the impact of AI administrative systems through comprehensive monitoring that includes clinical operations metrics, patient satisfaction assessments, and compliance verification. Real-time tracking of system interactions, automated alerts for potential privacy breaches, and regular HIPAA compliance checks should be integrated into daily operations. Annual external audits provide independent verification of system performance, while quarterly risk assessments help identify and address potential vulnerabilities before they impact operations.

## Organization Component: Enabling Factors and Implementation Framework

The organizational factors enabling responsible AI adoption emerged as crucial determinants of implementation success. Rangavittal (2022) found that organizational culture and leadership play pivotal roles in AI adoption, with particular emphasis on ethical considerations and change management. In this study, MHLs emphasized that successful implementation requires more than technical capability—it demands systematic organizational change that addresses both structural and cultural elements.

A finding centers on the role of organizational culture in fostering ethical AI adoption. The research reveals that organizations lacking awareness of ethical implications may inadvertently engage in practices that violate ethical standards. This situation creates potential ethical dissonance, which Barkan et al. (2015) defined as the internal conflict experienced when actions do not align with personal or professional values. In professional contexts, this dissonance can occur when external forces, such as organizational policies or technological requirements, conflict with ethical standards (Lauria & Long, 2019).

The influence of early adopters emerges as a significant factor in shaping organizational attitudes. Early adopters can act as a positive force to promote and encourage the adoption of technologies by the mainstream majority (Latif, 2017). This peer influence proves particularly powerful in healthcare settings, where professional trust and credibility significantly impact adoption decisions. The research demonstrates that successful implementations often begin with a core group of enthusiastic practitioners who can effectively demonstrate the benefits and address their colleagues' concerns.

Education and training of the healthcare workforce represent another critical factor, often emerging as a significant barrier to AI adoption (Ahmed et al., 2023). MHLs emphasized the importance of comprehensive training programs that address both technical competency and ethical awareness. Furthermore, psychological safety emerges as a crucial element that promotes integrity and ethical behavior in the organization (Ferrère et al., 2022), creating an environment where staff feel comfortable voicing concerns and seeking guidance about AI implementation.

## Organizational Recommendations: ADKAR Framework Implementation

The research findings revealed that the participants emphasized the need for systematic organizational change to implement AI responsibly. They identified critical components for successful implementation, including the need for a clear understanding of AI's implications, building staff buy-in through early adopters, providing comprehensive training, and ensuring ongoing support with feedback mechanisms. These identified needs align naturally with the ADKAR change management framework of awareness, desire, knowledge, ability, and reinforcement (Zine et al., 2023). The ADKAR framework is a tool that companies use to make a change, increasing productivity and improving working conditions. The usage of Prosci's ADKAR model as the change management framework is ideal, as its main priority is people over process (DePodesta, 2024). Therefore, ADKAR provides an appropriate structure for organizing the organizational recommendations for AI implementation in mental health settings, addressing both the technical and human aspects of change that MHLs identified as crucial.

### *Awareness: Creating an Ethics-First AI Vision*

The awareness stage of ADKAR is crucial for laying the foundation for successful AI adoption in mental health organizations. According to Ball and Hiatt (2024), Awareness addresses fundamental questions about why change is needed and the risks of not changing. MHLs emphasized that organizations must focus first on developing a deep understanding of why change is necessary and how AI will impact care delivery. As one MHL noted, "First of all, there needs to be fully immersed to understand the depth of our operations would change if it was to come in." The Awareness phase should create an ethics-first AI vision, focusing on communicating how AI fits into the evolving landscape of mental health care, emphasizing its role in enhancing rather than replacing human clinical judgment while demonstrating the organization’s commitment to ethical implementation.

Organizations should establish regular communication channels through newsletters, meetings, and forums to share successful mental health AI implementation case studies and address ethical concerns using the principlism framework. This communication must highlight specific ways AI will address current service delivery challenges while maintaining the crucial human connection in mental health care. Leadership must maintain transparency about potential challenges and their mitigation strategies, fostering an environment where staff feel informed and included in the transformation process. To ensure effective awareness building, organizations should implement clear metrics for measuring progress and engagement, as staff understanding assessments and engagement levels provide insights into the effectiveness of awareness campaigns.

***Desire: Fostering Organizational Buy-In***

This study found that MHLs recognize the importance of fostering genuine willingness among staff to engage with AI technologies beyond mere compliance, and the findings revealed the crucial role of early adopters in creating organizational desire. Desire, according to (Ball & Hiatt, 2024), represents an individual’s personal choice to engage in and support the change based on their understanding of its personal impact and motivating factors. As one MHL observed,

> There was a group in our clinic, early adopters where they were sort of excited to try it out, the pilot that did it, they loved it. And then the best way is when you hear it from someone else that you trust.

Organizations can leverage early adopters as an AI champion network to engage with all potential users, not just to address operational concerns, but to strengthen their commitment to ethical principles in AI adoption. By aligning AI implementation with healthcare's established principlism framework (autonomy, beneficence, non-maleficence, and justice), organizations can help staff envision how AI can enhance patient care while maintaining ethical standards. This approach can be achieved by developing an AI champion network, which creates a foundation where early adopters serve as both technical ambassadors and ethical champions.

Developing an AI champions network begins with cultivating a genuine desire among staff to implement AI ethically and responsibly. The champions network's success depends on creating an environment that promotes both psychological safety and ethical innovation. Organizations should establish safe spaces for staff to experiment with AI tools while continuously evaluating their alignment with ethical principles. Mentorship programs should pair AI champions with hesitant colleagues, focusing not just on technical competency but on ethical decision-making. This requires clear selection criteria for AI champions that emphasize both technical aptitude and ethical awareness, along with protected time for innovation and ethical consideration. Regular champion network meetings should include discussions of ethical implications, while anonymous feedback mechanisms ensure continuous improvement in both technical and ethical aspects. Organizations should develop reward systems that specifically recognize ethical AI use, encouraging staff to prioritize responsible implementation that upholds mental health care's core values.

### ***Knowledge: Implementing Ethical AI Learning***

The knowledge stage of ADKAR focuses on providing staff with the specific information, training, and education needed to understand how to effectively implement AI in mental health care practice. The knowledge stage addresses how to change during the transition and how to perform effectively in the future state (Ball & Hiatt, 2024). The participants emphasized this critical need for comprehensive training and understanding. One highlighted, “There’s a lot of training that gets involved. … You have to have enough people to understand what’s going on to train them to continuously monitor that change management.” To address knowledge requirements, organizations should establish a comprehensive ethical AI learning framework that provides two essential types of knowledge: the user aspects of AI implementation and the ethical frameworks guiding its use. This dual focus aligns with the emphasis on both tactical knowledge (how to use the technology) and behavioral knowledge (how to work within new ethical guidelines). Organizations should develop tiered training programs that progress logically from fundamental AI concepts to advanced applications in mental health care, incorporating ethical decision-making protocols based on the principlism framework’s principles of autonomy, beneficence, non-maleficence, and justice.

The successful transfer of knowledge requires structured support systems, including a cross-functional ethics committee and regular workshops on AI ethics and governance. Following ADKAR’s framework guidance on knowledge transfer, organizations must invest in robust learning management systems to track progress, assess competency, and ensure that training translates into practical application. This systematic approach to knowledge building ensures staff acquire both the technical skills and ethical understanding needed to effectively implement AI in mental health care settings, emphasizing comprehensive knowledge development for successful change implementation.

***Ability: Fostering Competence Through Psychological Safety and Resource Support***

While the knowledge stage provides the foundation for understanding AI tools, the ability stage addresses the actual execution and development of proficiency in using these tools effectively (Ball & Hiatt, 2024). The findings in this research highlighted the importance of fostering psychological safety for staff, enabling them to develop competence in AI use in mental health settings. The MHLs emphasized creating an environment where staff feel comfortable asking questions and voicing concerns without fear of judgment. As one MHL explained, “There is a level of psychological safety where [staff] can feel like they can come forward and ask those questions.” highlighting the importance of a supportive environment for developing new capabilities.

To build ability effectively, organizations should implement a structured safe practice environment that allows staff to develop and demonstrate competence in using AI tools. This environment includes sandbox settings where staff can practice with AI applications without risk to patient care, converting theoretical knowledge into practical ability. Following Prosci’s emphasis on demonstrated capability, organizations should implement a graduated practice system that moves from basic AI interactions to more complex scenarios specific to mental health care. This includes real-time technical support, structured practice scenarios, and clear performance metrics to measure progress toward required competency levels.

In addition to psychological safety, resource allocation plays a critical role in the ability phase. Successful implementation requires access to technical support resources that can address staff questions and resolve issues in real time after the knowledge stage. Practice scenarios and simulations, specifically designed for mental health applications, provide structured opportunities for staff to gain practical experience in a safe environment. Performance support tools, such as step-by-step guides or quick-reference materials, should be readily accessible, allowing staff to navigate AI applications with confidence. Finally, establishing clear success metrics will help staff understand progress and competency goals, offering a concrete way to measure their development and the overall success of the organization’s AI initiatives.

By emphasizing psychological safety and ensuring ample resources, mental health organizations can foster the ability required for staff to work effectively with AI. This approach builds technical proficiency and nurtures a culture of openness and support, empowering staff to engage with AI tools confidently and responsibly.

***Reinforcement: Cultural Sustainability and Ethical Governance***

For a meaningful lasting change, reinforcement is essential to maintain new behaviors and ensure they are consistently supported. According to Ball and Hiatt (2024), reinforcement focuses on solidifying change in organizational culture by recognizing achievements, creating feedback loops, and addressing resistance. The participants have recognized the need for lasting change that includes ongoing support and feedback. As one MHL put it, “You can implement something, and people will just do it because they think it has to be done, but they’re not really wholeheartedly wanting to [follow ethical guidelines].” This statement highlights the importance of true engagement over mere compliance for sustainable change.

To establish a sustainable AI culture in mental health organizations, reinforcing ethical practices and governance is critical as AI becomes an integrated part of providing services. MHLs emphasize the importance of regular ethics audits and structured reviews to keep AI tools aligned with core organizational values and ethical standards. Continuous feedback loops enable staff to voice their experiences and concerns with AI systems, fostering a proactive approach to addressing any ethical or operational issues. Programs that recognize and reward ethical AI use can further embed these values, creating incentives for thoughtful engagement with AI tools. Specific measures, such as clear audit schedules, well-defined recognition programs, and updated performance metrics that include AI ethics, help reinforce these values daily. Regular cultural assessments ensure that commitment to ethical AI remains strong, allowing mental health organizations to build a resilient, ethically grounded culture that aligns with the vision of MHLs.

## ADKAR Critical Success Factor

This study identified four critical success factors for AI implementation: strong leadership commitment, accountability to uphold principlism bioethics, psychological safety in organizations, and sound resource management. These interconnected elements provide the foundation for responsible AI adoption in mental healthcare organizations.

### *Leadership Commitment*

According to Zona and Thaib (2020), a leader plays a very important role in managing relationships, coordinating change, adapting operational activities to strategies, building structures, and developing rewards. The success of AI implementation in mental health organizations is deeply rooted in strong leadership commitment. Visible executive sponsorship is essential, as it signals to staff that AI initiatives are both valued and prioritized at the highest levels. Leaders must maintain consistent ethical messaging, reinforcing a commitment to using AI in ways that respect the dignity, privacy, and autonomy of individuals. Furthermore, the allocation of resources, including budget and personnel, must be assured from the outset to sustain AI initiatives. Regular progress reviews by leadership provide accountability, ensuring AI projects adhere to ethical standards and achieve their intended impact.

### ***Ethical Framework Integration***

Integrating principlism ethical framework into AI practices is crucial for maintaining trust and accountability in the healthcare setting. Aligning AI initiatives with the ethical principles of principlism is a healthcare ethical framework that guides decision-making through four core principles: autonomy (patient's right to self-determination), beneficence (promoting patient good), non-maleficence (avoiding harm), and justice (ensuring fairness in care delivery) (Pasricha, 2022). Regular ethics committee meetings serve as a forum for discussing potential challenges and aligning AI use with organizational values. Documented decision-making protocols enhance transparency, allowing stakeholders to understand how ethical considerations influence AI-related decisions. Continuous ethical assessment ensures that AI applications remain sensitive to ethical issues as they evolve, reinforcing a responsible and principled approach to AI in mental health settings.

### ***Psychological Safety Measures***

Creating an environment of psychological safety is key to fostering open, honest engagement with AI tools. Anonymous feedback channels enable staff to voice concerns without fear of retribution, promoting transparency and early identification of potential issues (Awar et al., 2023). Organizations should establish a no-blame learning environment that encourages individuals to report and learn from mistakes or challenges in AI implementation. Clear escalation procedures provide a structured way to address ethical or operational concerns as they arise. Regular assessments of the organizational safety climate ensure that employees feel secure and supported, which is vital for the successful and responsible adoption of AI in mental health services.

### ***Resource Management***

Effective resource management is foundational for sustainable AI implementation. A dedicated budget for AI initiatives is necessary to support ongoing development, updates, and troubleshooting of AI systems (Marotta & Au, 2022). Organizations should protect time for training to ensure staff have the opportunity to develop the skills needed to interact with and utilize AI effectively. Organizations should also establish a strong technical support infrastructure that further aids in resolving issues and minimizing downtime and frustration. Finally, organizations should invest in ongoing AI-related education that allows the organization to keep pace with advancements in AI, ensuring that the workforce remains informed and equipped to use AI in safe and ethically sound ways.

## Environment Component: External Influences and Strategic Responses

The adoption of AI in mental health care is significantly influenced by a complex web of external factors that shape organizational decision-making and implementation strategies. This study found that MHLs must navigate a multifaceted environment encompassing regulatory requirements, funding constraints, and stakeholder expectations. In this environment, several key factors emerge as particularly influential in AI adoption decisions.

Government policies and professional guidelines serve as critical environmental factors shaping AI adoption (Neumann et al., 2022). MHLs emphasized the importance of aligning AI initiatives with regulatory frameworks, even as these frameworks continue to evolve. The current lack of specific AI guidelines has led many organizations to look to professional associations for guidance, creating a situation where implementation strategies must remain flexible enough to adapt to emerging regulatory requirements.

Funding and economic considerations emerged as significant environmental factors in program and technology implementation, particularly for community mental health organizations (Razzouk, 2023). Historical precedent suggests potential challenges in obtaining funding for technological implementation, as evidenced by one participant's observation about EHRs: "EHR [and electronic medical record] technology was conceived around the 1960s, yet we didn't get funding to implement it until the 2010s." This lag in funding access raises significant equity concerns (Zabelski et al., 2024), as organizations with greater financial resources can invest in more advanced AI tools, potentially widening the gap in service quality and accessibility between well-funded and resource-constrained organizations.

The influence of public managed care organizations, particularly Medicare, emerged as a crucial environmental factor. Medicare's role as the largest healthcare payer in the United States creates a ripple effect throughout the industry. As private insurers increasingly use AI for claims processing (Jaffe, 2023), community mental health organizations face both pressure and opportunities in their AI adoption journey. The growing prevalence of private insurance plans in providing Medicare services through programs like Medicare Advantage adds another layer of complexity to this dynamic (Thomson et al., 2020).

Patient perspectives and public perception also emerged as significant environmental factors influencing AI adoption decisions. MHLs observed varying levels of receptivity among different patient groups, with factors such as technological literacy, cultural background, and generational differences playing important roles in acceptance and engagement with AI-enhanced services.

## Environmental Strategy Recommendations

Based on these findings, I recommend a comprehensive approach to addressing environmental challenges that focuses on strategic engagement with external stakeholders and systematic resource management. Mental health organizations must develop adaptive strategies that balance AI implementation with regulatory compliance, resource limitations, and ethical care standards in an evolving healthcare environment. We identified patterns suggesting that successful AI adoption requires organizations to simultaneously navigate relationships with multiple stakeholders, including government agencies, funding sources, insurance providers, and patient communities. This approach encompasses three key areas: strategic funding and policy engagement, collaborative networks, and patient-centered implementation strategies. Each area addresses specific environmental challenges identified in this study while recognizing the interconnected nature of these challenges in mental health care settings. These recommendations are particularly relevant for community mental health organizations that must balance innovation with fiscal constraints and regulatory requirements. By addressing these three key areas comprehensively, organizations can build resilient systems that support responsible AI adoption while maintaining their commitment to accessible, high-quality mental health care.

## Strategic Funding and Regulatory Compliance

Organizations can consider establishing a public healthcare-aligned AI documentation initiative to address the challenges posed by Medicare and other public healthcare systems. This initiative should focus on creating standardized documentation processes that align with current requirements while maintaining flexibility to adapt to future changes. As major public healthcare systems like Medicare and Medicaid increasingly utilize AI for risk assessment and claims processing, community mental health organizations must align their AI implementation strategies accordingly. The research findings revealed that MHLs are particularly concerned about maintaining compliance while maximizing reimbursement opportunities.

This initiative begins with forming a dedicated Medicare compliance team to review current AI documentation requirements and create standardized templates that align with Medicare documentation rules. Organizations may want to systematically track common documentation challenges that could be addressed through AI solutions, providing data-driven justification for technology investments. Additionally, proactive discussions with EHR vendors about AI integration capabilities can help organizations prepare for future implementation requirements while ensuring system compatibility.

The MHLs in this study noted that historical delays in technology funding, such as with EHR implementation, require organizations to find creative ways to maximize limited resources while maintaining regulatory compliance. The implementation of blockchain technology may offer a promising solution for managing the complex requirements of multiple stakeholders in the mental healthcare environment (Mercer & Khurshid, 2021). This decentralized approach can help organizations meet Medicare and Medicaid documentation requirements while maintaining efficient operations. Through blockchain’s transparent audit trails (Regueiro et al., 2021), organizations can demonstrate compliance with regulatory requirements while reducing administrative overhead. The technology’s ability to create immutable records particularly benefits community mental health organizations dealing with strict public healthcare documentation requirements and multiple funding sources (Ettaloui et al., 2023). The benefits of this approach include reduced documentation errors and improved Medicare reimbursement rates through better compliance. Organizations will gain a clearer understanding of current and emerging compliance requirements, positioning them to adapt more quickly to changes in the public healthcare landscape. This proactive stance on documentation and compliance creates a strong foundation for future AI implementations while addressing one of the most significant environmental pressures community mental health organizations face.

**Community Mental Health Alliance Network**

The participants recognize that collaboration, not competition, is critical to successful AI implementation in mental health organizations. To facilitate this collaboration, one recommendation is to create a community mental health alliance network with two main objectives: fostering structured collaborative relationships between organizations and building diverse, unbiased AI training datasets through collective zero-knowledge proof data sharing.

The implementation begins by partnering with two to three organizations of different sizes. Organizations with existing alliances should establish dedicated AI subcommittees. Through monthly virtual meetings, partners share experiences and solutions, while a shared repository captures AI resources and lessons learned. This systematic documentation of challenges and solutions builds a collective knowledge base serving all participants.

The network delivers immediate benefits through shared learning experiences and cost reduction through resource sharing. It also creates a stronger collective voice for advocacy in the mental health sector. This collaborative structure proves particularly valuable for resource-limited community mental health organizations, enabling them to leverage shared expertise for more effective and responsible AI adoption. By working together, organizations avoid duplicating efforts and strengthen their overall approach to AI implementation.

**Patient Engagement and Public Trust-Building**

Patient engagement and public trust-building initiatives should form the third pillar of environmental strategy, requiring a thoughtful approach that acknowledges the sensitive nature of mental health care and the varying levels of AI receptivity among different patient populations. This study found that successful AI adoption depends significantly on patient acceptance and trust, particularly given the intimate nature of mental health treatment and the heightened privacy concerns in this field.

Organizations should implement a multilayered communication and engagement strategy that begins before AI implementation and continues throughout the service delivery process. At the pre-implementation stage, organizations should partner with community mental health advocacy groups like the National Alliance on Mental Illness to conduct community forums and focus groups that bring together diverse patient populations to understand their concerns, preferences, and needs regarding AI in mental health care. These discussions should specifically address cultural perspectives on mental health treatment, varying comfort levels with technology, and specific privacy concerns that may affect different communities.

Cultural competency must be embedded in AI implementation through ongoing dialogue with different community groups and adaptations to align with cultural norms and values. Organizations should provide tiered support systems offering varying levels of assistance in engaging with AI-enhanced services, ensuring that technological barriers do not create new disparities in access to mental health care. Through these engagement strategies, organizations can build the trust necessary for successful AI adoption while maintaining their commitment to accessible, high-quality mental health care for all patients.

The research suggests that successful navigation of environmental challenges requires organizations to maintain flexibility while building strong collaborative relationships. By developing robust documentation systems, fostering inter-organizational collaboration, and maintaining focus on patient needs, mental health organizations can better position themselves to leverage AI technologies while managing external constraints and pressures.

## Directions for Future Research

As the integration of AI in mental health care continues to evolve, several critical areas emerge as priorities for future research. These directions address the current gaps in understanding and pave the way for more effective and ethical implementation of AI technologies in mental health services.

### Long-Term Effects of AI Adoption on Patient Outcomes

A pressing need in the field is for longitudinal studies that assess the impact of AI-enhanced mental health care on patient outcomes over an extended period. While current research provides insights into the initial stages of AI implementation, the true value and potential risks of these technologies can only be fully understood through long-term observation and analysis. Future studies should track patients receiving AI-augmented care over several years, comparing their outcomes with those receiving traditional care. These longitudinal studies could examine various factors, including symptom reduction, recovery rates, relapse prevention, and overall quality of life improvements. Additionally, researchers should investigate potential unintended consequences or side effects of long-term exposure to AI-driven interventions. Such comprehensive, long-term studies would validate the presumed benefits of AI in mental health care and help in refining and optimizing AI systems for better patient care over time. The findings from these studies could guide policy decisions, inform best practices, and ensure that the integration of AI truly serves the best interests of patients in the long run.

## Comparative Studies Across Different Mental Health Settings

The effectiveness and challenges of AI adoption likely vary significantly across different mental health care settings. Future research should focus on comparative studies that examine AI implementation and outcomes across various contexts, such as inpatient facilities, outpatient clinics, community mental health centers, and private practices. These studies should investigate how factors like resource availability, patient populations, organizational structures, and technological infrastructure influence the adoption and efficacy of AI technologies. For instance, researchers could explore how AI-driven diagnostic tools perform in high-pressure emergency psychiatric settings compared to routine outpatient visits. Similarly, studies could examine the effectiveness of AI-assisted therapy in community clinics serving diverse populations versus specialized private practices. Such comparative research would provide crucial insights into the adaptability and scalability of AI solutions across different mental health care environments. The findings would be instrumental in developing tailored implementation strategies that account for each setting's challenges and opportunities. This nuanced understanding would enable MHLs and policymakers to make more informed decisions about AI adoption, ensuring that these technologies are deployed in ways that maximize benefits while minimizing risks across the entire spectrum of mental health care services.

## Development of AI-Specific Ethical Frameworks for Mental Health Care

As AI becomes more prevalent in mental health care, there is an urgent need for research focused on developing and refining ethical frameworks specifically tailored to this context. While general ethical guidelines for AI and healthcare exist, the sensitive nature of mental health data and the potential vulnerability of mental health patients necessitate specialized ethical considerations. Future studies should explore the ethical implications of AI use in areas such as patient privacy, informed consent in AI-assisted diagnoses, the potential for AI bias in mental health assessments, and the impact of AI on the therapeutic relationship. Researchers should collaborate with ethicists, mental health professionals, AI developers, legal experts, and patient advocates to create comprehensive ethical guidelines that address the nuances of AI applications in mental health. This research could involve case studies of ethical dilemmas in AI implementation, surveys of stakeholder perspectives on ethical issues, and the development and testing of ethical decision-making models for AI use in mental health care. The resulting frameworks should provide clear guidance on issues such as data handling, algorithm transparency, accountability for AI-driven decisions, and safeguarding patient autonomy. By establishing robust, context-specific ethical frameworks, this research would help ensure that the advancement of AI in mental health care aligns with core ethical principles and prioritizes patient well-being.

**Exploration of Patient Experiences With AI-Augmented Mental Health Services**

A critical area for future research lies in conducting in-depth studies of patient perspectives and experiences with AI in mental health care. While much of the current focus has been on the viewpoints of healthcare providers and the technical aspects of AI implementation, understanding the patient’s experience is crucial for developing truly patient-centered AI solutions. Future studies should employ a mix of qualitative and quantitative methods to explore how patients perceive, interact with, and benefit from AI technologies in their mental health care journey. Researchers could conduct interviews and focus groups to gather rich, detailed accounts of patient experiences with AI-assisted therapy, diagnostic tools, and monitoring systems.

Additionally, large-scale surveys could be developed to collect quantitative data on patient satisfaction, perceived effectiveness, and comfort levels with various AI applications in mental health care. These studies should investigate patient preferences regarding the balance between AI and human interaction in their care, their trust in AI-generated insights, and any concerns they may have about privacy or the depersonalization of care. Researchers should also examine how different patient demographics—considering factors such as age, cultural background, and type of mental health condition—might influence attitudes toward and experiences with AI in mental health care. The insights gained from these studies would be invaluable in informing the design and implementation of more acceptable, effective, and patient-centered AI interventions. By prioritizing the patient voice in AI development, this research would help ensure that technological advancements in mental health care truly align with patient needs, preferences, and values.

To summarize, these directions for future research—examining long-term patient outcomes, comparing AI adoption across different settings, developing specialized ethical frameworks, and exploring patient experiences—are critical for advancing our understanding of AI in mental health care. By pursuing these research avenues, the field can work toward more effective, ethical, and patient-centered integration of AI technologies in mental health services. As AI continues to evolve and shape the landscape of mental health care, ongoing research in these areas will be essential for realizing the full potential of AI to improve mental health outcomes while addressing challenges and ethical concerns. The findings from these future studies will contribute to the body of scientific knowledge and provide guidance for policymakers, healthcare providers, and technology developers in shaping the future of AI-augmented mental health care.

## Conclusion

The integration of AI in mental health care represents a transformative shift in the field, offering unprecedented opportunities to enhance patient care, improve accessibility, and potentially revolutionize treatment approaches. This study has explored the multifaceted landscape of AI adoption in mental health organizations through the lens of the TOE framework, providing insights into the perceptions, challenges, and potential pathways for responsible implementation.

Our examination of the MHLs' perceptions of AI technologies revealed a nuanced landscape characterized by cautious optimism. Leaders recognize the immense potential of AI to personalize treatment strategies, enhance predictive analytics for early intervention, and improve operational efficiency. The promise of AI in democratizing access to mental health support and potentially reducing societal stigma associated with mental health issues is particularly noteworthy. However, this optimism is tempered by significant concerns regarding the accuracy, reliability, and potential biases of AI systems. The complexity of mental health work, with its inherent nuances and the critical importance of human empathy, poses unique challenges for AI integration that cannot be overlooked.

The organizational factors enabling responsible AI adoption emerged as critical elements in this study. Our findings underscore the importance of cultivating an organizational culture that prioritizes ethical considerations, fosters psychological safety, and promotes transparent communication about AI initiatives. The role of leadership in setting a clear vision for AI adoption, allocating resources for ethical governance, and nurturing a culture of innovation cannot be overstated. The influence of early adopters in shaping organizational attitudes toward

AI highlights the potential for strategically leveraging these champions to facilitate broader acceptance and responsible implementation.

External factors significantly influence the decision-making process for AI adoption in mental health care. The evolving regulatory landscape, funding constraints, and the influence of managed care organizations all play crucial roles in shaping the environment in which AI is adopted. The competitive pressures and collaborative opportunities in the healthcare ecosystem create a complex backdrop against which organizations must navigate their AI strategies. Patient perspectives and public trust emerge as critical factors, emphasizing the need for transparent, ethical, and patient-centered approaches to AI implementation.

Looking to the future, several key areas demand further research and attention. Longitudinal studies examining the long-term effects of AI adoption on patient outcomes are essential to validate the presumed benefits and identify any unforeseen consequences. Comparative studies across different mental health settings will provide insights for tailoring implementation strategies to diverse contexts. Developing AI-specific ethical frameworks for mental health care remains a critical priority to ensure that AI adoption aligns with the core values and ethical standards of the mental health profession.

The practical recommendations outlined in this study, structured around the ADKAR change management framework, offer a roadmap for mental health organizations embarking on the journey of AI adoption. From creating awareness and fostering a desire for change to building knowledge, developing abilities, and reinforcing new practices, these guidelines provide a comprehensive approach to responsible AI implementation. The emphasis on maintaining human-centered care while leveraging AI's capabilities reflects a balanced approach that respects the nature of mental health work.

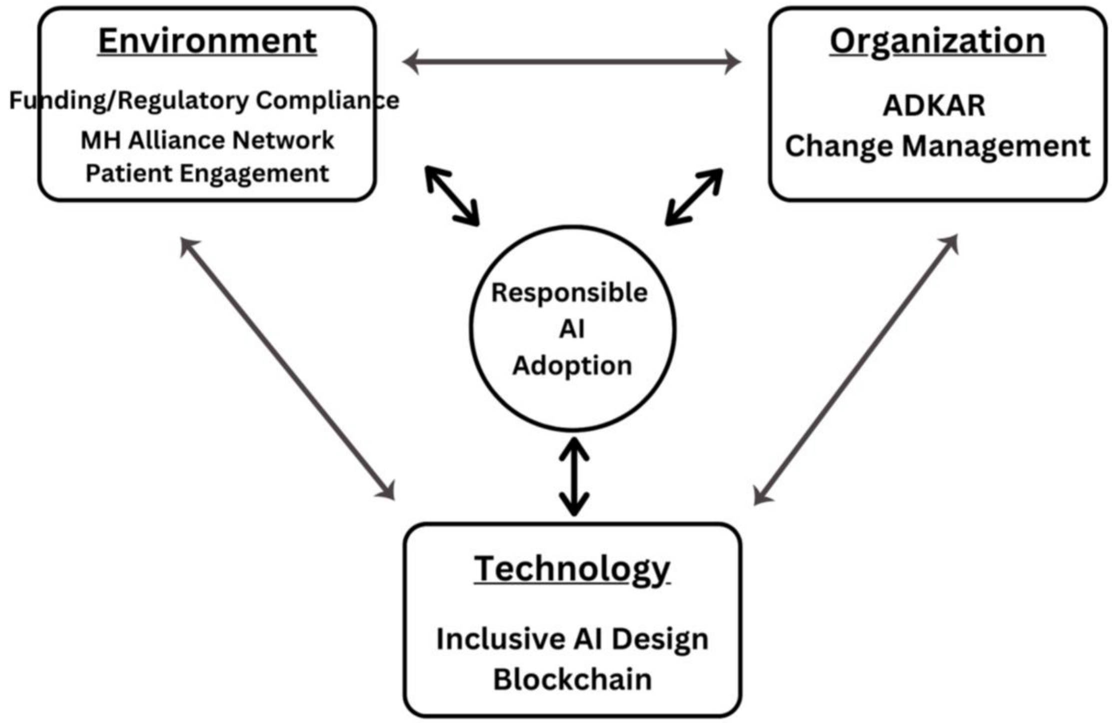
The adoption of AI in mental health care stands at a critical juncture. The potential benefits are immense, offering hope for more effective, accessible, and personalized mental health services. However, the challenges and ethical considerations are equally significant. Responsible AI adoption in mental health care is not merely a technological challenge but a multifaceted endeavor that requires careful navigation of organizational, ethical, and societal considerations.

Moving forward, it is imperative that the mental health community approaches AI adoption with a commitment to ethical practice, patient-centered care, and continuous learning. By fostering collaboration between mental health professionals, AI developers, policymakers, and patients can work toward a future where AI enhances and supports high-quality, ethical, and accessible mental health care. The journey ahead is complex, but with thoughtful implementation and ongoing research, AI could significantly improve mental health outcomes and contribute to a more effective and compassionate mental health care system.

The future of mental health care lies not in choosing between human expertise and AI but in finding the optimal synergy between the two. Continuing to explore and refine the role of AI in mental health care will require remaining grounded in the core values of the profession: empathy, ethical practice, and the unwavering commitment to improving the lives of those struggling with mental health challenges. Doing so will harness the power of AI to create a more responsive, effective, and inclusive mental health care system for all (Figure 2).

**Figure 2**

*Model for Responsible AI Adoption*



## References

Aali, G., Kariotis, T., & Shokraneh, F. (2020). Avatar therapy for people with schizophrenia or related disorders. *The Cochrane Library*, *2020*(5), Article CD011898 https://doi.org/10.1002/14651858.CD011898.pub2

Abd-Alrazaq, A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, Article 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review*, *23*(1), 1–33. https://doi.org/10.1016/s0272-7358(02)00146-0

Adams, D. (2024, January 30). Updating HIPAA security to respond to artificial intelligence. *Journal of AHIMA*. https://journal.ahima.org/page/updating-hipaa-security-to-respond-to-artificial-intelligence

Adelakun, N. B. O., Majekodunmi, N. T. G., & Akintoye, N. O. S. (2024). AI and ethical accounting: Navigating challenges and opportunities. *International Journal of Advanced Economics*, *6*(6), 224–241. https://doi.org/10.51594/ijae.v6i6.1230

Adigozel, O., Awad, N., Nayak, S., & Rabelo, I. (2023). *Are payers and providers ready for digital change in health care?* BCG Global. https://www.bcg.com/publications/2023/navigating-digital-healthcare-change

Adler-Milstein, J., Redelmeier, D. A., & Wachter, R. M. (2024). The limits of clinician vigilance as an AI safety bulwark. *Journal of the American Medical Association*, *331*(14), 1173–1174. https://doi.org/10.1001/jama.2024.3620

Agbo, C. C., Mahmoud, Q. H., & Eklund, J. M. (2019). Blockchain technology in healthcare: A systematic review. *Healthcare*, *7*(2), 56. https://doi.org/10.3390/healthcare7020056

Ahmed, M. I., Spooner, B., Isherwood, J., Lane, M., Orrock, E., & Dennison, A. (2023). A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus*, *15*(10), Article e46454. https://doi.org/10.7759/cureus.46454

Ahsan, S. M. M., Dhungel, A., Chowdhury, M., Hasan, M. S., & Hoque, T. (2024). *Hardware accelerators for artificial intelligence*. Cornell University. https://doi.org/10.48550/arxiv.2411.13717

Akbarighatar, P. (2024). Operationalizing responsible AI principles through responsible AI capabilities. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00524-4

Al-Abdullah, M., Alsmadi, I., AlAbdullah, R., & Farkas, B. (2020). Designing privacy-friendly data repositories: a framework for a blockchain that follows the GDPR. *Digital Policy Regulation and Governance*, *22*(5/6), 389–411. https://doi.org/10.1108/dprg-04-2020-0050

Alam, A., & Prybutok, V. R. (2024). Use of responsible artificial intelligence to predict health insurance claims in the USA using machine learning algorithms. *Exploration of Digital Health Technologies*, *2024*(2), 30–45. https://doi.org/10.37349/edht.2024.00009

Alami, H., Lehoux, P., Denis, J., Motulsky, A., Petitgand, C., Savoldelli, M., Rouquet, R., Gagnon, M., Roy, D., & Fortin, J. (2020). Organizational readiness for artificial intelligence in health care: Insights for decision-making and practice. *Journal of Health Organisation and Management*, *35*(1), 106–114. https://doi.org/10.1108/jhom-03-2020-0074

Alhur, A. (2024). Overcoming electronic medical records adoption challenges in Saudi Arabia. *Cureus*, *16*(2), Article e53827. https://doi.org/10.7759/cureus.53827

Alhuwaydi, A. (2024). Exploring the role of artificial intelligence in mental healthcare: Current trends and future directions – A narrative review for a comprehensive insight. *Risk Management and Healthcare Policy*, *17*, 1339–1348. https://doi.org/10.2147/rmhp.s461562

Ali, K., Garcia, A., & Vadsariya, A. (2024). Impact of the AI dependency revolution on both physical and mental health. *Journal of Strategic Innovation and Sustainability*, *19*(2). https://doi.org/10.33423/jsis.v19i2.7006

Allyn, B. (2023, July 2). Hollywood actors are pushing back against studios using AI to clone them. *NPR*. https://www.npr.org/2023/07/02/1185684635/hollywood-actors-are-pushing-back-against-studios-using-ai-to-clone-them

Alonso, S. G., De La Torre Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D. C., Nozaleda, L. M., & Franco, M. (2018). Data mining algorithms and techniques in mental health: A systematic review. *Journal of Medical Systems*, *42*(9), Article 161. https://doi.org/10.1007/s10916-018-1018-2

Alto, V. (2023). *Modern generative AI with ChatGPT and OpenAI models: Leverage the capabilities of OpenAI's LLM for productivity and innovation with GPT3 and GPT4*. Packt Publishing.

American Psychological Association. (2024, November 21). *Artificial intelligence in mental health care*. https://www.apa.org/practice/artificial-intelligence-mental-health-care

Andero. (2023, July 5). *How the world's leading blockchain company Guardtime turns trust into digital truth from its Tallinn office*. Invest in Estonia. https://investinestonia.com/how-the-worlds-leading-blockchain-company-guardtime-turns-trust-into-digital-truth-from-its-tallinn-office/

Anyanwu, E. C., Okongwu, C. C., Olorunsogo, T. O., Ayo-Farai, O., Osasona, F., & Daraojimba, O. D. (2024). Artificial intelligence in healthcare: A review of ethical dilemmas and practical applications. *International Medical Science Research Journal*, *4*(2), 126–140. https://doi.org/10.51594/imsrj.v4i2.755

Appenzeller, A., Hornung, M., Kadow, T., Krempel, E., & Beyerer, J. (2022). Sovereign digital consent through privacy impact quantification and dynamic consent. *Technologies*, *10*, 35. https://mdpi-res.com/technologies/technologies-10-00035/article_deploy/technologies-10-00035-v2.pdf?version=1645519830

Arigbabu, A. T., Olaniyi, O. O., Adigwe, C. S., Adebiyi, O. O., & Ajayi, S. A. (2024). Data governance in AI-enabled healthcare systems: A case of the project Nightingale. *Asian Journal of Research in Computer Science*, *17*(5), 85–107. https://doi.org/10.9734/ajrcos/2024/v17i5441

Atlam, H. F., Shafik, M., Kurugollu, F., & Elkelany, Y. (2022). Emotions in mental healthcare and psychological interventions: Towards an inventive emotions recognition framework using AI. In *Advances in Transdisciplinary Engineering*, *25*, 317–322 https://doi.org/10.3233/atde220609

Awar, D. T. A., Abdulla, F. I. M., Bakhamis, S. a. A., Rashid, M. a. A., Saleh, A. A., Mamalac, A. D., & Laja, N. (2023). Fostering a safe psychological environment and encouraging speak-up culture in primary care setups. *International Journal of Research in Medical Sciences*, *11*(12), 4583–4589. https://doi.org/10.18203/2320-6012.ijrms20233740

Bai, T., Hu, Y., He, J., Fan, H., & An, Z. (2022). Health-ZKIDM: a healthcare identity system based on fabric blockchain and Zero-Knowledge Proof. *Sensors*, *22*(20), Article 7716. https://doi.org/10.3390/s22207716

Baker, J. (2011). The technology–organization–environment framework. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), *Information systems theory: Explaining and predicting our digital society* (Vol. 1, pp. 231–245). Springer. https://doi.org/10.1007/978-1-4419-6108-2_12

Bal, S. B. (2008). An introduction to medical malpractice in the United States. *Clinical Orthopaedics and Related Research*, *467*(2), 339–347. https://doi.org/10.1007/s11999-008-0636-2

Ball, K., & Hiatt, J. (2024). *The ADKAR advantage: Your new lens for successful change*. Prosci Publications.

Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, *6*, 157–161. https://doi.org/10.1016/j.copsyc.2015.08.001

Battaglia, J. (2019). Supportive psychotherapy and importance of therapeutic alliance. *Psychiatric News*. https://doi.org/10.1176/appi.pn.2019.10b24

Baum, S. D. (2016). On the promotion of safe and socially beneficial artificial intelligence. *AI & Society*, *32*(4), 543–551. https://doi.org/10.1007/s00146-016-0677-0

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

Beauchamp, T. L., & Childress, J. F. (2012). *Principles of biomedical ethics* (7th ed.). Oxford University Press

Belani, S., Tiarks, G. C., Mookerjee, N., & Rajput, V. (2021). “I agree to disagree”: Comparative ethical and legal analysis of big data and genomics for privacy, consent, and ownership. *Cureus*, *13*(10), Article e18736. https://doi.org/10.7759/cureus.18736

Bench-Capon, T., Araszkiewicz, M., Ashley, K. D., Atkinson, K., Bex, F., Borges, F., Bourcier, D., Bourgine, P., Conrad, J. G., Francesconi, E., Gordon, T. F., Governatori, G., Leidner, J. L., Lewis, D., Loui, R. P., McCarty, L. T., Prakken, H., Schilder, F., Schweighofer, E., . . . Wyner, A. (2012). A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law. *Artificial Intelligence and Law, 20*(3), 215–319. https://doi.org/10.1007/s10506-012-9131-x

Bernardone, L. (2024, November). *‘Please die’: Google’s AI abuses grad student*. Information Age. https://ia.acs.org.au/article/2024/-please-die---google-s-ai-abuses-grad-student.html

Berre, A. J., Tsalgatidou, A., Francalanci, C., Ivanov, T., Pariente-Lobo, T., Ruiz-Saiz, R., Novalija, I., & Grobelnik, M. (2022). Big data and AI pipeline framework: Technology analysis from a benchmarking perspective. In E. Curry, S. Auer, A. J. Berre, A. Metzger, M. S. Perez & S. Zillner (Eds.), *Technologies and applications for big data value* (pp. 63–88). Springer https://doi.org/10.1007/978-3-030-78307-5_4

Bhargava, C., & Sharma, P. K. (2021). *Artificial intelligence: Fundamentals and Applications*. CRC Press.

Bhatia, K., Tanch, J., Chen, E. S., & Sarkar, I. N. (2020). Applying FAIR principles to improve data searchability of emergency department datasets: A case study for HCUP-SEDD. *Methods of Information in Medicine, 59*(01), 048–056. https://doi.org/10.1055/s-0040-1712510

Bickman, L. (2020). Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. *Administration and Policy in Mental Health*, *47*(5), 795–843. https://doi.org/10.1007/s10488-020-01065-8

Biswas, A., & Talukdar, W. (2024). Intelligent clinical documentation: Harnessing generative AI for patient-centric clinical note generation. *International Journal of Innovative Science and Research Technology*, *9*(5), 994–1008. https://doi.org/10.38124/ijisrt/ijisrt24may1483

BlackDeer, A. A., & Beeler, S. (2024). Decolonizing big data: addressing data colonialism in social work's grand challenges. *Journal of Ethnic & Cultural Diversity in Social Work*, 1–7. https://doi.org/10.1080/15313204.2024.2321440

Bordelon, B. (2023, June 21). *Schumer launches new phase in push for AI bill*. POLITICO. https://www.politico.com/news/2023/06/21/schumer-launches-new-phase-in-push-for-ai-bill-00102871

Borghouts, J., Eikey, E. V., Mark, G., De Leon, C., Schueller, S. M., Schneider, M., Stadnick, N. A., Zheng, K., Mukamel, D. B., & Sorkin, D. H. (2021). Barriers to and facilitators of user engagement with digital mental health interventions: Systematic review. *Journal of Medical Internet Research*, *23*(3), Article e24387. https://doi.org/10.2196/24387

Bostrom, N. (2015). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bradley, S. (2023, August 24). I tried the Woebot AI therapy app to see if it would help my anxiety and OCD. *Verywell Mind*. https://www.verywellmind.com/i-tried-woebot-ai-therapy-app-review-7569025

Bright, R. A. (2000). What is FDA doing about medical errors and adverse events related to medical devices? *Pharmacoepidemiology and Drug Safety*, *9*(5), 437–440. https://doi.org/10.1002/1099-1557(200009/10)9:5

Browne, G. (2022, October 1). The problem with mental health bots. *WIRED UK*. https://www.wired.co.uk/article/mental-health-chatbots

Buckwalter, J. G., & California Psychological Association. (2023, July 21). *Artificial intelligence (AI) and psychology: The ethics of the future are needed now* [Webinar]. Friday Webinar Series.

Bucky, S. F., Callan, J. E., & Stricker, G. (2013). *Ethical and legal issues for mental health professionals*. Routledge. https://doi.org/10.4324/9781315821160

Bukaty, P. (2019). *The California Consumer Privacy Act (CCPA)*. https://doi.org/10.2307/j.ctvjghvnn

Buruk, B., Ekmekçi, P. E., & Arda, B. (2020). A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Medicine Health Care and Philosophy*, *23*(3), 387–399. https://doi.org/10.1007/s11019-020-09948-1

Caetano, R. (2011). There is potential for cultural and social bias in DSM-V. *Addiction*, *106*(5), 885–887. https://doi.org/10.1111/j.1360-0443.2010.03308.x

Cahn, S. M., & Markie, P. J. (2011). *Ethics: History, theory, and contemporary issues*. Oxford University Press. http://ci.nii.ac.jp/ncid/BB07954360

Caldarini, G., Jaf, S., & Mcgarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, *13*(1), Article 41. https://doi.org/10.3390/info13010041

California Legislative Information. (2024a, September 28). *Bill Text - AB-3030 Health care services: artificial intelligence.* Retrieved December 9, 2024, from https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB3030

California Legislative Information. (2024b). *SB-1120 Health care coverage: utilization review.* https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1120

Cannarsa, M. (2021). Ethics guidelines for trustworthy AI. In L. A. DiMatteo, A. Janssen, P. Ortolani, F. de Elizalde, M. Cannarsa, & M. Dukovic (Eds.), *The Cambridge handbook of lawyering in the digital age* (pp. 283–297). Cambridge University Press. https://doi.org/10.1017/9781108936040.022

Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2023). How to talk when a machine is listening: Corporate disclosure in the age of AI. *Review of Financial Studies*, *36*(9), 3603–3642. https://doi.org/10.1093/rfs/hhad021

Carr, S. (2020). 'AI gone mental': Engagement and ethics in data-driven technology for mental health. *Journal of Mental Health*, *29*(2), 125–130. https://doi.org/10.1080/09638237.2020.1714011

Carta, S., Podda, A. S., Recupero, D. R., & Stanciu, M. M. (2022, October 4–8). *Explainable AI for financial forecasting* [Paper presentation]. Machine Learning, Optimization, and Data Science: 7th International Conference, Grasmere, UK. https://doi.org/10.1007/978-3-030-95470-3_5

Cecula, P., Yu, J., Dawoodbhoy, F. M., Delaney, J., Tan, J., Peacock, I., & Cox, B. (2021). Applications of artificial intelligence to improve patient flow on mental health inpatient units - Narrative literature review. *Heliyon*, *7*(4), Article e06626. https://doi.org/10.1016/j.heliyon.2021.e06626

Cesar, L. B., Callejo, M. Á. M., & Cira, C. (2023). BERT (Bidirectional Encoder Representations from Transformers) for Missing Data Imputation in Solar Irradiance Time Series. *Engineering Proceedings*, *39*(1), Article 26. https://doi.org/10.3390/engproc2023039026

Chatterjee, S., Rana, N. P., Dwivedi, Y. K., & Baabdullah, A. M. (2021). Understanding AI adoption in manufacturing and production firms using an integrated TAM-TOE model. *Technological Forecasting and Social Change*, *170,* Article 120880. https://doi.org/10.1016/j.techfore.2021.120880

Chen, Y., Huerta, E. A., Duarte, J., Harris, P., Katz, D. S., Neubauer, M. S., Diaz, D., Mokhtar, F., Kansal, R., Park, S. E., Kindratenko, V. V., Zhao, Z., & Rusack, R. (2022). A FAIR and AI-ready Higgs boson decay dataset. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-021-01109-0

Childress, J. F. (1983). *Principles of biomedical ethics*. Oxford University Press, USA.

Chiruvella, V., & Guddati, A. K. (2021). Ethical issues in patient data ownership. *Interactive Journal of Medical Research*, *10*(2), Article e22269. https://doi.org/10.2196/22269

Cirqueira, D., Helfert, M., & Bezbradica, M. (2021, July 24–29). *Towards design principles for user-centric explainable AI in fraud detection* [Paper presentation]. Artificial Intelligence in HCI: Second International Conference, online. https://doi.org/10.1007/978-3-030-77772-2_2

Clark, T. R. (2020). *The 4 stages of psychological safety: Defining the path to inclusion and innovation*. Berrett-Koehler Publishers.

Cogito Corporation. (2022, December 19). *Customer experience improvement*. https://cogitocorp.com/solutions/improve-my-cx/

Coglianese, C. (2002). Empirical analysis and administrative law. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.327520

CompanionMx. (2019, September 26). The U.S. Department of Defense selects CompanionMXTM for suicide prevention study in active duty naval personnel. *Business Wire*. https://www.businesswire.com/news/home/20190926005130/en/The-U.S.-Department-of-Defense-Selects-CompanionMx%E2%84%A2-for-Suicide-Prevention-Study-in-Active-Duty-Naval-Personnel

Cooper, M. (2023). Open Source is good for AI, but is AI good for Open Source? *Itnow*, *65*(2), 50–51. https://doi.org/10.1093/combul/bwad062

Cordina, J., Gilbert, G., Mph, N. G., & Kumar, R. (2019). *Next-generation member engagement during the care journey*. McKinsey & Company. https://www.mckinsey.com/industries/healthcare/our-insights/next-generation-member-engagement-during-the-care-journey

Couldry, N., & Mejias, U. A. (2018). Data colonialism: Rethinking big data’s relation to the contemporary subject. *Television & New Media*, *20*(4), 336–349. https://doi.org/10.1177/1527476418796632

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, Quantitative, and mixed methods approaches*. SAGE Publications.

Dakanalis, A., Wiederhold, B. K., & Riva, G. (2023). *Artificial intelligence: A game-changer for mental health care*. Cornell University. https://doi.org/10.31234/osf.io/ubdxz

Dale, S. (2023). A critique of principlism. *Voices in Bioethics, 9*. https://doi.org/10.52214/vib.v9i.10522

Dalvi, C., Rathod, M., Patil, S., Gite, S., & Kotecha, K. (2021). A survey of AI-based facial emotion recognition: Features, ML & DL techniques, age-wise datasets and future directions. *IEEE Access*, *9*, 165806–165840. https://doi.org/10.1109/access.2021.3131733

Dawoodbhoy, F. M., Delaney, J., Cecula, P., Yu, J., Peacock, I., Tan, J., & Cox, B. (2021). AI in patient flow: Applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units. *Heliyon*, *7*(5), Article e06993. https://doi.org/10.1016/j.heliyon.2021.e06993

DeAngelis, T. (2021, September 1). *Can real-world data lead to better interventions*. American Psychological Association. https://www.apa.org. https://www.apa.org/monitor/2021/09/news-real-world-data

De Carvalho, M. R., Freire, R. C., & Nardi, A. E. (2010). Virtual reality as a mechanism for exposure therapy. *World Journal of Biological Psychiatry*, *11*(2–2), 220–230. https://doi.org/10.3109/15622970802575985

Delahanty, R., Kaufman, D., & Jones, S. S. (2018). Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. *Critical Care Medicine*, *46*(6), e481–e488. https://doi.org/10.1097/ccm.0000000000003011

Del Re, A., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance–outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology Review*, *32*(7), 642–649. https://doi.org/10.1016/j.cpr.2012.07.002

Denecke, K., Gabarron, E., Grainger, R., Konstantinidis, S. T., Lau, A., Rivera-Romero, O., Miron-Shatz, T., & Merolli, M. (2019). Artificial intelligence for participatory health: Applications, impact, and future implications. *Yearbook of Medical Informatics*, *28*(01), 165–173. https://doi.org/10.1055/s-0039-1677902

DePodesta, M. (2024). The development of leadership communities of practice. *Nursing Administration Quarterly*, *48*(3), 225–233. https://doi.org/10.1097/naq.0000000000000648

DeSouza, D. D., Tang, S. X., & Danilewitz, M. (2022). The burgeoning role of speech and language assessment in schizophrenia spectrum disorders. *Psychological Medicine*, *53*(10), 4825–4826. https://doi.org/10.1017/s0033291722001325

DeSouza, D. D., Xu, M., Fidalgo, C., Robin, J., & Simpson, W. S. (2021). Psychometric approach to speech feature analysis as an objective measure of anxiety. *Biological Psychiatry*, *89*(9), S130. https://doi.org/10.1016/j.biopsych.2021.02.335

Diprose, W. K., Buist, N., Hua, N., Thurier, Q., Shand, G. B., & Robinson, R. (2020). Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *Journal of the American Medical Informatics Association*, *27*(4), 592–600. https://doi.org/10.1093/jamia/ocz229

Di Sarno, L., Caroselli, A., Tonin, G., Graglia, B., Pansini, V., Causio, F. A., Gatto, A., & Chiaretti, A. (2024). Artificial intelligence in pediatric emergency medicine: Applications, challenges, and future perspectives. *Biomedicines*, *12*(6), Article 1220. https://doi.org/10.3390/biomedicines12061220

Došilović, F. K., Brčić, M., & Hlupić, N. (2018, May 21–25). *Explainable artificial intelligence: A survey* [Paper presentation]. 41st International Convention on Information and

Communication Technology, Electronics and Microelectronics, Opatija, Croatia. https://doi.org/10.23919/MIPRO.2018.8400040

Dos Santos, P. I. G., & Rosinhas, A. C. (2023). Artificial intelligence and mental health. *Seven Editora Academica*. https://doi.org/10.56238/innovhealthknow-020

Durneva, P., Cousins, K., & Chen, M. (2020). The current state of research, challenges, and future research directions of blockchain technology in patient care: Systematic review *Journal of Medical Internet Research*, *22*(7), Article e18619. https://doi.org/10.2196/18619

Ebers, M. (2021). Regulating explainable AI in the European Union. An overview of the current legal framework(s). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3901732

Ebert, C., & Louridas, P. (2023). Generative AI for software practitioners. *IEEE Software*, *40*(4), 30–38. https://doi.org/10.1109/MS.2023.3265877

Edlich, A., Ip, F., & Whiteman, R. (2018). *How bots, algorithms, and artificial intelligence are reshaping the future of corporate support functions*. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-bots-algorithms-and-artificial-intelligence-are-reshaping-the-future-of-corporate-support-functions

Edwards, B. (2023, January 11). Controversy erupts over non-consensual AI mental health experiment [Updated]. *Ars Technica*. https://arstechnica.com/information-technology/2023/01/contoversy-erupts-over-non-consensual-ai-mental-health-experiment/

Ejjami, R. (2024). AI-driven healthcare in France. *International Journal for Multidisciplinary Research*, *6*(3). https://doi.org/10.36948/ijfmr.2024.v06i03.22936

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, *48*(1), 43–49. https://doi.org/10.1037/a0022187

Engle, R. L., Mohr, D. C., Holmes, S. K., Seibert, M. N., Afable, M. K., Leyson, J., & Meterko, M. (2019). Evidence-based practice and patient-centered care: Doing both well. *Health Care Management Review*, *46*(3), 174–184. https://doi.org/10.1097/hmr.0000000000000254

Enqvist, L. (2024). Rule-based versus AI-driven benefits allocation: GDPR and AIA legal implications and challenges for automation in public social security administration. *Information & Communications Technology Law*, *33*(2), 222–246. https://doi.org/10.1080/13600834.2024.2349835

Ettaloui, N., Arezki, S., & Gadi, T. (2023). An overview of Blockchain-based electronic health records and compliance with GDPR and HIPAA. *Data & Metadata*, *2*, Article 166. https://doi.org/10.56294/dm2023166

European Commission. (2021). *Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

European Commission, Joint Research Centre, Delipetrev, B., Chrysi, T., & Uros, K. (2020). *Historical evolution of artificial intelligence: Analysis of the three main paradigm shifts in AI*. Publications Office. https://doi.org/10.2760/801580

Evans, D. R., Cowey, H. D., Gliksman, L., Csapo, K., & Heseltine, G. F. D. (1976). The automated psychological evaluation system (APES). *Behavior Research Methods & Instrumentation*, *8*(2), 108–111. https://doi.org/10.3758/bf03201755

Familoni, N. B. T. (2024). Ethical frameworks for AI in healthcare entrepreneurship: A theoretical examination of challenges and approaches. *International Journal of Frontiers in Biology and Pharmacy Research*, *5*(1), 057–065. https://doi.org/10.53294/ijfbpr.2024.5.1.0032

Federal Trade Commission. (2021, June 22). *FTC finalizes order with Flo Health, a fertility-tracking app that shared sensitive health data with Facebook, Google, and others.* https://www.ftc.gov/news-events/news/press-releases/2021/06/ftc-finalizes-order-flo-health-fertility-tracking-app-shared-sensitive-health-data-facebook-google

Federal Trade Commission. (2023, October 5). *Protecting the privacy of health information: A baker's dozen takeaways from FTC cases.* https://www.ftc.gov/business-guidance/blog/2023/07/protecting-privacy-health-information-bakers-dozen-takeaways-ftc-cases

Federal Trade Commission. (2024a, May 6). *FTC gives final approval to order banning BetterHelp from sharing sensitive health data for advertising, requiring it to pay $7.8 million*. https://www.ftc.gov/news-events/news/press-releases/2023/07/ftc-gives-final-approval-order-banning-betterhelp-sharing-sensitive-health-data-advertising

Federal Trade Commission. (2024b, July 17). *Proposed FTC order will prohibit telehealth firm cerebral from using or disclosing sensitive data for advertising purposes, and require it to pay $7 million.* https://www.ftc.gov/news-events/news/press-releases/2024/04/proposed-ftc-order-will-prohibit-telehealth-firm-cerebral-using-or-disclosing-sensitive-data

Feijóo-García, P. G., Wrenn, C. G., Stuart, J., De Siqueira, A. G., & Lok, B. (2023). Participatory design of virtual humans for mental health support among North American computer science students: Voice, appearance, and the similarity-attraction Effect. *ACM Transactions on Applied Perception*, *20*(3), 1–27. https://doi.org/10.1145/3613961

Ferrère, A., Rider, C., Renerte, B., & Edmondson, A. (2022, June 7). Fostering ethical conduct through psychological safety. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/fostering-ethical-conduct-through-psychological-safety/

Ferryman, K. (2021). The dangers of data colonialism in precision Public Health. *Global Policy*, *12*(S6), 90–92. https://doi.org/10.1111/1758-5899.12953

Finkelstein, A., & McKnight, R. (2005). *What did Medicare do (and was it worth it)?* https://doi.org/10.3386/w11609

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical Implications of Embodied Artificial intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research*, *21*(5), Article e13216. https://doi.org/10.2196/13216

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities & Social Sciences Communications*, *7*(1), 10. https://doi.org/10.1057/s41599-020-0494-4

Food and Drugs Administration. (2021, January 12). *FDA releases artificial intelligence/machine learning action plan*. https://www.fda.gov/. Retrieved December 10, 2024, from https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan

Fox Rothschild LLP. (2019, October 3). *Data privacy and bias concerns in AI health tech*. https://hipaahealthlaw.foxrothschild.com/2019/10/articles/hit-health-information-technol/data-privacy-and-bias-concerns-in-ai-health-tech

Franco, M. P., Monfort, C., Piñas-Mesa, A., & Rincón, E. (2021). Could avatar therapy enhance mental health in chronic patients? A systematic review. *Electronics*, *10*(18), Article 2212. https://doi.org/10.3390/electronics10182212

Furnham, A., & Sjokvist, P. (2017). Empathy and mental health literacy. *HLRP Health Literacy Research and Practice*, *1*(2). https://doi.org/10.3928/24748307-20170328-01

Gaba, G. S., Hedabou, M., Kumar, P., Braeken, A., Liyanage, M., & Alazab, M. (2022). Zero knowledge proofs based authenticated key agreement protocol for sustainable healthcare. *Sustainable Cities and Society*, *80*, 103766. https://doi.org/10.1016/j.scs.2022.103766

Gangwar, H., Date, H., & Ramaswamy, R. (2015). Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *Journal of Enterprise Information Management*, *28*(1), 107–130. https://doi.org/10.1108/jeim-08-2013-0065

Gao, M. (2023, March 10–12). *Research on the application of artificial intelligence in night tourism* [Paper presentation]. 3rd International Conference on Public Management and Intelligent Society, Shanghai, China. https://doi.org/10.2991/978-94-6463-200-2_57

Geertz, C. (1973). *Thick description: Towards an interpretive theory of culture*. Basic Books. https://doi.org/10.4324/9780203931950-11

Gega, L. (2017). The virtues of virtual reality in exposure therapy. *The British Journal of Psychiatry*, *210*(4), 245–246. https://doi.org/10.1192/bjp.bp.116.193300

Ghosh, A., Chakraborty, D., & Law, A. (2018). Artificial intelligence in Internet of things. *CAAI Transactions on Intelligence Technology*, *3*(4), 208–218. https://doi.org/10.1049/trit.2018.1008

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, *178*(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763

Giovanola, B., & Tiribelli, S. (2022). Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & Society*, *38*(2), 549–563. https://doi.org/10.1007/s00146-022-01455-6

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://dl.acm.org/citation.cfm?id=3086952

Gorre, N., Fuhrman, J., Carranza, E., Li, H., Madduri, R. K., Giger, M. L., & El Naqa, I. (2023). *User experience evaluation for MIDRC AI interface* [Paper presentation]. SPIE Medical Imaging, San Diego, CA, United States. https://doi.org/10.1117/12.2651812

Götzl, C., Hiller, S., Rauschenberg, C., Schick, A., Fechtelpeter, J., Abaigar, U. F., Koppe, G., Durstewitz, D., Reininghaus, U., & Krumm, S. (2022). Artificial intelligence-informed mobile mental health apps for young people: A mixed-methods approach on users' and stakeholders' perspectives. *Child and Adolescent Psychiatry and Mental Health*, *16*(1). https://doi.org/10.1186/s13034-022-00522-6

Graham, S., Depp, C. A., Lee, E., Nebeker, C., Tu, X., Kim, H., & Jeste, D. V. (2019). Artificial Intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, *21*(11), Article 116. https://doi.org/10.1007/s11920-019-1094-0

Grundy, Q., Chiu, K., Held, F., Continella, A., Bero, L., & Holz, R. (2019). Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. *BMJ*, *364*, l920–l920. https://doi.org/10.1136/bmj.l920

Gugerty, L. (2006). Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 880–884. https://doi.org/10.1177/154193120605000904

Guo, B., Lou, Y., & Pérez-Castrillo, D. (2015). Investment, duration, and exit strategies for corporate and independent venture capital-backed start-ups. *Journal of Economics & Management Strategy*, *24*(2), 415–455. https://doi.org/10.1111/jems.12097

Gupta, N. (2023). Artificial intelligence ethics and fairness: A study to address bias and fairness issues in AI systems, and the ethical implications of AI applications. *Revista Review Index Journal of Multidisciplinary*, *3*(2), 24–35. https://doi.org/10.31305/rrijm2023.v03.n02.004

Haber, Y., Levkovich, I., Hadar-Shoval, D., & Elyoseph, Z. (2024). The artificial third: A broad view of the effects of introducing generative artificial intelligence on psychotherapy (Preprint). *JMIR Mental Health*, *11*, e54781. https://doi.org/10.2196/54781

Hadar-Shoval, D., Asraf, K., Mizrachi, Y., Haber, Y., & Elyoseph, Z. (2024). Assessing the alignment of large language models with human values for mental health integration: Cross-sectional study using Schwartz's theory of basic values. *JMIR Mental Health*, *11*, Article e55988. https://doi.org/10.2196/55988

Haenlein, M., & Kaplan, A. (2019). A Brief History of artificial intelligence: on the past, present, and future of artificial intelligence. *California Management Review*, *61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Hagendorff, T. (2024). Deception abilities emerged in large language models. *Proceedings of the National Academy of Sciences*, *121*(24). https://doi.org/10.1073/pnas.2317967121

Halamka, J. D., Lippman, A., & Ekblaw, A. (2017, May 18). The potential for blockchain to transform electronic health records. *Harvard Business Review*. https://hbr.org/2017/03/the-potential-for-blockchain-to-transform-electronic-health-records?ab=at_art_art_1x4_s03

Hardy, A., & Allnutt, H. (2023, March 23). *Replika, a 'virtual friendship' AI chatbot, receives GDPR ban and threatened fine from Italian regulator over child safety concerns.* Lexology. https://www.lexology.com/library/detail.aspx?g=d028a39d-e65b-4d20-b5a5-9cfa75a7addf

Hasan, H. R., Salah, K., Jayaraman, R., Yaqoob, I., Omar, M., & Ellahham, S. (2021). Blockchain-Enabled telehealth services using smart contracts. *IEEE Access*, *9*, 151944–151959. https://doi.org/10.1109/access.2021.3126025

Hatem, R., Simmons, B., & Thornton, J. E. (2023). A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. *Cureus*, *15*(9), Artcile e44720. https://doi.org/10.7759/cureus.44720

Haupt, C. E., & Marks, M. (2024). FTC regulation of AI-generated medical disinformation. *JAMA: The Journal of the American Medical Association*, *332*(23), 1975–1976. https://doi.org/10.1001/jama.2024.19971

Heart, T., Ben-Assuli, O., & Shabtai, I. (2016). A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. *Health Policy and Technology*, *6*(1), 20–25. https://doi.org/10.1016/j.hlpt.2016.08.002

Helpline Associates United [@HLAUnited]. (2023, May 26). *Helpline Associates United Statement on the Termination of the NEDA Helpline*. Twitter. https://twitter.com/HLAUnited/status/1662075335385526273/photo/1

Hendler, J. A. (2008). Avoiding another AI winter. *IEEE Intelligent Systems*, *23*(2), 2–4. https://doi.org/10.1109/mis.2008.20

Henry, S., Yetisgen, M., & Uzuner, O. (2021). Natural language processing in Mental Health research and practice. In *Computers in health care* (pp. 317–353). https://doi.org/10.1007/978-3-030-70558-9_13

Hilliard, A. (2024, December 3). *The state of healthcare AI regulations in the US*. Holistic AI. Retrieved December 11, 2024, from https://www.holisticai.com/blog/healthcare-laws-us#:~:text=Healthtech%20laws%20already%20in%20effect,assist%20users%20with%20basic%20tasks.

Hodge, J. G., & Gostin, K. G. (2004). Challenging themes in American Health information privacy and the 'public's health: Historical and modern assessments. *The Journal of Law Medicine & Ethics*, *32*(4), 670–679. https://doi.org/10.1111/j.1748-720x.2004.tb01972.x

Hodson, H. (2019, March 21). DeepMind and Google: The battle to control artificial intelligence. *The Economist*. https://www.economist.com/1843/2019/03/01/deepmind-and-google-the-battle-to-control-artificial-intelligence

Hoffman, S., & Podgurski, A. (2020). *Artificial intelligence and discrimination in health care* (Report Paper 2020-29). Case Western Reserve University School of Law. https://case.edu/law/sites/default/files/2021-01/Sharona%20CLE%202-2021_0.pdf

Hollis, C., Morriss, R., Martin, J., Amani, S., Cotton, R., Denis, M., & Lewis, S. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry*, *206*(4), 263–265. https://doi.org/10.1192/bjp.bp.113.142612

Holm, S. (2002). Principles of Biomedical Ethics, 5th edn.: Beauchamp T L, Childress J F. Oxford University Press, 2001, pound19.95, pp 454. ISBN 0-19-514332-9. *Journal of Medical Ethics*, *28*(5), 332. https://doi.org/10.1136/jme.28.5.332-a

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, *8*(1). https://doi.org/10.1177/2053951720982012

Hutson, E., & Melnyk, B. M. (2022). An adaptation of the COPE intervention for adolescent bullying victimization improved mental and physical health symptoms. *Journal of the American Psychiatric Nurses Association*, *28*(6), 433–443. https://doi.org/10.1177/10783903221127687

IEEE. (2019). *Artificial intelligence*. https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf

Inkarbekov, M., Monahan, R., & Pearlmutter, B. A. (2023). Visualization of AI systems in virtual reality: A comprehensive review. *International Journal of Advanced Computer Science and Applications*, *14*(8). https://doi.org/10.14569/ijacsa.2023.0140805

Ive, J. (2022). Leveraging the potential of synthetic text for AI in mental healthcare. *Frontiers in Digital Health*, *4*, Article 1010202. https://doi.org/10.3389/fdgth.2022.1010202

Jaffe, S. (2023, October 9). Denied by AI: Medicare Advantage’s “predictive” software cuts off care, say feds. *BenefitsPRO*. https://www.benefitspro.com/2023/10/09/feds-rein-in-use-of-predictive-software-that-limits-care-for-medicare-advantage-patients/?slreturn=20250120142917

Joerin, A., Rauws, M., Fulmer, R., & Black, V. (2020). Ethical artificial intelligence for digital health organizations. *Cureus*. https://doi.org/10.7759/cureus.7202

Jones, A. (2019, September 26). *The U.S. Department of Defense Selects CompanionMx*™ for suicide prevention study in active duty naval personnel. *Business Wire*. https://www.businesswire.com/news/home/20190926005130/en/The-U.S.-Department-of-Defense-Selects-CompanionMx%E2%84%A2-for-Suicide-Prevention-Study-in-Active-Duty-Naval-Personnel

Joseph, A. P., & Babu, A. (2024). The unseen dilemma of AI in mental healthcare. *AI & Society*. https://doi.org/10.1007/s00146-024-01937-9

Kandula, S., Ranga, S., & Moda, V. (2023). Hardware strategies for network optimization supporting AI workloads. *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets45389

Kargbo, R. B. (2023). Pioneering changes in psychiatry: biomarkers, psychedelics, and AI. *ACS Medicinal Chemistry Letters*, *14*(9), 1134–1137. https://doi.org/10.1021/acsmedchemlett.3c00333

Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal Endoscopy*, *92*(4), 807–812. https://doi.org/10.1016/j.gie.2020.06.040

Keaton, C. C. (2022). *The health in mental health: Exploring individualized holistic health practices as a process to transform mental health* (Publication No. 29331360) [Doctoral dissertation, The University of Texas at Arlington]. ProQuest Dissertations and Theses Global.

Kelly, L. (2024, December 27). *Third party data marketplaces: All you need to know 2024*. https://www.monda.ai/blog/third-party-data-marketplaces

Kerasidou, A. (2020). Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bulletin of the World Health Organization*, *98*(4), 245–250. https://doi.org/10.2471/blt.19.237198

Khante, N., & Hande, K. N. (2019). A survey to chatbot system with knowledge base database by using artificial intelligence & expert systems. *International Research Journal of Engineering and Technology*, *6*(05). https://www.irjet.net/archives/V6/i5/IRJET-V6I5153.pdf

Kimiagari, S., & Baei, F. (2021). Promoting e-banking actual usage: Mix of technology acceptance model and technology-organisation-environment framework. *Enterprise Information Systems*, *16*(8-9), Article 1894356. https://doi.org/10.1080/17517575.2021.1894356

Klang, E., Apakama, D., Abbott, E. E., Vaid, A., Lampert, J., Sakhuja, A., Freeman, R., Charney, A. W., Reich, D., Kraft, M., Nadkarni, G. N., & Glicksberg, B. S. (2024). A strategy for cost-effective large language model use at health system-scale. *NPJ Digital Medicine*, *7*(1), Article 30. https://doi.org/10.1038/s41746-024-01315-1

Klein, N. (2023, May 12). AI machines aren't 'hallucinating'. But their makers are. *The Guardian*. https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein?CMP=share_btn_tw

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085x.2021.1927212

Krager, D., & Krager, C. (2016). *HIPAA for health care professionals*. Cengage Learning.

Kreps, S. E., & Kriner, D. L. (2023). How AI threatens democracy. *Journal of Democracy*, *34*(4), 122–131. https://doi.org/10.1353/jod.2023.a907693

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. *Biomedical Informatics Insights*, *11*, 117822261982908. https://doi.org/10.1177/1178222619829083

Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A. E., Paszkiewicz, I., Nawrat, Z., Cebula, M., & Gruszczyńska, K. (2023). What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. *Diagnostics*, *13*(15), Article 2582. https://doi.org/10.3390/diagnostics13152582

Kuran, C. H. A., Morsut, C., Kruke, B. I., Krüger, M., Segnestam, L., Orru, K., Nævestad, T. O., Airola, M., Keränen, J., Gabel, F., Hansson, S., & Torpan, S. (2020). Vulnerability and vulnerable groups from an intersectionality perspective. *International Journal of Disaster Risk Reduction*, *50*, Article 101826. https://doi.org/10.1016/j.ijdrr.2020.101826

Laijawala, V., Aachaliya, A., Jatta, H., & Pinjarkar, V. (2020). Mental health prediction using data mining: A systematic review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3561661

Lamarche-Toloza, A. (2020). *Digital humans, virtual humans, digital doubles. . . what's the difference?* Virtuals - Creative Collective. https://virtuals.co/digital-humans-virtual-humans-differences-overview/

Laney, D. B. (2024, May 9). AI ethics essentials: Lawsuit over AI denial of healthcare. *Forbes*. https://www.forbes.com/sites/douglaslaney/2023/11/16/ai-ethics-essentials-lawsuit-over-ai-denial-of-healthcare/

LaRose, C. (2024, August 23). *1557 Final rule protects against bias in health care algorithms - National Health Law Program*. National Health Law Program. https://healthlaw.org/1557-final-rule-protects-against-bias-in-health-care-algorithms/

Latif, F. (2017). TELFest: an approach to encouraging the adoption of educational technologies. *Research in Learning Technology*, *25*(0). https://doi.org/10.25304/rlt.v25.1869

Lauria, M., & Long, M. F. (2019). Ethical dilemmas in professional planning practice in the United States. *Journal of the American Planning Association*, *85*(4), 393–404. https://doi.org/10.1080/01944363.2019.1627238

Lee, E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S., Kim, H., 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. https://doi.org/10.1016/j.bpsc.2021.02.001

Lee, S., Kang, J., Kim, H., Chung, K., Lee, D., & Yeo, J. (2024). *COCOA: CBT-based conversational counseling agent using memory specialized in cognitive distortions and dynamic prompt*. Cornell University. https://doi.org/10.48550/arxiv.2402.17546

Lee, Y., & Park, W. (2022). Diagnosis of depressive disorder model on facial expression based on Fast R-CNN. *Diagnostics*, *12*(2), 317. https://doi.org/10.3390/diagnostics12020317

Lenert, L., Lane, S., & Wehbe, R. M. (2023). Could an artificial intelligence approach to prior authorization be more human? *Journal of the American Medical Informatics Association*, *30*(5), 989–994. https://doi.org/10.1093/jamia/ocad016

Leone, D. (2021). *Data Colonialism in Canada: Decolonizing Data Through Indigenous data governance*. https://doi.org/10.22215/etd/2021-14697

Liehner, G. L., Hick, A., Biermann, H., Brauner, P., & Ziefle, M. (2023). Perceptions, attitudes and trust toward artificial intelligence — An assessment of the public opinion. *AHFE International*. https://doi.org/10.54941/ahfe1003271

Liu, D. (2021). machinesMemory: Malleability of AI technique, the data generated by machine learning algorithms. *Electronic Workshops in Computing*. https://doi.org/10.14236/ewic/eva2021.31

Long, K. M., & Meadows, G. N. (2017). Simulation modelling in mental health: A systematic review. *Journal of Simulation, 12*(1), 76–85. https://doi.org/10.1057/s41273-017-0062-0

Lu, T., Liu, X., Sun, J., Bao, Y., Schuller, B. W., Han, Y., & Lu, L. (2023). Bridging the gap between artificial intelligence and mental health. *Science Bulletin, 68*(15), 1606–1610. https://doi.org/10.1016/j.scib.2023.07.015

Luxton, D. D. (2016). An introduction to artificial intelligence in behavioral and mental health care. In D. D. Luxton (Ed.), *Artificial intelligence in behavioral and mental health care* (pp. 1–26). Elsevier https://doi.org/10.1016/b978-0-12-420248-1.00001-5

Ma, X., & Jiang, C. (2023). On the ethical risks of artificial intelligence applications in education and its avoidance strategies. *Journal of Education Humanities and Social Sciences, 14*, 354–359. https://doi.org/10.54097/ehss.v14i.8868

Maddali, H. T., Dixon, E., Pradhan, A., & Lazar, A. (2022, April 28). I*nvestigating the potential of artificial intelligence powered interfaces to support different types of memory for people with dementia* [Paper presentation]. CHI Conference on Human Factors in Computing Systems, New Orleans, LA, United States.

Margam, R. (2023). ChatGPT: the silent partner in healthcare. *The Review of Contemporary Scientific and Academic Studies*, *3*(10). https://doi.org/10.55454/rcsas.3.10.2023.005

Marks, M., & Haupt, C. E. (2023). AI chatbots, health privacy, and challenges to HIPAA compliance. *JAMA*, *330*(4), 309–310. https://doi.org/10.1001/jama.2023.9458

Marotta, G., & Au, C. (2022). Budgeting in the age of artificial intelligence – new approaches, old challenges? *International Journal of Artificial Intelligence and Machine Learning*, *2*(2), 1–11. https://doi.org/10.51483/ijaiml.2.2.2022.1-11

Marr, B. (2023, July 6). AI in mental health: Opportunities and challenges in developing intelligent digital therapies. *Forbes*. https://www.forbes.com/sites/bernardmarr/2023/07/06/ai-in-mental-health-opportunities-and-challenges-in-developing-intelligent-digital-therapies/?sh=7e5dc2c45e10

Martin, A., Sharma, G., De Souza, S. P., Taylor, L., Van Eerd, B., McDonald, S. M., Marelli, M., Cheesman, M., Scheel, S., & Dijstelbloem, H. (2022). Digitisation and sovereignty in humanitarian space: Technologies, territories and tensions. *Geopolitics*, *28*(3), 1362–1397. https://doi.org/10.1080/14650045.2022.2047468

Martínez-Castaño, R., Htait, A., Azzopardi, L., & Moshfeghi, Y. (2021, September 21–24). BERT-based transformers for early detection of mental health illnesses [Paper presentation] 12th International Conference of the CLEF Association, online. https://doi.org/10.1007/978-3-030-85251-1_15

Masche, J., & Le, N. (2017, June 30–July 1). *A review of technologies for conversational systems* [Paper presentation]. 5th International Conference on Computer Science, Applied Mathematics and Applications. https://doi.org/10.1007/978-3-319-61911-8_19

Mautang, T. W. E., & Suarjana, I. W. G. (2023). The global and cultural context of using AI for mental health. *Journal of Public Health*, *46*(2), e343. https://doi.org/10.1093/pubmed/fdad262

May, J. (2018). *Regard for reason in the moral mind*. Oxford University Press.

Mayover, T. (2024, October 2). When AI technology and HIPAA collide. *The HIPAA Journal*. https://www.hipaajournal.com/when-ai-technology-and-hipaa-collide/

Mazziotti, R., & Rutigliano, G. (2020). Tele–Mental Health for Reaching Out to Patients in a Time of Pandemic: Provider Survey and Meta-analysis of Patient Satisfaction (Preprint). *JMIR Mental Health*, *8*(7). https://doi.org/10.2196/preprints.26187

McCradden, M. D., Joshi, S., Anderson, J. A., Mazwi, M., Goldenberg, A., & Shaul, R. Z. (2020). Patient safety and quality improvement: Ethical principles for a regulatory approach to bias in healthcare machine learning. *Journal of the American Medical Informatics Association*, *27*(12), 2024–2027. https://doi.org/10.1093/jamia/ocaa085

McDougall, R. (2018). Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics*, *45*(3), 156–160. https://doi.org/10.1136/medethics-2018-105118

McGraw, D. (2012). Building public trust in uses of Health Insurance Portability and Accountability Act de-identified data. *Journal of the American Medical Informatics Association*, *20*(1), 29–34. https://doi.org/10.1136/amiajnl-2012-000936

McGregor, B., Belton, A., Henry, T. L., Wrenn, G., & Holden, K. B. (2019). Improving behavioral health equity through cultural competence training of health care providers. *Ethnicity & Disease*, *29*(Supp2), 359–364. https://doi.org/10.18865/ed.29.s2.359

McLellan, E., MacQueen, K. M., & Neidig, J. L. (2003). Beyond the qualitative interview: Data preparation and transcription. *Field Methods*, *15*(1), 63–84. https://doi.org/10.1177/1525822x02239573

Meinke, A., Schoen, B., Scheurer, J., Balesni, M., Shah, R., & Hobbhahn, M. (2024). Frontier Models are Capable of In-context Scheming. *arXiv.org*. https://arxiv.org/abs/2412.04984

Melcher, J. R., & Torous, J. (2020). Smartphone apps for college mental health: A concern for privacy and quality of current offerings. *Psychiatric Services*, *71*(11), 1114–1119. https://doi.org/10.1176/appi.ps.202000098

Mello, M. M., & Rose, S. (2024). Denial—Artificial intelligence tools and health insurance coverage decisions. *JAMA Health Forum*, *5*(3), e240622. https://doi.org/10.1001/jamahealthforum.2024.0622

Melnyk, B. M., Hoying, J., & Tan, A. (2020). Effects of the MINDSTRONG© CBT-based program on depression, anxiety and healthy lifestyle behaviors in graduate health sciences students. *Journal of American College Health*, *70*(4), 1001–1009. https://doi.org/10.1080/07448481.2020.1782922

Memishi, B., Appuswamy, R., & Paradies, M. (2019). *Cold storage data archives: more than just a bunch of tapes*. Cornell University. https://doi.org/10.48550/arxiv.1904.04736

Mendes-Santos, C., Nunes, F., Weiderpass, E., Santana, R., & Andersson, G. (2022). Understanding mental health professionals' perspectives and practices regarding the implementation of digital mental health: Qualitative study. *JMIR Formative Research*, *6*(4), Article e32558. https://doi.org/10.2196/32558

Mercer, T., & Khurshid, A. (2021). Advancing health equity for people experiencing homelessness using blockchain technology for identity management: A research agenda. *Journal of Health Care for the Poor and Underserved*, *32*(2S), 262–277. https://doi.org/10.1353/hpu.2021.0062

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). John Wiley & Sons.

Misra, S., & Mondal, A. (2011). Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment. *Mathematical and Computer Modelling*, *53*(3–4), 504–521. https://doi.org/10.1016/j.mcm.2010.03.037

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Moczuk, E., & Płoszajczak, B. (2020). Artificial intelligence – Benefits and threats for society. *Humanities and Social Sciences Quarterly*. https://doi.org/10.7862/rz.2020.hss.22

Mody, V., & Mody, V. (2019). *Mental health monitoring system using artificial intelligence: A review*. IEEE. https://doi.org/10.1109/i2ct45611.2019.9033652

Moggia, D., Lutz, W., Brakemeier, E., & Bickman, L. (2024). Treatment personalization and precision mental health care: Where are we and where do we want to go? *Administration and Policy in Mental Health and Mental Health Services Research*, *51*(5), 611–616. https://doi.org/10.1007/s10488-024-01407-w

Molnár-Gábor, F. (2019). Artificial intelligence in healthcare: Doctors, patients, and liabilities. In T. Wischmeyer & T. Rademacher (Ed.), *Regulating artificial intelligence* (pp. 337–360). https://doi.org/10.1007/978-3-030-32361-5_15

Moradi, P., & Levy, K. (2020). The future of work in the age of AI: Displacement or risk-shifting? In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *Oxford handbook of ethics of AI* (pp. 271–287). Oxford University Press.

Moreau, J. T., Baillet, S., & Dudley, R. W. (2020). Biased intelligence: on the subjectivity of digital objectivity. *BMJ Health & Care Informatics*, *27*(3), Article e100146. https://doi.org/10.1136/bmjhci-2020-100146

Morris, C. (2023, May 26). National Eating Disorder Association replaces human helpline staff with an AI chatbot. *Fortune Well*. https://fortune.com/well/2023/05/26/national-eating-disorder-association-ai-chatbot-tessa/

Na, S., Heo, S., Han, S., Shin, Y., & Roh, Y. (2022). Acceptance model of artificial intelligence (AI)-based technologies in construction firms: Applying the technology acceptance model (TAM) in combination with the technology–organisation–environment (TOE) framework. *Buildings*, *12*(2), Article 90. https://doi.org/10.3390/buildings12020090

Nakao, M., Shirotsuki, K., & Sugaya, N. (2021). Cognitive–behavioral therapy for management of mental health and stress-related disorders: Recent advances in techniques and technologies. *BioPsychoSocial Medicine*, *15*(1), 16. https://doi.org/10.1186/s13030-021-00219-w

Napolitano, E. (2023, November 21). *UnitedHealth uses faulty AI to deny elderly patients medically necessary coverage, lawsuit claims*. CBS News.

https://www.cbsnews.com/news/unitedhealth-lawsuit-ai-deny-claims-medicare-advantage-health-insurance-denials/

National Eating Disorders Association. (2023, September 22). *Our work*. https://www.nationaleatingdisorders.org/about-us/our-work

Neumann, O., Guirguis, K., & Steiner, R. (2022). Exploring artificial intelligence adoption in public organizations: A comparative case study. *Public Management Review*, *26*(1), 1–28. https://doi.org/10.1080/14719037.2022.2048685

Newhouse, J. P., & Sinaiko, A. D. (2007). Productivity adjustment in the Medicare physician fee schedule update. *Health Care Financing Review*, *29*(2), 5–14. https://pmc.ncbi.nlm.nih.gov/articles/PMC4195023/

Nguyen, N., Labonté-LeMoyne, É., Grégoire, Y., Radanielina-Hita, M. L., & Sénécal, S. (2022, June 26–July 1). *Understanding the* ’*patients’ adoption and usage of AI solution in mental health: A scoping review* [Paper presentation]. 24th International Conference on Human-Computer Interaction, online. https://doi.org/10.1007/978-3-031-19682-9_85

Niebel, C. (2021). The impact of the general data protection regulation on innovation and the global political economy. *Computer Law & Security Report*, *40*, Article 105523. https://doi.org/10.1016/j.clsr.2020.105523

Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819346

Noguchi, Y. (2023, January 19). Therapy by chatbot? The promise and challenges in using AI for mental health. *NPR*. https://www.npr.org/sections/health-shots/2023/01/19/1147081115/therapy-by-chatbot-the-promise-and-challenges-in-using-ai-for-mental-health

Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy*, *55*(4), 303–315. https://doi.org/10.1037/pst0000193

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

O'Connor, J., & Matthews, G. (2011). Informational privacy, public health, and state laws. *American Journal of Public Health*, *101*(10), 1845–1850. https://doi.org/10.2105/ajph.2011.300206

Ofulue, J., & Benyoucef, M. (2022). Data monetization: insights from a technology-enabled literature review and research agenda. *Management Review Quarterly*, *74*(2), 521–565. https://doi.org/10.1007/s11301-022-00309-1

Oguamanam, C. (2020). Indigenous peoples, data sovereignty and self-determination: Current realities and imperatives. *The African Journal of Information and Communication*, *2020*(26), 31–50. https://doi.org/10.23962/10539/30360

Oliveira, T., & Martins, M. F. O. (2010). Understanding e-business adoption across industries in European countries. *Industrial Management and Data Systems*, *110*(9), 1337–1354. https://doi.org/10.1108/02635571011087428

Oliveira, T., Thomas, M. A., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management*, *51*(5), 497–510. https://doi.org/10.1016/j.im.2014.03.006

Ooijen, P., Darzidehkalani, E., & Dekker, A. (2022). *AI technical considerations: Data storage, cloud usage and AI pipeline*. Cornell University. https://doi.org/10.48550/arxiv.2201.08356

Pandey, M., Kamrul, R., Michaels, C. R., & McCarron, M. (2021). Identifying barriers to healthcare access for new immigrants: a qualitative study in Regina, Saskatchewan, Canada. *Journal of Immigrant and Minority Health*, *24*(1), 188–198. https://doi.org/10.1007/s10903-021-01262-z

Paniagua, F. A. (2018). ICD-10 versus DSM-5 on cultural issues. *SAGE Open*, *8*(1), 215824401875616. https://doi.org/10.1177/2158244018756165

Park, H. J. (2024). Patient perspectives on informed consent for medical AI: A web-based experiment. *Digital Health*, *2024*(10). https://doi.org/10.1177/20552076241247938

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, *5*(5), Article 100988. https://doi.org/10.1016/j.patter.2024.100988

Parsons, R. D., & Dickinson, K. L. (2016). *Ethical practice in the human services: From Knowing to Being*. SAGE Publications.

Pasricha, S. (2022). AI ethics in smart Healthcare. *IEEE Consumer Electronics Magazine*, *12*(4), 12–20. https://doi.org/10.1109/mce.2022.3220001

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *PubMed*, *34*(5 Pt 2), 1189–1208. https://pubmed.ncbi.nlm.nih.gov/10591279

Patton-López, M. (2022). Communities in action: Pathways to health equity. *Journal of Nutrition Education and Behavior*, *54*(1), 94–95. https://doi.org/10.1016/j.jneb.2021.09.012

Payne, K. (2024, October 25). AI chatbot pushed teen to kill himself, lawsuit alleges. *AP News*. https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdbc3c90d1456bd0

Percy, S. (2024, August 8). Do the right thing: 4 ways to navigate the ethics of AI. *Forbes*. https://www.forbes.com/sites/sallypercy/2024/08/05/do-the-right-thing-4-ways-to-navigate-the-ethics-of-ai/

Perlis, R. (2023, February 6). Mindstrong's demise and the future of mental health care. *STAT*. https://www.statnews.com/2023/02/06/mindstrong-demise-future-mental-health-care/

Petersson, L., Larsson, I., Nygren, J. M., Nilsen, P., Neher, M., Reed, J., Tyskbo, D., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research, 22*(1), Article 850. https://doi.org/10.1186/s12913-022-08215-8

Place, S., Blanch-Hartigan, D., Smith, V., Erb, J. L., Marci, C. D., & Ahern, D. K. (2020). Effect of a mobile monitoring system vs usual care on depression symptoms and psychological health. *JAMA Network Open, 3*(1), Article e1919403. https://doi.org/10.1001/jamanetworkopen.2019.19403

Polzer, A., Fleiß, J., Ebner, T., Kainz, P., Koeth, C., & Thalmann, S. (2022). Validation of AI-based information systems for sensitive use cases: using an XAI approach in pharmaceutical engineering. *Proceedings of the . . . Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2022.186

Poufinas, T., Gogas, P., Papadimitriou, T., & Zaganidis, E. (2023). Machine learning in forecasting motor insurance claims. *Risks, 11*(9), Article 164. https://doi.org/10.3390/risks11090164

Powell, J., & Kleiner, A. (2023). *The AI dilemma: 7 principles for responsible technology*. Berrett-Koehler Publishers.

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*, *7*(4), 351–367. https://doi.org/10.1007/s12553-017-0179-1

Pratap, A., Homiar, A., Waninger, L., Herd, C., Suver, C., Volponi, J., Anguera, J. A., & Areán, P. (2022). Real-world behavioral dataset from two fully remote smartphone-based randomized clinical trials for depression. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01633-7

Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M. S., Baiocchi, M., Pajarito, S., & Robinson, A. (2021). A therapeutic relational agent for reducing problematic substance use (WOEBOT): Development and usability study. *Journal of Medical Internet Research*, *23*(3), Article e24850. https://doi.org/10.2196/24850

Psychiatrist.com. (2023, June 5). *NEDA suspends AI chatbot for giving harmful eating disorder advice*. https://www.psychiatrist.com/news/neda-suspends-ai-chatbot-for-giving-harmful-eating-disorder-advice/

Pujari, S., Reis, A., Zhao, Y., Alsalamah, S., Serhan, F., Reeder, J. C., & Labrique, A. (2023). Artificial intelligence for global health: cautious optimism with safeguards. *Bulletin of the World Health Organization*, *101*(06), 364–364A. https://doi.org/10.2471/blt.23.290215

Rangavittal, P. B. (2022). Evolving role of AI in enhancing patient care within digital health platforms. *Journal of Artificial Intelligence & Cloud Computing*, 1–6. https://doi.org/10.47363/jaicc/2022(1)241

Räsänen, M., & Nyce, J. M. (2013). The raw is cooked. *Science Technology & Human Values*, *38*(5), 655–677. https://doi.org/10.1177/0162243913480049

Razzouk, D. (2023). Editorial: Mental health economics and public mental health policy: mental health services costs, quality and its impact on reducing the burden of mental illness. *Frontiers in Health Services*, *3,* Article 1267580. https://doi.org/10.3389/frhs.2023.1267580

Regueiro, C., Seco, I., Gutiérrez-Agüero, I., Urquizu, B., & Mansell, J. (2021). A blockchain-based audit trail mechanism: Design and implementation. *Algorithms*, *14*(12), Article 341. https://doi.org/10.3390/a14120341

Rezaeikhonakdar, D. (2023). AI chatbots and challenges of HIPAA compliance for AI developers and vendors. *The Journal of Law Medicine & Ethics*, *51*(4), 988–995. https://doi.org/10.1017/jme.2024.15

Rizzo, A., Hartholt, A., & Mozgai, S. (2021). From combat to COVID-19 – Managing the impact of trauma using virtual reality. *Journal of Technology in Human Services*, *39*(3), 314–347. https://doi.org/10.1080/15228835.2021.1915931

Rizzo, A., Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., Williams, J., Difede, J., Rothbaum, B. O., Reger, G., Parsons, T., & Kenny, P. (2011). SimCoach: an intelligent virtual human system for providing healthcare information and support. *International Journal on Disability and Human Development*, *10*(4), 277–281. https://doi.org/10.1515/IJDHD.2011.046

Rizzo, A., & Shilling, R. (2017). Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. *European Journal of Psychotraumatology*, *8*(sup5). https://doi.org/10.1080/20008198.2017.1414560

Rizzo, A. S., & Koenig, S. T. (2017). Is clinical virtual reality ready for primetime? *Neuropsychology*, *31*(8), 877–899. https://doi.org/10.1037/neu0000405

Robin, J., Xu, M., Kaufman, L. D., & Simpson, W. S. (2021). Using digital speech assessments to detect early signs of cognitive impairment. *Frontiers in Digital Health*, *3*, Article 749758. https://doi.org/10.3389/fdgth.2021.749758

Rocher, L., Hendrickx, J. M., & De Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10933-3

Rosenfeld, S., Bernasek, C., & Mendelson, D. (2005). Medicare's next voyage: encouraging physicians to adopt health information technology. *Health Affairs*, *24*(5), 1138–1146. https://doi.org/10.1377/hlthaff.24.5.1138

Russell, C. (2023, February 8). Health care bias is dangerous. but so are 'fairness' algorithms. *WIRED*. https://www.wired.com/story/bias-statistics-artificial-intelligence-healthcare/

Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson Education Limited.

Russell, S. J., & Norvig, P. (1995). Artificial intelligence: a modern approach. *Choice Reviews Online*, *33*(03), 33–1577. https://doi.org/10.5860/choice.33-1577

Sandu, I., Wiersma, M., & Manichand, D. (2022). Time to audit your AI algorithms. *Maandblad Voor Accountancy En Bedrijfseconomie*, *96*(7/8), 253–265. https://doi.org/10.5117/mab.96.90108

Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science*, *3*(2). https://doi.org/10.1007/s42979-022-01043-x

Scheuch, K. (2024, December 2). *News Release: Utah Department of Commerce's Office of Artificial Intelligence announces first regulatory mitigation agreement* [Press release].

https://blog.commerce.utah.gov/2024/12/02/news-release-utah-department-of-commerces-office-of-artificial-intelligence-announces-first-regulatory-mitigation-agreement/

Schueller, S. M., Hunter, J. F., Figueroa, C., & Aguilera, A. (2019). Use of digital mental health for marginalized and underserved populations. *Current Treatment Options in Psychiatry*, *6*(3), 243–255. https://doi.org/10.1007/s40501-019-00181-z

Seger, E. (2022). In defence of principlism in AI ethics and governance. *Philosophy & Technology*, *35*(2). https://doi.org/10.1007/s13347-022-00538-y

Shahriari, K., & Shahriari, M. (2017). IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. *IEEE*. https://doi.org/10.1109/ihtc.2017.8058187

Sharfstein, S. S. (2019). Community mental health and deinstitutionalization. *Psychiatric News*, *54*(24). https://doi.org/10.1176/appi.pn.2019.12b28

Sharma, R. (2020). Artificial Intelligence in Healthcare: a review. *Türk Bilgisayar Ve Matematik Eğitimi Dergisi*, *11*(1), 1663–1667. https://doi.org/10.61841/turcomat.v11i1.14628

Shi, L., & Singh, D. A. (2004). *Essentials of the U.S. health care system*. https://openlibrary.org/books/OL22552128M/Essentials_of_the_U.S._health_care_system

Shilton, K., Moss, E., Gilbert, S. A., Bietz, M. J., Fiesler, C., Metcalf, J., Vitak, J., & Zimmer, M. (2021). Excavating awareness and power in data science: A manifesto for trustworthy pervasive data research. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211040759

Shu, J. (2024). *Data Storage Architectures and Technologies*. Springer.

Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Informaion Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/fitee.1700826

Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine*, *296*, 114782. https://doi.org/10.1016/j.socscimed.2022.114782

Silverman, J. J., Galanter, M., Jackson-Triche, M., Jacobs, D. G., Lomax, J. W., Riba, M., Tong, L., Watkins, K. E., Fochtmann, L. J., Rhoads, R. S., & Yager, J. (2015). The American Psychiatric Association practice guidelines for the psychiatric evaluation of adults. *American Journal of Psychiatry*, *172*(8), 798–802. https://doi.org/10.1176/appi.ajp.2015.1720501

Singh, R., & Gill, S. S. (2023). Edge AI: A survey. *Internet of Things and Cyber-Physical Systems*, *3*, 71–92. https://doi.org/10.1016/j.iotcps.2023.02.004

Singh, S., & Thakur, H. K. (2020). Survey of Various AI Chatbots Based on Technology Used. *IEEE*. https://doi.org/10.1109/icrito48877.2020.9197943

Sinha, C., Meheli, S., & Kadaba, M. (2023). Understanding digital mental health needs and usage with an artificial intelligence–led mental health app (WYSA) during the COVID-19 pandemic: Retrospective analysis. *JMIR Formative Research*, *7*, e41913. https://doi.org/10.2196/41913

Sipola, T., Alatalo, J., Kokkonen, T., & Rantonen, M. (2022). Artificial intelligence in the IoT Era: A review of edge AI hardware and software. *FRUCT Oy*, 320–331. https://doi.org/10.23919/fruct54823.2022.9770931

Šlepecký, M., Škobrtal, P., Novotný, M., Bazinková, E., & Kotianová, A. (2018). Exposure therapy by virtual reality and its monitoring by biofeedback. *Cognitive Remediation Journal*, *7*(1), 4–9. https://doi.org/10.5507/crj.2018.001

Sofia, S., Montserrat, L. C., Emilia, G. G., Giuditta, D. P., Martínez-Plumed, F., & Delipetrev, B. (2020). AI WATCH. Defining Artificial Intelligence. *Publications Office of the European Union*. https://doi.org/10.2760/382730

State of Colorado. (2024). *Consumer protections for artificial intelligence*. Colorado General Assembly. Retrieved December 11, 2024, from https://leg.colorado.gov/bills/sb24-205

Stix, C. (2018). *3 ways AI could help our mental health*. World Economic Forum. http://www.weforum.org/agenda/2018/03/3-ways-ai-could-could-be-used-in-mental-health/

Stoeklé, H., Mamzer-Bruneel, M., Vogt, G., & Hervé, C. (2016). 23andMe: A new two-sided data-banking market model. *BMC Medical Ethics*, *17*(1). https://doi.org/10.1186/s12910-016-0101-9

Straw, I., & Callison-Burch, C. (2020). Artificial Intelligence in mental health and the biases of language based models. *PLoS ONE*, *15*(12), e0240376. https://doi.org/10.1371/journal.pone.0240376

Stringer, H. (2025, January). *Technology is reshaping practice to expand psychology's reach*. https://www.apa.org. https://www.apa.org/monitor/2025/01/trends-technology-shaping-practice

Sutton, J., & Austin, Z. (2015). Qualitative research: data collection, analysis, and management. *The Canadian Journal of Hospital Pharmacy*, *68*(3). https://doi.org/10.4212/cjhp.v68i3.1456

Suver, C., Harper, J., Loomba, J., Saltz, M., Solway, J., Anzalone, A. J., Walters, K., Pfaff, E., Walden, A., McMurry, J., Chute, C. G., & Haendel, M. (2023). The N3C governance ecosystem: A model socio-technical partnership for the future of collaborative analytics at scale. *Journal of Clinical and Translational Science*, *7*(1). https://doi.org/10.1017/cts.2023.681

Terranova, C., Cestonaro, C., Fava, L., & Cinquetti, A. (2024). AI and professional liability assessment in healthcare. A revolution in legal medicine? *Frontiers in Medicine*, *10*. https://doi.org/10.3389/fmed.2023.1337335

Thomson, S., Sagan, A., & Mossialos, E. (2020). *Private health insurance: History, politics and performance*. Cambridge University Press.

Timmons, A. C., Duong, J. B., Fiallo, N. S., Lee, T., Vo, H. P. Q., Ahle, M. W., Comer, J. S., Brewer, L. C., Frazier, S. L., & Chaspari, T. (2022). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916221134490

Timofeyev, Y., & Jakovljević, M. (2020). Fraudster's and victims' profiles and loss predictors' hierarchy in the mental healthcare industry in the US. *Journal of Medical Economics*, *23*(10), 1111–1122. https://doi.org/10.1080/13696998.2020.1801454

Toosi, A., Bottino, A., Saboury, B., Siegel, E. L., & Rahmim, A. (2021). A brief history of AI: How to prevent another winter (A critical review). *PET Clinics*, *16*(4), 449–469. https://doi.org/10.1016/j.cpet.2021.07.001

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. (1990). *Processes of technological innovation*. Lexington Books.

Tornero-Costa, R., Martinez-Millana, A., Azzopardi-Muscat, N., Lazeri, L., Traver, V., & Novillo-Ortiz, D. (2023). Methodological and quality flaws in the use of artificial intelligence in mental health research: Systematic review. *JMIR Mental Health*, *10*, Article e42045. https://doi.org/10.2196/42045

Trang, B. (2024, December 21). As Congress waffles on AI, state legislatures step up to fill the void. *STAT*. https://www.statnews.com/2024/12/23/regulating-health-ai-state-laws-increase-compliance-burden/

Tseng, E., Meza, K., Marsteller, J. A., Clark, J. M., Maruthur, N. M., & Smith, K. (2022). Engaging payors and primary care physicians together in improving diabetes prevention. *Journal of General Internal Medicine*, *38*(2), 309–314. https://doi.org/10.1007/s11606-022-07788-8

Turner Lee, N., Resnick, P., & Barton, G. (2023, June 27). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. The Brookings Institution. https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Tutun, S., Johnson, M., Ahmed, A., Albizri, A., Irgil, S., Yesilkaya, I., Ucar, E. N., Sengun, T., & Harfouche, A. (2022). An AI-based decision support system for predicting mental health disorders. *Information Systems Frontiers*, *25*(3), 1261–1276. https://doi.org/10.1007/s10796-022-10282-5

U.S. Department of Health and Human Services. (2022, October 9). *Summary of the HIPAA Privacy Rule*. https://www.hhs.gov. Retrieved December 8, 2024, from https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

Utah State Legislature. (2024). *S.B. 149 Artificial Intelligence Amendments*. https://le.utah.gov/~2024/bills/static/SB0149.html

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, *15*(11), Article e1002689. https://doi.org/10.1371/journal.pmed.1002689

Virginia's Legislative Information System. (2021, March 18). *2021 Special Session I*. https://legacylis.virginia.gov/cgi-bin/legp604.exe?212+ful+CHAP0219

Wang, J. (2023). The power of AI-Assisted diagnosis. *ICST Transactions on e-Education and e-Learning*, *8*(4), e3. https://doi.org/10.4108/eetel.3772

Wang, T., & Bashir, M. (2020). Privacy considerations when predicting mental health using social media. *Proceedings of the Association for Information Science and Technology*, *57*(1). https://doi.org/10.1002/pra2.244

Warraich, H. J., Tazbaz, T., & Califf, R. M. (2024, October 15). FDA perspective on the regulation of artificial intelligence in health care and biomedicine. *JAMA*. https://doi.org/10.1001/jama.2024.21451

Weber, M., Engert, M., Schaffer, N., Weking, J., & Krcmar, H. (2022). Organizational capabilities for AI implementation—Coping with inscrutability and data dependency in AI. *Information Systems Frontiers*, *25*(4), 1549–1569. https://doi.org/10.1007/s10796-022-10297-y

Wells, K. (2023, June 9). An eating disorders chatbot offered dieting advice, raising fears about AI in health. *NPR*. https://www.npr.org/sections/health-shots/2023/06/08/1180838096/an-eating-disorders-chatbot-offered-dieting-advice-raising-fears-about-ai-in-hea

The White House. (2023). *Blueprint for an AI Bill of Rights*. Retrieved March 16, 2023, from https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Wieder, P., & Nolte, H. (2022). Toward data lakes as central building blocks for data management and analysis. *Frontiers in Big Data*, *5*, Article 945720. https://doi.org/10.3389/fdata.2022.945720

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18

Williams, R. (2023, May 24). Innovative app uses AI to battle rising veteran suicide rates. *Spectrum News 1*. https://spectrumnews1.com/ca/la-west/mental-health/2023/05/24/innovative-app-uses-ai-to-battle-rising-veteran-suicide-rates#

Wimbarti, S., Kairupan, B. H. R., & Tallei, T. E. (2024). Critical review of self-diagnosis of mental health conditions using artificial intelligence. *International Journal of Mental Health Nursing*, *33*(2), 344–358. https://doi.org/10.1111/inm.13303

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. https://www.who.int/publications/i/item/9789240029200

Zabelski, S., Hollander, M., & Alexander, A. (2024). Addressing inequities in access to mental healthcare: A policy analysis of community mental health systems serving minoritized populations in North Carolina. *Administration and Policy in Mental Health and Mental Health Services Research*, *51*(4), 543–553. https://doi.org/10.1007/s10488-024-01344-8

Zanna, K., Sridhar, K., Yu, H., & Sano, A. (2022). Bias reducing multitask learning on mental health prediction. *IEEE*. https://doi.org/10.1109/acii55700.2022.9953850

Zeberga, K., Attique, M., Shah, B., Ali, F., Jembre, Y. Z., & Chung, T. S. (2022). A novel text mining approach for mental health prediction using Bi-LSTM and BERT model. *Computational Intelligence and Neuroscience*, *2022*, 1–18. https://doi.org/10.1155/2022/7893775

Zhang, M., Scandiffio, J., Younus, S., Jeyakumar, T., Karsan, I., Charow, R., Salhia, M., & Wiljer, D. (2023). The Adoption of AI in Mental Health Care–Perspectives from Mental Health Professionals: Qualitative Descriptive study. *JMIR Formative Research*, *7*, e47847. https://doi.org/10.2196/47847

Zhao, J., Chang, Y., Li, D., Xia, C., Cui, H., Zhang, K., & Feng, X. (2018). On retargeting the AI programming framework to new hardwares. In *Lecture notes in computer science* (pp. 39–51). https://doi.org/10.1007/978-3-030-05677-3_4

Zidaru, T., Morrow, E., & Stockley, R. (2021). Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. *Health Expectations*, *24*(4), 1072–1124. https://doi.org/10.1111/hex.13299

Zine, M., Harrou, F., Terbeche, M., Bellahcene, M., Dairi, A., & Sun, Y. (2023). E-Learning readiness assessment using machine learning methods. *Sustainability*, *15*(11), 8924. https://doi.org/10.3390/su15118924

Zona, M. A., & Thaib, I. (2020, November 14–15). *Leadership and commitment to organizational change of government employee in Padang, West Sumatra* [Paper presentation]. Sixth Padang International Conference on Economics Education, Economics, Business and Management, Accounting and Entrepreneurship, Padang, Indonesia. https://doi.org/10.2991/aebmr.k.210616.074

Zou, C. (2024). Revolutionizing machine learning: Harnessing hardware accelerators for enhanced AI efficiency. *Applied and Computational Engineering*, *47*(1), 141–146. https://doi.org/10.54254/2755-2721/47/20241256

## Appendix A: Interview Protocol

This section outlines the research questions, conceptual foundations, questions, respondent criteria, an introduction, structured interview questions and probes, and an introduction and conclusion for the subjects. The interview questions align with the research aim to elicit insights into perceptions, organizational elements, and external factors of AI adoption in mental healthcare settings.

### Brief Statement of the Problem

While artificial intelligence promises advancements in mental health assessments, interventions, and accessibility, adopting AI risks displacing human relationships and creating harm if implemented without proper precautions (Fiske et al., 2019; Sinha et al., 2023). This research explores organizational barriers and facilitators to ethically integrating AI in a way that complements clinicians' irreplaceable therapeutic role. Rather than focusing on technical details, it investigates MHLs' perceptions, organizational factors, and environmental pressures to identify considerations for responsibly adopting AI to augment human-centered practice. This exploratory research aims to help MHLs, ethicists, and technologists strategize how to leverage AI's potential in a way that preserves the humanistic core of mental healthcare.

### Brief Statement of the Conceptual Framework

This study applied two complementary frameworks—the TOE framework—to develop a comprehensive understanding of the factors influencing the ethical integration of AI in mental healthcare settings. The TOE framework (Tornatzky et al., 1990) provides a lens to analyze the technological, organizational, and environmental dimensions that shape AI adoption, including organizational culture, resources, regulations, and ethical landscape.

## Research Questions

1. What are the MHLs' perceptions of adopting AI technologies in their organizations?
2. What organizational factors influence MHLs to adopt AI technologies responsibly in their organization?
3. What external factors are MHLs concerned about the most when evaluating their organization's readiness to adopt AI technologies?

## Respondent Type

This research utilized purposive sampling to recruit 11 U.S.-based leaders at various levels of responsibility from community mental health and group practice organizations to participate in interviews. Eligible leadership roles included frontline supervisors overseeing employees, middle managers supervising frontline supervisors, senior executives supervising middle managers, and thought leaders in advisory/consultative roles providing expertise on technology, ethics, and mental healthcare policy. Participants were MHLs with a background in community mental health or group practice settings and at least 1 year of experience in a leadership position related to clinical operations and clinical technology, which may intersect with a clinical license, though licensure is not mandatory for inclusion.

The goal is to target leaders with extensive mental health experience who are currently in organizational decision-making roles related to administration and care delivery workflows. Exclusions focus the sample on those in community-based organizations and group practice settings rather than academics or solo private practices. Exclusion criteria are as follows:

- mental health professionals whose leadership experience is exclusive to self-owned, single-provider practices.
- entry-level practitioners without leadership or management duties

**Introduction to the Interview**

Thank you for taking the time to participate in this interview. I am Mabel Yiu, and I am a leadership and organizational change doctoral student from USC. I am conducting a study exploring perspectives on adopting artificial intelligence technologies in mental healthcare organizations. You have been selected for this interview because of your valuable insights as a leader overseeing technology decisions that shape clinical practice.

The aim of this research was to uncover factors that influence AI adoption to develop best practices for its ethical and effective implementation. My goal is to gather mental health leaders' views, motivations, and concerns to guide appropriate integration of AI that thoughtfully augments human expertise and therapeutic relationships.

This semi-structured interview will take around 60 minutes and will be recorded for analysis if you consent. All responses will remain confidential, and your identity will be protected. There are no right or wrong answers; I am simply interested in understanding your honest perspectives. I would like to emphasize that you are in control of this interview. If there are any questions you do not wish to answer or topics you do not wish to discuss, please let me know. The interview may also be terminated at any time. Our conversation will be kept confidential; no identifiable information is recorded and any information you provide will be used only for this research.

If you have any other questions, please let me know before we begin. May I have your permission to record our discussion? Thank you, I truly appreciate your participation today.

**Table A1**

*Interview Protocol*

| Interview questions | Potential probes | RQ addressed | Key concept addressed |
|---|---|---|---|
| What is your current understanding of how AI is being applied in mental healthcare settings? | Can you give me an example? | RQ1 | TOE, technology |
| How would you describe your general perceptions and attitudes toward adopting AI technologies in your organization? | What factors shape your overall opinion of AI tools? | RQ1 | TOE, technology |
| In what ways could AI technologies benefit your organizations? | What opportunities excite you the most about AI tools?<br>Can you give me an example? | RQ1 | TOE, technology |
| What potential risks do you associate with AI adoption? | Can you give me an example? | RQ1 | TOE, technology |
| What are some of the main workflow or process needs you feel AI could address in your organization? | Where are current pain points that technology could help? | RQ2 | TOE, organizational |
| In your view, what barriers currently exist that may inhibit effective AI adoption? | What infrastructure gaps would need to be addressed? | RQ2 | TOE, organizational |
| How could leadership better support adoption of new technologies like AI? | Can you give me an example? | RQ2 | TOE, organizational |
| How might introducing AI tools impact the culture of clinical teams and their openness to integrating new technologies into practice? | What cultural shifts or training would be needed? | RQ2 | TOE, organizational |
| Is there any reluctance or resistance within your organization to adopt AI? | Can you give me an example? | RQ2 | TOE, organizational |
| How would your organization respond if your employee secretly used AI technology such as using ChatGPT to complete clinical notes? | Are there existing policies in place to detect and address unauthorized use of AI technology? | RQ2 | TOE, organizational |
| Could you share some insight into who takes the lead on | What are the roles and backgrounds of those | RQ2 | TOE, organizational |

| Interview questions | Potential probes | RQ addressed | Key concept addressed |
|---|---|---|---|
| establishing guidelines for the ethical use of AI technologies within your organization? | helping craft responsible AI policies? | | |
| What best practices in change management would you emphasize to successfully integrate AI innovations into your mental health services? | What training or support would leadership need to provide? | RQ2 | TOE, organizational |
| What external pressures or competitive forces may influence the need to adopt AI technologies in your organization? | Which environmental factors most significantly impact technology decisions? | RQ3 | TOE, environmental |
| How does patient demand shape your approach to AI tools? | Can you give me an example? | RQ3 | TOE, environmental |
| How significant are regulatory requirements and ethical obligations in evaluating whether to adopt an AI technology? | Which regulations or ethical codes are most relevant? | RQ3 | TOE, environmental |
| How could AI tools impact clinical effectiveness compared to current human-delivered practices? | What tasks or roles do you feel should remain strictly human-delivered? | RQ3 | TOE, environmental |
| What challenges, risks or unintended consequences are you most concerned about in adopting AI technologies? | How might AI negatively impact workplace culture and human roles? | RQ3 | TOE, environmental |
| Have you or your organization seen vendors marketing their AI tools for mental health? | Can you give me some examples? | RQ3 | TOE, environmental |
| On an industry-wide level, what groups or organizations do you trust to lead the guidelines development for implementing AI in mental healthcare? | Who are the major players influencing those conversations? | RQ3 | TOE, environmental |
| Is there anything I didn’t ask that you think is important? | What is your thought on AI in the future of mental health? | Closing | Closing |

## Conclusion to the Interview

Thank you so much for taking the time to speak with me today and share your valuable perspectives on AI adoption in mental healthcare organizations. I truly appreciate you providing such thoughtful insights into the opportunities and challenges associated with integrating emerging technologies into clinical practice and workflows. Your opinions and experiences will be incredibly helpful for developing best practices to guide leaders in ethically and effectively leveraging AI tools to augment human expertise. If any other reflections or considerations come to mind after our discussion, please feel free to contact me. I sincerely thank you for contributing your first-hand knowledge to this research and participating in this interview. It was a pleasure speaking with you, and I wish you the very best.

# Appendix B: Institutional Review Board Approval

| | |
|---|---|
| Date: | Dec 20, 2023, 02:57pm |
| Action Taken: | **Approve** |
| Principal Investigator: | Mabel Yiu<br>ROSSIER SCHOOL OF EDUCATION |
| Faculty Advisor: | Patricia Tobey<br>OFFICE OF THE PROVOST |
| Co-Investigator(s): | |
| Project Title: | **Responsible AI adoption in Mental Health Organization** |
| Study ID: | **UP-23-01193** |
| Funding: | |

Your submission was reviewed and determined to be exempt §46.104(d)(2) on 12/20/2023 by the University of Southern California Institutional Review Board (IRB).

Any Unanticipated Problems Involving Risks to Subjects or Others, Deviations from the approved submission, or complaints, must be reported to the IRB in accordance with University of Southern California Human Research Protection Program policies and procedures.

If there are significant changes that increase the risk to subjects, revisions to study materials/documents, or if the funding has changed, you must submit an amendment to the IRB for review. For other revisions to the application, use the "Send Message to IRB" link.

iStar Application and contents dated 12/08/2023 were reviewed.
Attachments
:

Social-behavioral health-related interventions or health-outcome studies must register with **clinicaltrials.gov** or other International Community of Medical Journal Editors (ICMJE) approved registries in order to be published in an ICJME journal. The ICMJE will not accept studies for publication unless the studies are registered prior to enrollment, despite the fact that these studies are not applicable "clinical trials" as defined by the Food and Drug Administration (FDA). For support with registration, go to www.clinicaltrials.gov or contact Jean Chan (jeanbcha@usc.edu, 323-442-2825).