Measuring Perceived Empathy in Generative AI in the Context of Mental Health Applications

Debra Jordan -T00463444

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Nursing from Thompson Rivers University (TRU) Kamloops BC

Primary Supervisor: Dr. Florriann Fehr, PhD, RN Associate Professor, School of Nursing, TRU

Committee Members Dr. Anila Virani, PhD, RN Assistant Professor, School of Nursing, TRU

Committee Members: Dr. Amy Beck, PhD, RN Assistant Professor, School of Nursing, TRU

July 30, 2025

Abstract

The rising popularity of generative artificial intelligence and chatbot features in systems like ChatGpt for supporting mental health requires research into long-term risks and benefits. Mental health concerns are on the rise around the world. This heightened demand has exacerbated wait times, impacted access, and generally made it harder for those seeking services to be referred to and connected to mental health support. ChatGPT has the potential to become a simple, cost-effective, and easy-to-access solution to the increased demand. ChatGpt may eventually fill a gap in services for many suffering people. Empathy is the cornerstone of the therapeutic relationship, and understanding how users perceive ChatGPT's empathy is imperative to building safe and effective mental health support.

Eight participants engaged in an interactive mental health scenario with ChatGPT. Then they participated in a virtual semi-structured interview to share their reflections on how empathy was experienced during the exchange. To support and prompt further reflection, they also completed the Perceived Empathy of Technology Scale (PETS), a brief survey focused on their perceptions. Interview narratives formed the primary source of insight, with survey responses used to enrich and triangulate emerging themes. Inductive thematic analysis was used to identify key patterns in participants' perceptions of empathy, and basic descriptive statistics were used to triangulate and validate the integration of study findings. Results from the qualitative-driven study indicate that the perception of empathy is nuanced, but overall, participants felt that when their feelings were mirrored and validating language was used in the responses, they thought they were experiencing cognitive empathy from ChatGPT. These findings suggest that ChatGPT can be perceived as empathic by users seeking mental health support, with the subjective experience reflected in the post-activity interviews as well as from the survey.

Keywords: generative AI; mental health; empathy; chatbot; access; rapport; ChatGPT

Table of Contents

Measuring Perceived Empathy in Generative AI in the Context of Mental Health Applications

Chapter 1

Recent data highlights a significant rise in mood and anxiety disorders among Canadians, with many reporting unmet needs in mental health care, particularly counselling services (Stephenson, 2023). Factors like long wait times, workforce shortages, and high costs necessitate alternative solutions such as AI-driven mental health alternatives (Boucher et al., 2021). These alternatives offer accessible, cost-effective support, particularly valuable in rural areas where traditional services are scarce (Rad & Rad, 2023,Banerjee et al., 2024). A study from the Fraser Institute (Esmail, 2023) highlighted that the median waiting time between a referral from a general practitioner to receiving treatment was 27.4 weeks, the longest recorded in their survey's history. The wait time for urgent referral to a psychiatrist is up 30% from 2003-2005, and the wait time for treatment after a consult now averages 15.5 weeks, an increase of 49% from 2003-2005. Despite rising mental health needs, access to timely support remains a systemic challenge. In 2023, only 43% of Canadians with mental health concerns reported receiving any form of care (Esmail, 2023). These prolonged wait times and lack of access can have severe consequences, such as increased suffering and potential deterioration of mental health conditions (Kretzschmar et al., 2019). This underscores the urgent need for alternative solutions like AI-driven mental health apps, which may be able to provide more immediate support and help bridge the gap in access to care.

As a Registered Psychiatric Nurse working in the interior of British Columbia, I have seen firsthand the challenges people have in accessing timely mental health care. I have direct experience with the structural and logistical challenges influencing access to counselling

services. This professional insight provided valuable contextual understanding during the development of this research, particularly regarding the real-world barriers to timely mental health support. While national-level statistics suggest a median wait time of 28 days for community counselling (Canadian Institute for Health Information, 2024), firsthand experience of intake services within Interior Health reveals that wait times often extend significantly longer, more in line with the wait time for access to general psychiatric care as outlined in Emasil (2023). The public healthcare system in the province is stretched, and there are limited options for people who are struggling with their mental health unless they have the means to seek private care, and even then, the wait times for private counselling can vary. Within the publicly funded mental healthcare system, there are rigid referral and access pathways, a lack of programs, understaffing, and the lingering stench of stigma. The trauma and stigma associated with mental health can, at times, make accessing services extremely difficult. I have worked with adults who refused to seek emergency mental health care because of the treatment they had previously received in the Emergency Departments of the local hospitals, many of them stating they would rather die than be judged, dismissed or told they were being dramatic by hospital staff and left for hours in the waiting rooms because they were not deemed as critical in the triage process. The access to timely mental healthcare has reached a crisis point (Esmail, 2023), and maybe it is time to look at other options for mental health access and support. Perhaps it is time to let AI help? The use of digital mental health options, including mood tracking apps, text-based crisis lines, and guided self-help modules, is not novel and has been used to augment mental health services for many years (Inkster et al., 2018). Generative AI, however, offers unique, interactive, language-based support that mimics human conversation, potentially offering emotional presence and genuine rapport. ChatGPT, developed by OpenAI, is a widely used example of a large

language model-based generative AI that simulates conversation with human-like fluency. The effectiveness of digital interventions needs further validation against traditional methods through broader studies. As a mental healthcare provider, I am interested in seeing if these generative AI-driven interventions can provide empathic care to mental healthcare consumers. This study does not evaluate ChatGPT's diagnostic accuracy or therapeutic reliability. Instead, it explores how users perceive empathic presence during conversations with ChatGPT, a critical component of therapeutic rapport (Austin & Boyd, 2015). I have chosen to position this study within the framework of Jean Watson's Theory of Human Caring because the theory embodies the values that I, as a Registered Psychiatric Nurse, hold dear. A key concept in this theory is that empathy is a fundamental aspect of humanistic care that enhances the quality of nursing (Savieto &Leao, 2016). The theory promotes the establishment of a deep, trusting bond essential for empathetic care, emphasizing the importance of compassionate and ethical relationships that uphold patient dignity and foster healing through genuine rapport and empathy.

This lens is appropriate for examining the subjective experiences of participants as they engage with ChatGPT, a generative AI tool, to support them with a mental health challenge. Effective mental health support relies on the user's sense of connection (Austin & Boyd, 2015). Without a perceived sense of empathy, such interventions are unlikely to be effective.

Given the importance of perceived empathy in the effectiveness of mental health support (Austin & Boyd, 2015), it is concerning that there is a noticeable gap in Canadian research on the use of these technologies. Most existing studies have been conducted in the United States, Asia, and Europe (Xian et al., 2024), highlighting the need for localized research exploring how these tools simulate empathetic interactions and their potential impact within Canadian mental health care. This gap underscores the importance of exploring how ChatGpt is being used to support

mental health challenges within the Canadian context. To provide a foundation for this exploration, this chapter will define key terms, outline the problem and significance of using generative AI in mental health care, and review the purpose and objectives of the current research.

Definition

AI Chatbots: AI chatbots are automated systems designed to engage in conversation with users, typically providing mental health support by simulating human interactions (Boucher et al., 2021; Abd-alrazaq et al., 2019).

Carative Factors: Watson originally introduced ten carative factors as the core elements of her theory, later evolving them into clinical caritas processes. These factors guide nurses in practising caring behaviours. The key carative factors related to empathy and holistic care include:

- **Formation of a humanistic-altruistic** system of values: Encourages nurses to practice kindness and compassion.
- **Instillation of faith-hope**: Supports emotional and spiritual well-being, fostering trust between patient and nurse.
- **Cultivation of sensitivity to oneself and others:** Encourages the development of empathy, where the nurse remains aware of the patient's emotional and psychological state.
- **Development of a helping-trusting relationship**: Promotes the establishment of a deep bond through trust, critical for empathetic care.
- **Promotion of interpersonal teaching-learning**: Emphasizes communication and understanding, integral to empathetic care (Cara, 2003).

Caring Moment: A caring moment is a significant interaction between nurse and patient, allowing both individuals to experience personal growth. These moments are characterized by the nurse's authentic presence and empathetic engagement, allowing for a mutual transformation. The caring moment creates an environment where the patient feels safe and understood, promoting holistic healing (Cara, 2003).

ChatGPT: ChatGPT is an advanced artificial intelligence chatbot developed by OpenAI that is designed to engage in natural, human-like conversations. The name "ChatGPT" stands for "Chat Generative Pre-trained Transformer." This means it is built to generate meaningful and relevant text responses in conversations. The model was pre-trained on a vast amount of text from books, articles, and websites, allowing it to understand language patterns and context before being fine-tuned for more specific tasks. It operates using a Transformer architecture, a type of neural network that enables it to process and understand complex relationships between words, making its responses more accurate and contextually appropriate. This design allows ChatGPT to interact with users naturally and confidently and assist with various tasks, from answering questions to providing information and creative writing support. (OpenAI,2025; Radford et al., 2019)

Clinical Caritas Processes: These are the evolved version of Watson's carative factors and include practical steps that nurses can take to implement caring behaviors in clinical practice. They stress the importance of love, kindness, and authentic emotional presence. Empathy is embedded in several of these processes, particularly in fostering a trusting and emotionally supportive patient environment (Cara, 2003).

Cognitive Behavioral Therapy (CBT): CBT is a widely used psychological intervention that identifies and alters dysfunctional thinking and behavior patterns (Inkster et al., 2018; Boucher et al., 2021).

Emotional Responsiveness: Emotional responsiveness refers to how well an AI system reacts to its users' emotional states (Concannon & Tomalin, 2023).

Empathy: In the context of Watson's Theory of Human Caring, empathy is defined as the ability of the nurse to be emotionally present and fully engage with the patient's experiences, feelings, and emotions. Watson's approach to empathy emphasizes the nurse's responsibility to understand the patient as a whole person, beyond just their illness (Savieto & Leão, 2016). Empathy can be broken down into three distinct types.

I. Cognitive empathy is understanding another person's emotions and perspective without necessarily feeling the same emotions. It involves thinking about what someone else is experiencing and recognizing their feelings based on verbal and non-verbal cues. Cognitive empathy helps caregivers anticipate patient needs and respond appropriately while maintaining emotional objectivity. (Savieto & Leão, 2016).

II. Emotional empathy is the ability to share and feel another person's emotions. It allows a caregiver to connect emotionally with a patient, offering compassion and warmth. This type of empathy helps build trust and emotional support. (Savieto & Leão, 2016).

III. Behavioural empathy is the visible expression of empathy through actions, words, and gestures. It involves active listening, compassionate communication, and appropriate physical gestures that show understanding and support. Behavioural empathy ensures that empathy is not just felt but demonstrated through meaningful interactions. (Savieto & Leão, 2016).

Ethical Concerns in AI: Key ethical issues surrounding AI chatbots in mental health include data privacy, security, and the authenticity of emotional engagement. Concerns also arise around the potential for AI to deliver harmful or inappropriate responses, as well as its limitations in handling severe mental health issues (Rad & Rad, 2023).

Generative AI: Generative AI systems dynamically create responses based on user input by drawing on patterns learned from vast datasets. Generative artificial intelligence (AI) refers to machine learning models capable of producing new content that mimics human-created material. Unlike earlier AI systems that relied on structured, rule-based outputs, generative AI models use large datasets and deep learning to create naturalistic, contextually adaptive responses (Banerjee et al., 2024, Abd-alrazaq et al., 2019).

Large Language Model (LLM): An advanced type of artificial intelligence designed to understand and generate human-like text. It is trained on massive amounts of text data from diverse sources such as books, articles, and websites to learn language patterns, grammar, and context. The term "large" refers to both the vast size of the training data and the model's complex structure, often containing billions of parameters (adjustable parts of the model that influence how it generates text). LLMs use a neural network architecture called a Transformer, which allows them to process and understand the relationships between words and concepts over long text passages. This design enables LLMs to perform a wide range of language-based tasks, including answering questions, writing essays, translating languages, and even generating creative content, all while producing coherent and contextually relevant responses. (Open AI, 2025, Radford et al., 2019)

Perceived Empathy of Technology Scale (PETS): The PETS is a validated tool designed to measure users' perceptions of empathy in technological systems, including AI chatbots such as

ChatGPT. The Perceived Empathy of Technology Scale (PETS) is a 10-item, two-factor instrument developed to measure users' perceptions of a system's empathy. The two sub-scales are Emotional Responsiveness (PETS-ER) and Understanding and Trust (PETS-UT). Developed through expert interviews, focus groups, and user testing, PETS was validated across 22 distinct scenarios to ensure broad applicability. It offers a standardized tool for evaluating and advancing emotionally intelligent technologies (Schmidmaier et al., 2024).

Rule-Based Systems: Rule-based systems use predefined rules or decision trees to guide their interactions (Abd-alrazaq et al., 2019).

Stigma: Stigma refers to the negative stereotypes and social disapproval associated with mental health conditions, which can prevent individuals from seeking help (Kretzschmar et al., 2019; Balcombe, 2023).

Therapeutic Alliance: Therapeutic Alliance refers to the bond between a therapist and a patient, which is critical for effective therapy (Rubin et al., 2024).

Transpersonal Caring: Transpersonal care is central to Watson's theory and refers to a deeper, spiritual connection between nurse and patient that goes beyond the physical aspects of care (Cara, 2003; Savieto & Leão, 2016).

Challenges and Significance

There is a growing number of easily accessible and free generative artificial intelligence (AI) systems. Open AI developed ChatGPT, but Google, Meta and Amazon are rapidly developing competing technologies. These systems use generative AI and Large Language Models (LLM) to mimic human interactions and are designed to invoke empathy in their responses to users (Cuadra et al., 2024). The use of generative AI tools, such as ChatGPT, in mental health contexts is expanding rapidly, with applications ranging from wellness chatbots to

AI-enhanced therapy platforms (Lai et al., 2023). These tools are designed to simulate therapeutic dialogue, offer psychoeducation, and guide users through self-help techniques, such as cognitive-behavioural or mindfulness-based interventions (Boucher et al., 2021). AI-driven applications are also being used to triage symptoms, deliver mood tracking, or provide brief supportive interactions during emotional distress. These interventions appeal to users who value anonymity, affordability, and immediacy, particularly younger demographics who are already accustomed to digital interactions (Abd-alrazaq et al., 2019). Previous to the development of generative AI, mental health chatbots did not access generative AI and worked in a closed algorithmic system that made their responses predictable (Xian et al., 2024). In 2023, the National Eating Disorders Association was embroiled in controversy after it shut down its human-manned help line and put a chatbot named Tessa in its place. Tessa began as a closed algorithmic system and was deemed safe because she could not generate unique answers; however, when the system was allowed to update to a generative system, the service had to be shut down as it was offering dangerous weight-loss advice (Algre, 2025). Also, in an article from the New York Post, it was reported that the most recent update of ChatGPT had to be rolled back as it was noticeably more sycophantic, validating doubts, fueling anger and urging impulsive actions (Bardolf, 2025). With the rise of this newer technology, the safety rails have come down, and there is no longer any way to ensure that the responses to users of these open AI systems will not inadvertently cause harm (Balcomb, 2023). The research by Rubin et al. (2024) broke down Generative AI's strengths and weaknesses in the context of empathy and found that systems such as ChatGPT do a great job of mimicking cognitive empathy but fall short when trying to portray emotional empathy and lack the capacity for genuine emotional engagement.

Studying generative AI and empathy in mental health is crucial due to the significant positive and negative potential that challenges this quickly evolving technology. Unlike rule-based models, generative AI systems are designed to dynamically create responses based on user input by drawing on patterns from vast datasets. This flexibility allows for more nuanced and context-sensitive conversations, which is particularly valuable in mental health support, where patients often present complex emotional and psychological needs (Abd-alrazaq et al., 2019).

One of the primary challenges in using generative AI for mental health support lies in its limited ability to replicate the full scope of empathy, a core component of effective mental health care (Cara, 2003). While AI systems can mimic empathetic language, they often lack the depth of emotional engagement required to establish a therapeutic alliance comparable to that between human therapists and patients (Rubin et al., 2024). According to Jean Watson's Theory of Human Caring, empathy is not merely an expression of kind words but a foundational element in building trust, understanding, and genuine human connection that supports healing (Savieto and Leão, 2016). AI's inability to consistently recognize and respond to patients' deeper emotional and psychological needs raises concerns about the risk of offering superficial, or even harmful, responses to complex emotional issues, ultimately limiting their effectiveness in delivering truly compassionate care (Concannon & Tomalin, 2023).

While generative AI can simulate aspects of therapeutic relationships by providing supportive dialogue, it lacks the true capacity to replicate the depth of connection necessary for sustained therapeutic success. The therapeutic alliance is a vital predictor of positive outcomes in mental health care, and the inability of AI to truly "understand" a patient's emotions and motivations poses a barrier to its effectiveness in this field (Rubin et al., 2024). Drawing on Watson's Theory of Human Caring (Watson, 2008), authentic healing relationships are grounded

in transpersonal connections, these are interactions that go beyond surface-level exchanges to honor the whole person. These caring moments, rooted in empathy, presence, and mutual trust, are essential for meaningful therapeutic engagement. The study of empathy in AI is pivotal because it addresses a gap in the current AI system's ability to provide emotionally resonant care. By focusing on developing technologies that can better mimic or even embody empathetic behaviours, researchers can help bridge the emotional gap between AI and human therapists, making AI more effective in mental health interventions (Schmidmaier et al., 2024). Addressing empathy within generative AI has implications for improving patient outcomes, building trust in AI systems, and ensuring that these technologies can ethically and responsibly support individuals in their mental health journeys.

Another significant challenge involves ethical concerns about privacy and data security. Generative AI systems like ChatGPT rely on large amounts of personal data to provide contextually relevant responses, which raises questions about how this data is stored, used, and protected. These concerns are even more pressing in mental health contexts, where sensitive personal information is often shared (Rad & Rad, 2023). While generative AI shows promise in expanding access to mental health support, it also raises questions about emotional safety, misinformation, and lack of formal regulation (Balcombe, 2023). Bodies such as the Canadian Centre for Cyber Security and WHO have emphasized the need for oversight in AI-driven health tools, but generative models like ChatGPT remain largely unregulated in mental health contexts (Cuadra et al., 2024). Generative AI holds great promise for enhancing mental health care, its current limitations regarding ethical concerns make it essential to continue studying and refining these systems.

Despite these challenges, the potential significance of generative AI in mental health care cannot be overlooked. AI chatbot applications and generative AI such as ChatGPT offer scalable, accessible mental health support, especially in areas with shortages of mental health professionals (Balcombe, 2023). They can reduce stigma by providing anonymous, non-judgmental interactions, making it easier for people to seek help without fear of societal disapproval (Kretzschmar et al., 2019). Generative AI offers the possibility of continuous support, helping individuals manage symptoms of conditions like depression and anxiety outside traditional therapy sessions (Boucher et al., 2021). However, addressing ethical concerns such as privacy and data security is essential to building the trust necessary for users to fully engage with these systems. Watson's Theory of Human Caring emphasizes the moral and ethical responsibility of care providers to protect the dignity and well-being of those they serve (Watson, 2008). In the context of AI, this means ensuring that users have confidence that their sensitive personal information will be handled with integrity, especially when disclosing vulnerable aspects of their mental health (Rad & Rad, 2023). Without robust safeguards, AI systems risk alienating users and undermining the very accessibility and relational trust they are designed to enhance (Kretzschmar et al., 2019).

Thus, the study of empathy in AI is crucial because it addresses a gap in current AI systems' ability to provide emotionally resonant care. As noted in the development of the Perceived Empathy of Technology Scale (PETS), focusing on the emotional responsiveness and trust-building aspects of AI systems can help improve user engagement and therapeutic outcomes (Schmidmaier et al., 2024). By integrating better empathetic responses and addressing ethical and relational challenges, AI chatbots with generative capabilities, such as ChatGPT, can more effectively support individuals in managing mental health conditions. Generative AI shows

promise in supporting mental health conditions such as depression and anxiety by providing accessible, continuous care that complements human therapists and broadens access to support (Boucher et al., 2021; Rubin et al., 2024).

Purpose and Research Objectives and Questions

This study explores ChatGPT's capacity to exhibit empathy, as perceived by human users, through the lens of Jean Watson's Theory of Human Caring. This theory provides a robust framework for understanding the essential role of empathy in human interactions, emphasizing holistic care that addresses emotional, spiritual, and relational dimensions (Watson,2008). By applying Watson's framework, this study aims to assess how well AI systems align with the principles of empathy, trust, and respect for human dignity; these elements serve as a benchmark for assessing whether generative AI can offer interactions that reflect the depth and quality of human-centered care.

The first research objective in this qualitative study is to explore the level of perceived empathy in ChatGPT's interactions, using post-activity interviews to gather participants' subjective experiences. To enrich and support these findings, responses from the Perceived Empathy of Technology Scale (PETS) (Schmidmaier et al., 2024) will be incorporated as it captures user perceptions of key empathic qualities such as emotional responsiveness and trust. Although numerical in format, the PETS data will be interpreted descriptively and used to triangulate the interview findings, offering an additional lens through which to understand participants' experiences of empathy in their interactions with ChatGPT. The second objective is to apply Jean Watson's Theory of Human Caring as a framework for interpreting these interactions, focusing on how ChatGPT might replicate or fall short of the humanistic, relational, and caring dimensions of empathy that Watson's theory emphasizes.

The research questions guiding this inquiry are as follows: How is empathy perceived and experienced by human users during interactions with ChatGPT? And how can Jean Watson's Theory of Human Caring inform the assessment of empathetic responses from ChatGPT? By framing the study within Watson's theory, the research aims to maintain the integrity of empathy, ensuring that ChatGPT's interventions are technically sound and resonate with the deeper human need for emotional and spiritual connection in caregiving.

Summary

In summary, Chapter 1 has introduced the growing need for alternative mental health solutions due to increasing wait times and unmet care demands in traditional healthcare systems. The chapter outlined the definitions of key terms, presented the challenges and potential of generative AI in mental health, and emphasized the importance of empathy as viewed through Jean Watson's Theory of Human Caring. The purpose and objectives of this research have been established, aiming to evaluate perceived empathy in AI interactions and apply Watson's theoretical framework to guide the assessment of these technologies. With these foundational concepts in place, Chapter 2 will delve deeper into the existing literature, providing a comprehensive review of current research on AI, empathy, and mental health interventions and highlighting the gaps this study seeks to address.

Chapter 2: Narrative Literature Review

This chapter will review current research on perceived empathy in AI-driven mental health care, focusing on how these insights inform this study. This review includes an outline of the process used to identify relevant literature, a discussion of key findings, and a presentation of

the main themes that emerged. Additionally, it will address limitations and gaps in existing research to emphasize the significance of this study, particularly in evaluating empathy in generative AI through the lens of Jean Watson's Theory of Human Caring.

Databases Searched

This literature review was conducted using an iterative approach, beginning with broad searches in Google Scholar and the Royal Inland Hospital Library Services, accessing academic databases such as CINAHL, PsychINFO and PubMed. As relevant articles were identified, search terms were refined and expanded to capture emerging themes. Sources included major publishers (e.g., Springer, Wiley, Elsevier), open-access journals, university repositories, professional associations, and government websites. This flexible and exploratory process supported the development of a diverse and relevant body of literature to inform the study.

Utilization of Search Terms

To capture the relevant studies, search terms such as "generative AI," "mental health," "patients," "empathy," "nursing & generative AI", “ChatGPT”, “Chatbot” and “mental health apps” were used. This approach allowed for the inclusion of a wide range of perspectives and findings, encompassing academic research and practical insights from various fields related to the application of generative AI in mental health and nursing contexts. The selection criteria focused on identifying studies that provided valuable insights into the interaction between AI and patients, explicitly examining aspects of empathy and the potential role of AI in supporting mental healthcare.

Inclusion and Exclusion Criteria

Each general search on Google Scholar yielded approximately 100,000 hits. To narrow this down (except for seminal works or works to support other aspects of this thesis), only peer-

reviewed articles on generative AI published in 2018 or after met the criteria for use. In order to ensure the relevance and applicability of the literature reviewed, only studies published from 2018 onward were included. This five-year time frame aligns with standard academic guidance, which suggests that “current” or “recent” research typically refers to publications within the past five years (TRU Library Guides, n.d.) This is particularly important in technology-related fields, where developments evolve rapidly and older research may no longer reflect contemporary tools, ethical considerations, or user contexts.

Furthermore, ChatGPT and other generative AI tools only became publicly available in late 2022, with widespread access and public engagement beginning in 2023. As such, earlier studies would not have been able to capture the emergence of large language models in mental health applications. Limiting the review to post-2018 literature ensured that the sources reviewed were technologically relevant and reflective of the current digital landscape.

Search Outcomes

From the initially discovered articles, they were screened by title, abstract and articles that could be accessed for free, either published in or translated and published in English. A comprehensive iterative literature review was conducted using Google Scholar and related academic databases, initially identifying approximately 150 articles related to perceived empathy in AI-driven mental health care. These articles were further screened to ensure relevance to this study's objectives, focusing on recent, primary research that aligns with the theoretical framework of Jean Watson's Theory of Human Caring. Studies that didn’t match the six-year research timeframe and were not centred on AI interactions in mental health were excluded. This rigorous selection process resulted in the curated list of references used in this thesis, totalling 42 articles that relate directly to generative AI or AI systems in the context of mental health.

Figure 1 is a flow diagram that demonstrates the literature review process. The articles explicitly related to Artificial Intelligence and Mental Health are themed in Figure 2.

Figure 1

Flow Diagram for Literature Review

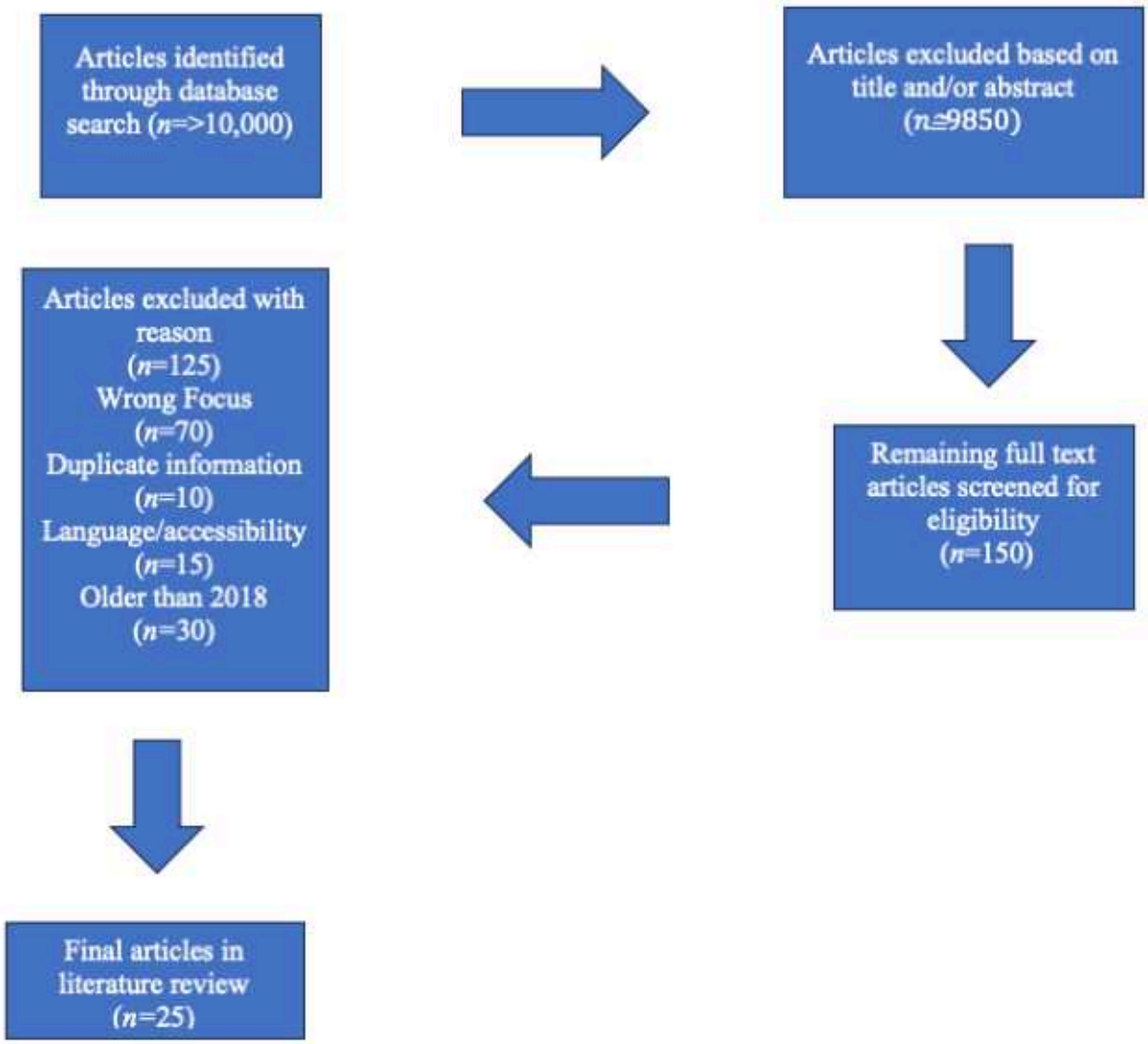
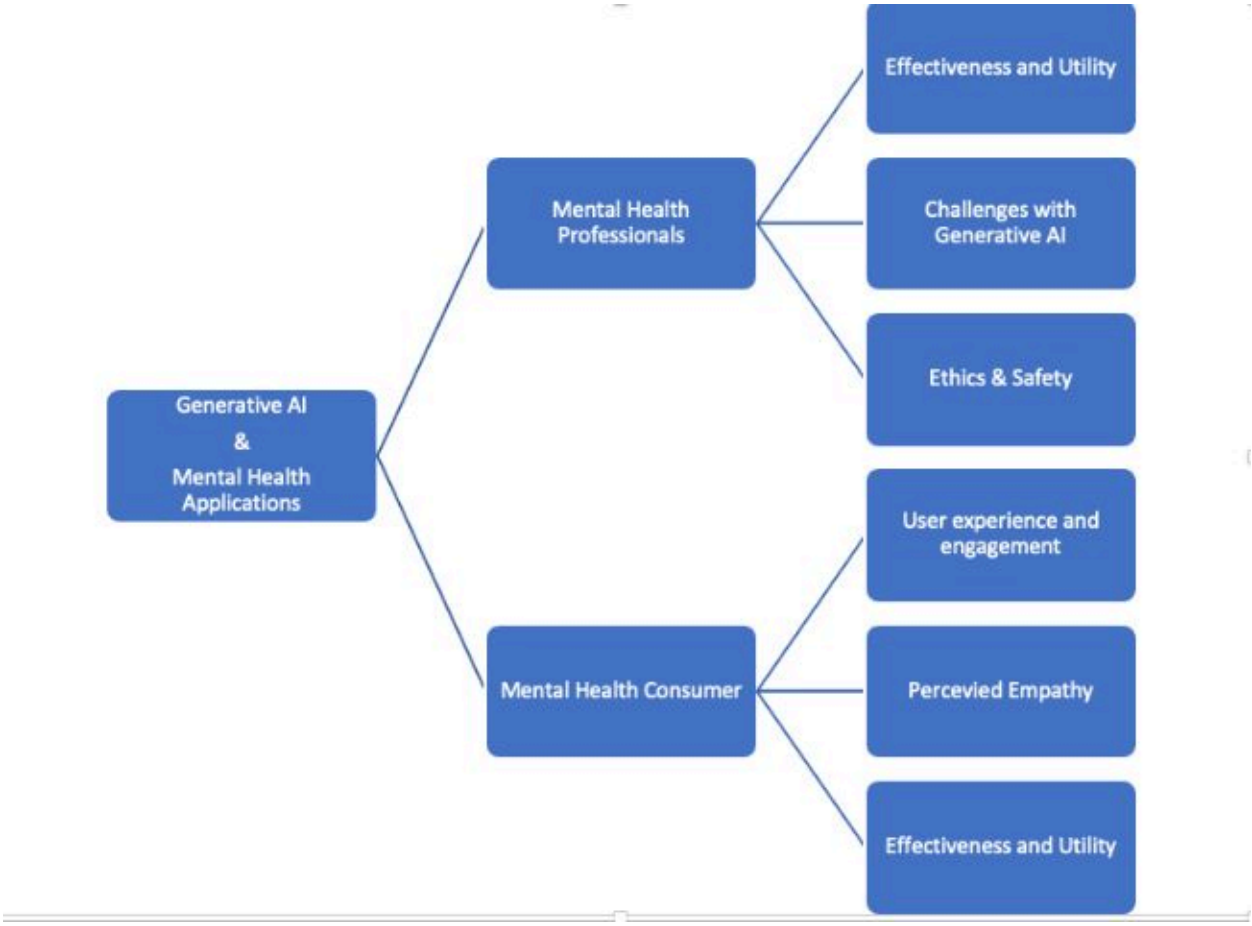


Figure 2

Themes of Literature Review



Themes from Literature Review

Introduction to AI in Mental Health Care

The integration of artificial intelligence (AI) in mental health care is rapidly evolving, with conversational agents (CAs) and generative AI models playing an increasingly prominent role. This development addresses several pressing global mental health issues, including access limitations, societal stigma, and the shortage of mental health professionals (Abd-alrazaq et al., 2019; Balcombe, 2023). Through scalable, accessible, and personalized support, AI has the potential to fill critical gaps in traditional mental health services. Nevertheless, ethical and

practical considerations surrounding AI's empathy and reliability are essential concerns, as accuracy, user trust, and over-reliance on technology may introduce new risks (Torous & Blease, 2023; Solaiman, 2024).

Effectiveness and Utility of AI Chatbots in Mental Health

AI chatbots have proven valuable in providing mental health support, augmenting access to therapeutic resources, and supplementing human-led care. Abd-alrazaq et al. (2019) conducted a scoping review categorizing mental health chatbots by purpose, platform, response generation methods, and therapeutic targets. Findings indicated that rule-based chatbots, though limited in flexibility, are still the most commonly implemented due to ease of use. However, they often lack the flexibility and empathy required for complex cases. Complementing this, Balcombe (2023) asserted that AI chatbots improve access to mental health care but should serve as supplements to human therapists. Similarly, Boucher et al. (2021) explored the diagnostic and monitoring roles of AI chatbots, noting that while accessibility improves, effectiveness in delivering empathetic care remains limited due to technological constraints.

More advanced systems, such as Lai et al.'s (2024) Psy-LLM, highlight AI's utility in mental health support. Psy-LLM effectively generates timely responses to psychological inquiries, potentially alleviating some of the burden on mental health professionals. However, challenges persist, as this model and similar LLMs struggle to address nuanced and complex mental health issues. This underscores that while AI can support therapists, it cannot yet replace the depth of understanding offered by human practitioners.

User Experience and Engagement

User engagement significantly influences the effectiveness of AI chatbots in mental health care. Abd-alrazaq et al. (2019) identified that elements such as avatars and virtual

representations can enhance user engagement by creating a sense of presence and attentiveness. Boucher et al. (2021) reported that users appreciate the non-judgmental nature of AI chatbots, which fosters openness and self-disclosure. However, they also noted that repetitive or scripted responses could lead to disengagement, suggesting a need for chatbots to employ adaptive interaction methods to maintain user interest.

Maurieello et al. (2021) further emphasized engagement in their study of the Popbots suite, a collection of AI tools offering daily micro-interventions for stress management. Findings showed frequent engagement was associated with reductions in depressive symptoms, indicating that personalized, brief interventions can promote mental well-being. Chin et al. (2023) expanded this perspective, revealing cultural differences in user engagement, with Eastern users expressing stronger emotions in chatbot interactions than Western users. This highlights the need for culturally sensitive, personalized responses to sustain engagement across diverse user groups.

Empathy and Emotional Safety in AI-Driven Mental Health Tools

Empathy is crucial for effective mental health interventions, yet remains challenging for AI systems to replicate. Inkster et al. (2018) evaluated Wysa, a chatbot employing empathy-driven techniques, and noted mood improvements among users, although some perceived response repetition as detracting from the chatbot's empathy. Rubin et al. (2024) analyzed AI's empathetic capacities, distinguishing between cognitive empathy (recognizing emotions) and emotional empathy (resonating with emotions). The study concluded that AI could simulate cognitive empathy to a degree but struggles to provide authentic emotional engagement, limiting its therapeutic potential.

Concannon and Tomalin (2023) introduced the Empathy Scale for Human-Computer Communication (ESHCC) to measure perceived empathy in AI, finding that superficial empathy

claims can lead to "credibility fallacies" that reduce user trust. Schmidmaier et al. (2024) built on this with the Perceived Empathy of Technology Scale (PETS), measuring empathy dimensions such as emotional responsiveness and trust. Findings indicate that while AI can mimic empathy, the lack of authentic emotional resonance remains a barrier to handling complex mental health needs effectively.

Messina et al. (2013) reinforces the significance of empathy in mental health care, noting that physiological concordance between therapist and patient enhances perceived empathy, especially among trained professionals. This insight underlines that empathy in mental health care is not merely linguistic but involves complex physiological and relational dynamics that current AI systems cannot fully replicate.

Ethical and Safety Considerations in AI for Mental Health

The ethical implications of AI in mental health are substantial, encompassing issues such as data privacy, interaction authenticity, and response reliability. Kretzschmar et al. (2019) explored young people's ethical concerns regarding mental health chatbots like Woebot and Wysa, noting concerns about data privacy, especially on platforms like Facebook, and limitations in the chatbots' emotional depth. Solaiman (2024) examined regulatory gaps in generative AI (GenAI) within mental health, emphasizing the limitations of frameworks like the EU AI Act, which currently lack specific mental health standards, exposing users to privacy risks and misinformation.

Torous and Blease (2023) further argued for mental health-specific regulations to address the unique risks associated with generative AI. While this technology could democratize access to mental health resources, it may also perpetuate biases and compromise user safety. Elyoseph

et al. (2024) highlighted that corporate ownership of GenAI technology may prioritize profit over patient-centred care, raising concerns about data misuse and user manipulation.

Challenges with Generative AI in Mental Health

Generative AI in mental health care faces unique challenges related to coherence, emotional depth, and context-sensitive responses. Lai et al. (2024) found that Psy-LLM, though effective for initial screenings, struggled with cases requiring nuanced emotional engagement. Yang et al. (2023) explored the interpretability of LLMs like GPT-3.5, noting strengths in emotion detection but variability in performance, especially with complex, context-dependent scenarios. Concannon and Tomalin (2023) highlighted "credibility fallacies" in empathy simulation that can erode therapeutic alliance, while Rubin et al. (2024) emphasized the limitations of generative AI in achieving deep emotional empathy, a core component of meaningful therapeutic relationships.

Theoretical Framework: Jean Watson's Theory of Human Caring

Jean Watson's Theory of Human Caring offers a valuable framework for assessing empathy in AI mental health interactions. Watson posits that empathy, a deep form of understanding and caring, is a foundational aspect of effective healing and is central to genuine human connection (Watson & Woodward, 2010). This theory emphasizes a humanistic-altruistic system of values, promoting the expression of both positive and negative feelings as essential to healing (Cara, 2003). Watson further argues that empathy is more than a cognitive process; it involves a holistic engagement that includes emotional and spiritual connections with patients, fostering authentic therapeutic relationships (Watson & Woodward, 2010).

In applying Watson's Theory to generative AI, this research evaluates AI’s capacity for empathy, using it as a lens to gauge AI’s ability to respond empathetically to users seeking

mental health support. By defining empathy as both cognitive and emotional, Watson's model sets a high standard, underscoring that empathy in mental health requires more than just linguistic responses; it involves presence, attunement, and the ability to form meaningful, compassionate relationships. In this context, AI's limitations in replicating emotional resonance become clear, reinforcing the importance of a human-centred approach to mental health care.

Summary

In summary, the literature underscores AI's potential to expand access to mental health support while highlighting persistent limitations in user engagement, empathy simulation, and ethical challenges. While capable of recognizing emotions and delivering supportive responses, current AI models often fall short in replicating deeper levels of emotional engagement. Many studies, such as Rubin et al. (2024), distinguish between cognitive empathy (understanding another's emotions) and emotional empathy (resonating with another's feelings). Most generative AI models, including ChatGPT, are primarily limited to cognitive empathy, responding based on language patterns without a genuine emotional connection. This gap highlights the difficulty of achieving true emotional resonance, a core aspect of effective therapeutic relationships, which limits AI's perceived authenticity and relational depth in mental health settings. Generative AI, while promising, still struggles with emotional depth and consistency, limiting its effectiveness in complex mental health contexts. Watson's Theory of Human Caring provides a robust framework for examining empathy in AI, setting a standard for understanding the limitations and potential of AI in fostering genuine therapeutic relationships. Addressing these challenges is essential to ensure that AI applications in mental health care remain safe, ethical, and supportive of user well-being. These research gaps underscore the need for continued investigation into how generative AI can be optimized for mental health support. Advances in empathy simulation,

adaptability for cultural contexts, ethical guidelines, and refined measurement tools are essential to fully realize AI's potential in providing empathetic, safe, and effective mental health care. By addressing these areas, future research can work toward developing AI systems that understand, resonate with, and support users authentically, fostering a new paradigm in mental health support.

Chapter 3: Methodology

Purpose

This study employed thematic analysis as outlined by Braun and Clarke (2006), a flexible and widely used approach for identifying and interpreting patterns of meaning within qualitative data. Thematic analysis was selected for its suitability in exploring subjective experiences and perceptions, particularly in relation to empathic responses from AI. The study followed Braun and Clarke's six-phase process, first becoming familiar with the data, then generating initial codes, before searching for and reviewing themes, defining and naming those themes, and finally producing the analytic narrative. This iterative process supports the development of rich, contextual themes grounded in participants' experiences.

Braun and Clarke (2021) also emphasize that reflexive thematic analysis (RTA) allows for rich, in-depth engagement with complex human experiences, rather than seeking to combine multiple methodological paradigms. Braun and Clark's original work with thematic analysis (2006) was expanded on, and RTA was born to better capture the process. While this study is rooted in a qualitative design, it incorporates structured scale data from the Perceived Empathy of Technology Scale (PETS) (Schmidmaier et al., 2024) to enrich and support the qualitative findings. This qualitative-driven approach investigates perceived empathy in ChatGPT's responses to a simulated mental health challenge by exploring participants' subjective

experiences through post-activity interviews and supplementing these insights with structured responses from the PETS. Although statistical analysis is not appropriate due to the small sample size, the PETS contributes to methodological triangulation, offering an additional lens for understanding participants' perceptions of key empathic qualities such as emotional responsiveness and trust in a format that complements the thematic data. This integration offers a more fulsome understanding of how empathy is perceived in AI-driven interactions and the factors that shape that perception, particularly in the context of mental health support.

Jean Watson's Theory of Human Caring (2008) was also used as a guiding theoretical framework in this study. The theory informed the design of the semi-structured interview guide, with questions shaped to explore key constructs such as empathy, presence, trust, and emotional connection, all elements central to Watson's carative factors. In this way, the framework was applied deductively, ensuring conceptual alignment with the study's focus on perceived empathy. However, the analytic approach followed Braun and Clarke's (2006, 2021) thematic analysis, which emphasizes researcher subjectivity, iterative engagement with the data, and the construction of meaning through interpretation. Importantly, themes were not imposed based on Watson's theory, but were developed inductively, allowing patterns to emerge directly from participants' descriptions of their experiences. This dual approach provided both conceptual grounding and analytical openness, aligning with qualitative research values of depth, reflexivity, and contextual understanding.

Description of Research Design and Approach

This study follows a qualitative research design using thematic analysis, as outlined by Braun and Clarke (2006), with methodological triangulation provided by the Perceived Empathy of Technology Scale (PETS). Methodological triangulation was employed by comparing

quantitative scores from the Perceived Empathy of Technology Scale (PETS) with qualitative themes that emerged from post-interaction interviews. Although participant responses were collected anonymously and could not be directly linked, both data types explored the same phenomenon, perceived empathy in AI interactions, allowing for an integrated understanding through cross-method comparison. In addition to semi-structured interviews, PETS (Schmidmaier et al., 2024) was employed to descriptively capture participants' perceptions of empathy during their interactions with ChatGPT. Although PETS is traditionally a quantitative instrument, it is used in this study to complement and enrich the qualitative analysis, offering an additional perspective to support the interpretation of emergent themes.

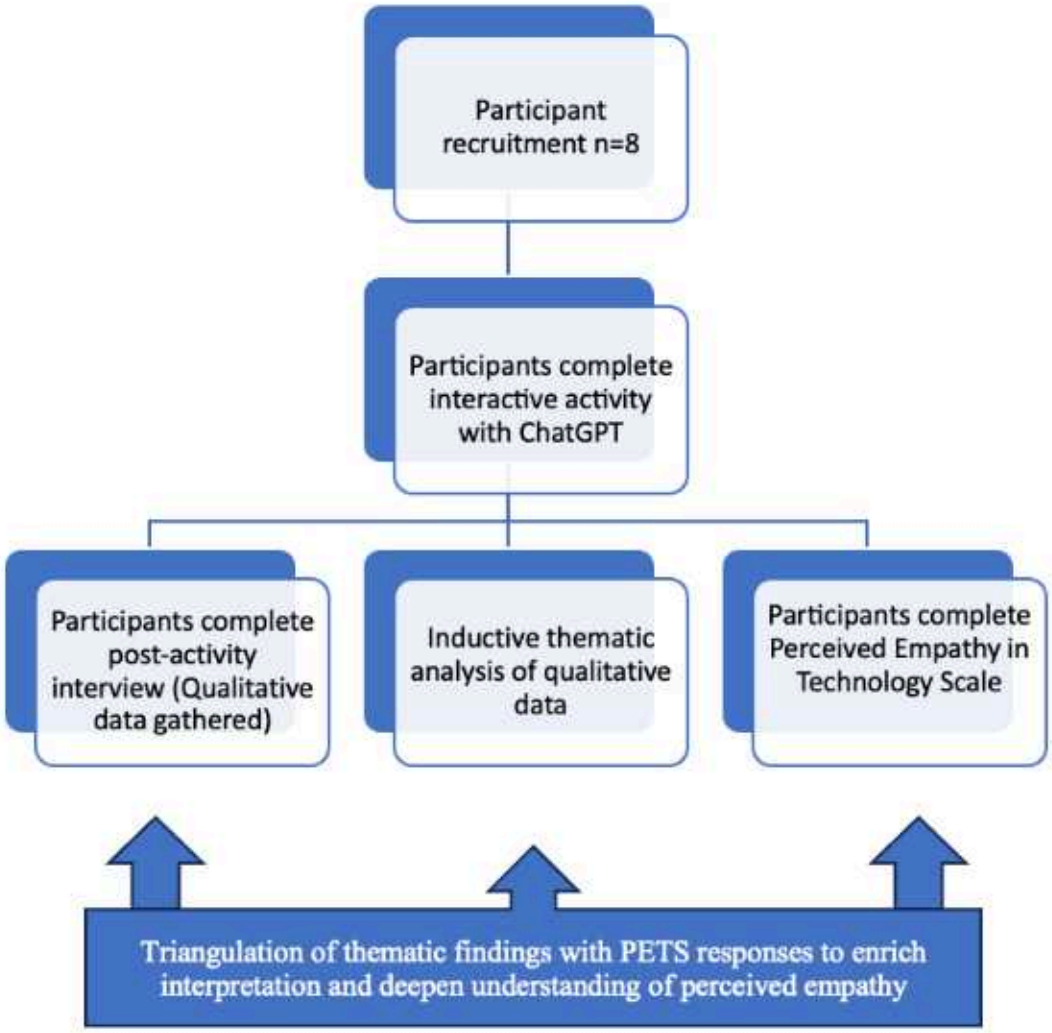
This study involves participants completing an interactive activity with ChatGPT (OpenAI., 2025), filling out a Perceived Empathy of Technology Scale (PETS) (Schmidmaier, et al., 2024) and engaging in a post-activity interview with the primary researcher. The PETS was completed as a survey on SurveyMonkey, an online survey platform (SurveyMonkey Inc., n.d.) and captured the subjective experiences of empathy, by using a 0–100 sliding scale, where 0=strongly disagree and 100=strongly agree. This means lower scores indicate less perceived empathy and higher scores indicate more. This offered a structured yet personal way for participants to capture and reflect on the subjective experience of the interactive activity and how emotionally understood and supported they felt during the exchange. For a full description of the PETS (Schmidmaier, et al., 2024) please refer to Appendix A. The virtual interview over Microsoft Teams (Microsoft Corporation, n.d.) following the interactive activity was used to gather participants' reflections on their experience interacting with ChatGPT. This narrative data was complemented by responses to the Perceived Empathy of Technology Scale (PETS), offering a structured account of their subjective experience during their interaction ChatGPT.

Using Microsoft Teams aligns with the recommended practices for remotely gathering qualitative data (Archibald et al., 2019). It also aligns with Jean Watson's framework (Cara, 2003) as this approach promoted psychological safety and emotional ease, creating conditions that support open, reflective dialogue essential for exploring perceptions of empathy. Similarly, the use of SurveyMonkey to administer the Perceived Empathy of Technology Scale (PETS) aligned with Watson's Theory of Human Caring by providing a secure and anonymous way for individuals to share their reflections at their own pace, in a manner that respected their comfort, autonomy, and emotional safety (Cara, 2003).

For this study, the generative AI is ChatGPT. This study used ChatGPT powered by GPT-4o, accessed via the OpenAI platform (OpenAI, 2025), and for the sake of consistency, it will be referred to as ChatGPT in the remainder of the writing. Collecting and analyzing participants' subjective experiences through both the post-activity interview and the PETS contributes to a more robust and credible study (Klassen et al., 2012). While the PETS is structured in format, it captures individual reflections on perceived empathy, which offers an additional layer of insight that supports methodological triangulation. This approach allows the researcher to explore areas where participants' PETS responses align with or differ from their narrative reflections, strengthening the interpretation of empathy-related themes. The integration of these two forms of participant input reinforces and validates the findings generated through inductive thematic analysis (Braun et al., 2006), offering a fuller picture of how empathy is experienced by participants interacting with ChatGPT. Figure 3 lays out the steps followed to gather and analyze data and illustrates the research design from participant interaction to thematic triangulation.

Figure 3

Research Design: From Participant Interaction to Thematic Triangulation



Ethical Considerations

The study has been approved by Thompson Rivers University's Research Ethics Board (REB) (REB Ref No: 10669), ensuring compliance with guidelines and ethical standards for human research set out by the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) (Tri- Council Policy Statement, 2022). The TCPS2 sets guidelines that address essential ethical concerns regarding consent, risk mitigation and confidentiality. This baseline for ethical research in Canada does not provide a framework for relational depth,

emotional presence or empathy, all critical research components that align with Jean Watson's Theory of Human Caring. Grounding the research in Jean Watson's Theory of Human Caring ensures that ethical considerations beyond consent and confidentiality are safeguarded. The Tri-Counsel Policy Statement (TCPS2) (2022) ensures participant well-being is prioritized throughout the research process, but Watson's Carative Factors emphasize the importance of creating a caring, emotionally safe, as well as an ethically sound environment (Cara, 2003). Both frameworks guided the study design, with participant protection as a central focus. This aligns with the ethical principles of the TCPS2, as well as Watson's Theory of Human Caring, which emphasizes promoting human dignity and a sensitivity to participants' emotional and relational needs (Watson & Woodward, 2010). Participants were provided detailed study information, including opportunities to ask questions before consenting. They were explicitly informed of their right to withdraw at any stage without penalty, reinforcing the ethical commitment to freedom of choice and respect for autonomy (Watson, 2008, Tri-Council Policy Statement, 2022). Recognizing the potential for sensitive or triggering content, the study incorporated mental health resources and referrals for participants who experienced distress. This reflects Watson's emphasis on creating a healing environment (Watson, 2008), prioritizing emotional and psychological safety, which goes beyond the TCPS2 mandates of risk, consent, justice and autonomy. To protect participant privacy, all data is securely stored, and the data used in the thesis document is anonymized, ensuring ethical handling of sensitive information following the principles of honesty and trust-building in therapeutic interactions, as well as the honesty and trust set out in the guidelines for ethical research practice (Watson, 2008, Tri-Council Policy Statement, 2022). Given the nature of the study, individuals with preexisting mental health conditions or vulnerabilities were excluded through a structured screening process. This aligns

with Watson's Carative Factor of assisting with human needs to prevent harm (Cara,2003); it also aligns with the TCPS2's mandate for the minimum standard of participant safety, ensuring that participation did not negatively impact participants.

Multiple safeguards were implemented to uphold these values and key ethical measures, which include:

- Informed Consent: Consent was obtained electronically prior to the start of the study, with verbal reconfirmation before screening and at the start of the post-activity interviews.
- Confidentiality: All participant data used in the thesis is anonymized, and each participant is assigned a unique code to protect personal information. PETS responses were collected via SurveyMonkey. The survey was distributed via an anonymous web link to ensure anonymity, and no identifying information such as names or email addresses were collected. Interview recordings are stored securely on Microsoft OneDrive. This platform uses encryption and two-factor authentication to access and is accessible only to the primary researcher and thesis supervisor.
- Participant Right to Withdraw: Participants were informed of their right to withdraw at any stage without penalty. If they chose to exit the study, they could decide to remove or retain their anonymized data up until the defence of the thesis. Each participant was made aware of the projected timeline for pulling their data from the study. This was outlined in the consent form and reiterated at the time of screening. Data Retention and Disposal: Following ethical guidelines set out in the TCPS2 (Tri- Council Policy Statement, 2022) and in compliance with the Thompson Rivers Ethics Review Board (REB File

N0:10669), data will be securely stored for five years after the study's completion. After that, digital data will be permanently erased, and any physical copies will be destroyed.

- Support Resources: To mitigate psychological risks, participants were given a list of mental health resources in case of distress. These resources, centred around the study's geographical location, include local crisis lines, the Interior Health intake number, and Kelty's Key, a free self-directed mental health platform. The resources are accessible to participants in the region where the study took place and include in-person and online resource options.

Participants

The study aimed to recruit ten healthy adult participants aged 19 to 65 with no mental health diagnosis. The age range was chosen by the researcher in consultation with the supervisor and was kept at this range for the following reasons. The study was approved for adults, so no one under the age of 19 was screened for the study, and the researcher wanted to ensure that participants would be able to manage the technology. As the activities were self-directed and completed on the participants' own electronic devices at their own convenience, setting the age range from 19 to 65 ensured that the participants would be safe and feel confident that they could complete the required activities without undue stress. Braun and Clark (2021) noted that small sample sizes are appropriate in qualitative-driven research, where the focus is on exploring subjective experiences. Additionally, the use of structured tools such as surveys can serve to support and contextualize emergent themes by offering complementary insights that enhance the interpretation of qualitative data (Polit & Beck, 2020). Braun and Clark (2021) also purport that a small sample size aligns with the qualitative research standard that seeks data adequacy. The goal is not to reach a saturation point where no new information emerges but to generate rich,

interpretive insights. Drawing on Braun and Clark's (2021) recommendations for qualitative sample sizes and considering the time constraints of the master's program, coupled with the focused scope of the research, a sample size of 10 participants was anticipated to achieve data adequacy. After two months of recruitment efforts, the study concluded with eight participants.

To expedite the selection process, the invitation to participate was first sent out through the email channels at Thompson Rivers University to those actively engaged in academia; from there, a snowball method to engage participants was employed, and anyone 19 to 65 who expressed an interest in participating was screened for eligibility. Participants are required to simulate a mental health challenge safely; thus, those currently experiencing active mental health issues were excluded. The screening ensured that participants could engage in the study without risk to their emotional well-being. Participants were screened to ensure they met these criteria, with an initial phone screening conducted by the primary researcher, a licensed mental health clinician with over a decade of clinical experience. At the end of the recruitment timeframe, eight healthy participants were identified from the screening process and accepted into the study.

Recruitment Process

Initial recruitment occurred through email channels at Thompson Rivers University, using an approved flyer outlining study goals, participant requirements, and the time commitment (See Appendix B). Interested candidates were invited to contact the primary researcher via email to express interest. Snowball sampling (word of mouth) was also employed to reach potential participants outside the university. Upon expressing interest, potential participants engaged in a 15-minute phone screening conducted by the primary researcher. This screening employed a brief psychosocial history (See Appendix C) to assess mental health status and confirm the absence of current active mental health conditions. To ensure the safety of

participants, eligibility was confirmed if the participants answered no to the list of historical mental health conditions and did not indicate they were experiencing any current challenges in their day-to-day functions. The psychosocial history screen assessed seven domains related to mental health and consisted of questions designed to paint a picture of the potential participants' present and past mental health. The questions asked about emotions, behaviour, social support, substance use, occupational or academic functioning, and previous mental health challenges, as well as assessed the risk of suicidal, homicidal or self-harming ideation or behaviour. If individuals did not meet the criteria, they received a written notification via email and a list of local mental health resources in case the screening experience had any impact. Following the psychosocial screening, once participants were deemed eligible and consented to participate in the study, the primary researcher conducted a role-play of a simulated mental health condition (see Appendix D). This demonstration was designed to help participants understand the structure of the interactive activity and what was expected of them during their conversation with ChatGPT.

Consent Process

Eligible participants received a comprehensive consent package via email, which included:

- Informed Consent Form (See Appendix E): This form outlined the study's purpose, detailed procedures, potential risks, the voluntary nature of participation, and confidentiality measures. It also includes information on the right to withdraw at any time.
- Study Procedure Overview: Participants were briefed on all study activities, estimated timelines, and data collection methods (AI interaction, PETS survey, post-activity

interview). If they were interested, they received an email with the consent form and this had the study procedures. As well, at time of screening this process was reviewed on the phone.

- Researcher Contact Information: The primary researcher's contact details were provided in the consent form for questions and additional clarification.

Before beginning the study, participants electronically signed and returned the informed consent form. During the phone screening, the primary researcher reaffirmed verbal consent and emphasized the participants' right to withdraw or refrain from any part of the study without penalty.

The Thompson Rivers University Ethics Review Board (REB) file N0:10669 approved the study's methodology. To comply with the ethics board and to maintain ethical integrity, several steps were taken to ensure that the potential risk of harm to participants was kept at an absolute minimum. Once a participant had completed the phone screening process, the accepted participant was assigned a unique code to anonymize all further data gathered by the primary researcher, (later in the findings section, the participants are given pseudonyms). Except for the demographic information gathered during screening, all other personal information collected to determine eligibility was not used in the study. The unique code replaced the names of the participants when they were screened for recruitment. There is a master key linking the unique code to participants that is known only to the research team and is saved on Thompson Rivers University Microsoft Cloud Drive attached to the primary researcher's university Microsoft account. All participants were informed of the purpose of the study, their right to withdraw from the study at any time, and their ability to pull their data from the study results up until the time of knowledge transfer. Consent was obtained in writing prior to starting the study and again

verbally each time the primary researcher interacted with the participant. Signatures on the consent form were acknowledgment that participants understood their participation in the study was strictly voluntary and that they had the right to withdraw at any time. In compliance with the REB, all raw data and any other information used for the purpose of this research will be saved electronically in the Thompson Rivers University Microsoft Cloud Drive for five years and then destroyed.

Data Collection Procedures

Following consent, participants engaged in a structured interaction with ChatGPT, assuming the role of "Sam", a 28-year-old teacher experiencing anxiety and burnout based on a case study given to each participant. During the initial screening, participants were given a brief verbal overview of the research and its purpose. At that time, I provided a general explanation of the activity and answered any questions. Participants were then invited to read the detailed written instructions independently (See Appendix F), which outlined how to engage with the AI. They were encouraged to decide for themselves how best to interact with the AI system, allowing for naturalistic and individualized responses. This approach supported participant autonomy and reflected the exploratory nature of the study. To support realistic therapeutic conversation, participants input four structured prompts designed to guide ChatGPT's responses through Sam's primary concerns (e.g., job-related stress, burnout, relationship impacts), which were consistent and standardized among all participants.

Participants were instructed to interact as 'Sam' by inputting each prompt, reflecting on the ChatGPT's responses, and continuing the dialogue in a conversational manner. Each interaction aimed for at least three to five exchanges per prompt, fostering a dynamic and

engaging conversation. The case study/pseudo-patient ‘Sam’ was purposefully kept gender-neutral to limit the possibility of gender bias in ChatGPT’s system.

The prompts covered various aspects of Sam's experience, from initial expressions of anxiety to exploring burnout's impact on relationships and seeking advice on managing these challenges. Each prompt was designed to help participants input information that ChatGPT would respond to empathically, offer support for, and provide actionable advice to. The prompts were designed to elicit responses from ChatGPT so that participants could reflect on subjective and objective measures of the perceived empathy experienced during the interactive activity.

Once the participants completed the interactive activity with ChatGPT, they moved on to the survey and the post-activity interview. Once the participant completed the interaction with ChatGPT, the suggested timeline for the interactive activity, survey and post-activity interview was one week; this would ensure that the information was fresh in the participant's minds and accommodated the busy schedules of the primary researcher and the participants. The link to the PETS was in the activity instructions, and participants were encouraged to complete the PETS as soon as the interaction with ChatGPT concluded, ensuring the interaction was fresh in their minds. When notification of the PETS completion came in from SurveyMonkey, and the participant had sent the ChatGPT interaction back to the primary researcher, the post-activity interview was booked as soon as possible.

Participants submit their PETS responses via SurveyMonkey, ensuring efficient and anonymous data collection (Regmi et al., 2017). The SurveyMonkey platform was chosen because of its convenient and reliable data management; Regmi et al. (2017) found that online survey formats such as SurveyMonkey protect against lost data, make data transfer easier and are user-friendly, making the surveys more attractive to complete for study participants. Along with

the survey, each participant also engaged in an approximately 30-minute follow-up interview conducted remotely via Microsoft Teams. While Archibald et al. (2019) specifically highlight the use of Zoom in qualitative research, many of the advantages they outline, such as accessibility, integrated recording/transcription features, and participant comfort also apply to Microsoft Teams. In addition, Teams offered enhanced security and compliance with institutional protocols, as it operates within the password-protected, two-factor authentication environment of the university's Microsoft 365 Suite the primary researcher has access to. These features make it a secure and practical choice for conducting interviews while respecting participants' privacy and maintaining data integrity. The semi-structured, post-activity interview (Appendix G) was used to collect the subjective experiences of the participants and their insights following their interaction with ChatGPT. The interview explored areas such as:

- Empathy Perception: Participants discuss whether they felt ChatGPT's responses were empathetic, supportive, and aligned with their needs as Sam.
- Support and Comfort: Participants reflected on how comfortable they felt with ChatGPT, including whether it provided an emotionally safe space for self-disclosure.
- Practicality of ChatGPT's Suggestions: Participants evaluate the usefulness of ChatGPT's advice, particularly in terms of effectively addressing Sam's concerns.

With the participants' consent, interviews were recorded and transcribed automatically through the Microsoft Teams platform. To protect privacy, all recordings are stored in Microsoft OneDrive, which uses encryption and strict access controls to protect sensitive data and are accessible only to the research team.

While the chat transcripts between participants and ChatGPT were not analyzed as part of the formal data set, they were collected for process verification purposes. These transcripts

allowed the researcher to confirm that each participant completed the interaction with the AI and engaged meaningfully with the task. Participants were not evaluated based on the content of their conversations, and no thematic or linguistic analysis of the transcripts was undertaken. This decision reflects the study's primary focus on participants' perceived empathy, as captured through PETS scores and post-activity reflections, rather than on the content or quality of the AI's specific responses.

To ensure rigour in this qualitative study, the trustworthiness framework proposed by Lincoln and Guba (1985), as interpreted by Tobin and Begley (2004), was adopted. Tobin and Begley argue that applying criteria developed for quantitative research may be inappropriate and can undermine the legitimacy of qualitative inquiry. Instead, Lincoln and Guba (1985) propose the criteria of credibility, transferability, dependability, and confirmability to evaluate the trustworthiness of qualitative research. In this study, participants' subjective experiences of perceived empathy were primarily captured through post-activity interviews. To enhance credibility and support metheological triangulation, participants also completed the Perceived Empathy of Technology Scale, (Schmidmaier, et al., 2024) which provides a structured yet subjective reflection on the empathic aspects of their interaction with ChatGPT. While the Perceived Empathy of Technology Scale is numeric in format, it was used to complement the thematic findings from the interviews. The PETS, is a 10-item instrument developed by Schmidmaier et al. (2023) and captures two distinct but related dimensions of perceived empathy: Emotional Responsiveness (The system considered my mental state) and Understanding and Trust (The system understood my needs). Each item is rated using a 100-point slider scale (1 = strongly disagree, 100 = strongly agree), allowing for sensitive measurement of user experience. This study employed a qualitative design focused on gathering

feedback that authentically reflected participants' perspectives. Efforts were made to ensure that interpretations emerged directly from the post-activity interviews and the PETS survey, rather than researcher bias supporting the criteria of credibility and confirmability.

Credibility (Tobin & Begley, 2004) in this study was supported through multiple forms of triangulation, each contributing to a layered and trustworthy interpretation of perceived empathy in AI-mediated mental health support.

Data triangulation was achieved by examining participants' subjective reflections during post-activity interviews alongside their responses on the Perceived Empathy of Technology Scale (PETS). The PETS survey offered a structured, individualized assessment of empathic perception, while the interviews provided richer, contextualized narratives. The PETS survey consisted of 10 questions, ranging from "The System understood my needs" to "I trusted the system," which allowed specific PETS items to be mapped against the emergent themes of the post-activity interview. This process allowed for comparison across data types, highlighting areas of convergence and divergence in participants' experiences.

Methodological triangulation was employed by integrating the data from the quantitative survey and the qualitative interview. This combination enhanced interpretive depth and provided multiple angles from which to examine the phenomenon, ensuring that findings were not dependent on a single method or perspective.

Theoretical triangulation was guided by Jean Watson's Theory of Human Caring, which emphasizes authentic presence, relational trust, and the honouring of human dignity (Watson & Woodward, 2010). This framework shaped both data collection and analysis by offering a consistent philosophical lens through which empathic interactions were evaluated, ensuring that

empathy was understood not only as a perceived quality but also as a relational and ethical construct.

To further enhance credibility and dependability (Tobin & Begley, 2004), ongoing consultation with the thesis supervisor was integrated throughout the research process. This included critical feedback on data interpretation, support for reflexivity in the analysis, and a review of the alignment between emerging findings and the theoretical framework, thereby contributing to the integrity and transparency of the study.

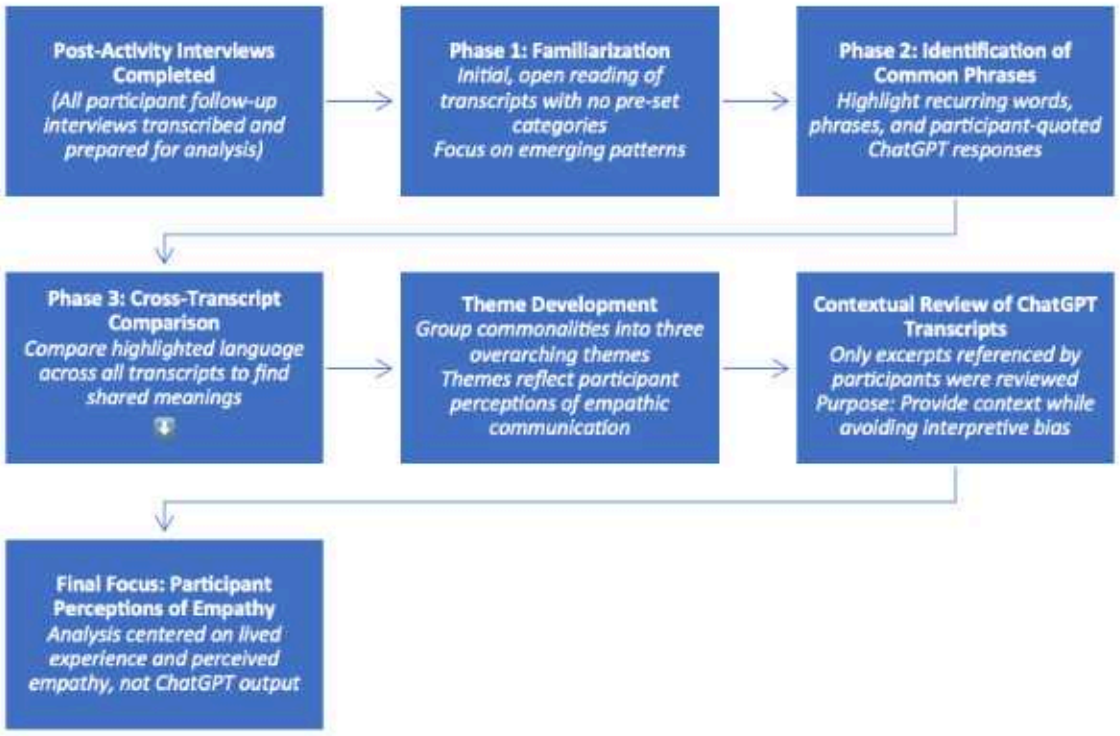
Additionally, a broad range of existing literature on AI, mental health and empathy underpins this research and can be used to compare and contrast the findings of this study within a wider body of evidence to enhance confirmability (Tobin & Begley, 2004). Laying out the methodology and documenting the research process provides transparency and allows others to follow the study's logic. Transferability (Tobin & Begley, 2004) was addressed by offering detailed descriptions of the research methodology and data collection methods, enabling others to assess the applicability of the findings in future research. Finally, reflexivity (Tobin & Begley, 2004) was an ongoing practice, with continuous self-examination of potential biases and perspectives, contributing to the trustworthiness (Tobin & Begley, 2004) of the findings and ensuring that conclusions were drawn authentically from participants' voices and the data collected. The primary researcher was cognizant of potential bias and as a result kept the analysis to only the reflective experience of the participants and did not analyze any of the transcripts of the interaction between the participants and ChatGPT. Applying trustworthiness as the scaffold for this study enhances the understanding of the data, giving it a richer overview of the complex interplay between ChatGPT and human perception of empathy.

Data Analysis Process

The analysis followed Braun and Clarke's (2006) six-phase process for thematic analysis, which provides a structured yet flexible framework for identifying and interpreting patterns within qualitative data. In the first phase, the researcher became deeply familiar with the data by reading and re-reading transcripts, making initial notes to capture potential areas of significance. The second phase involved generating initial codes from the data that captured meaningful features relevant to the research question. During the third phase, these codes were examined and grouped to identify broader patterns of meaning that reflected shared experiences or interpretations among participants. In the fourth phase, the themes were reviewed and refined to ensure they were coherent, distinct, and accurately represented the coded data. The fifth phase involved defining and naming the themes. In the sixth phase, the themes were synthesised into a final analytic narrative that illustrated how participants perceived and experienced empathy in their interactions with the AI. Braun and Clarke (2006) state that thematic analysis helps explore subjective experiences and is particularly useful for studies exploring perceptions. Thematic analysis is an active process (Braun & Clarke, 2006) that lends itself well to summarization and comparison across data sets while being suitable for researchers with varying levels of experience. Figure 4 lays out the process for the thematic analysis in steps 1 through 6.

Figure 4

Inductive Thematic Analysis Procedure



Summary

Chapter 3 outlines the methodology employed in conducting the research. It summarized the purpose, design, recruitment, participants, consent, ethical considerations, data collection procedures, and data analysis process. It illustrated how the research utilised Jean Watson's Theory of Human Caring as the framework to support the study's focus and reliability while investigating perceived empathy in responses from ChatGPT in the context of mental health challenges. Chapter 4 will lay out the findings by analyzing qualitative data collected and themed from the post-activity interviews and will use the survey responses to support the methodological triangulation.

Chapter 4: Findings

This study examines ChatGPT's perceived empathy through the lens of Jean Watson's Theory of Human Caring, which emphasizes holistic, relational, and emotionally attuned interactions (Watson & Woodward, 2010). The subjective impression from the reflections of the participant's post-activity interviews explores how ChatGPT's responses align with or diverge from Watson's principles of caring. Emotional responsiveness, trust, and relational depth are widely recognized as core components of empathy, particularly within therapeutic and supportive relationships (Austin & Boyd, 2015). These elements contribute to a person's sense of being emotionally understood and relationally safe, key aspects of perceived empathy. This study centres on how participants perceived empathy during their interactions with ChatGPT; it captures the reflections of presence, connection, and care in their subjective experience. Participants' interpretations of empathy reflected key elements of Jean Watson's Theory of Human Caring, such as relational trust, authentic presence, and dignity (Watson, 2008). These concepts were not used to guide coding but became relevant during interpretation. This post-analytic use of theory aligns with Braun and Clarke's (2021) reflexive approach, allowing theoretical ideas to emerge organically and enrich the data's understanding.

While not used as a primary measure, the Perceived Empathy of Technology Scale (PETS) served to triangulate participants' experiences, offering additional insight into how empathy was perceived in the interaction (Schmidmaier et al., 2024). The interview data were analyzed using inductive thematic analysis to explore participants' subjective experiences of empathy during their interaction with ChatGPT. Responses from the Perceived Empathy of Technology Scale (PETS) were reviewed descriptively and used to support the interpretation of

emerging themes. The survey responses complemented the interview findings by offering a structured reflection of how participants perceived emotional responsiveness and understanding/trust; enhancing the overall understanding of empathy within the ChatGPT interaction.

Participants

During the initial recruitment phase, 12 participants expressed interest in completing the research activity. Two potential participants declined screening, stating they were ineligible after looking at the eligibility criteria on the consent form. Eligibility screening was completed, and 10 potential participants were identified and accepted into the study. Of the 10 eligible and accepted screened participants, 8 completed the whole activity, and their experiences were recorded and used in the study. Braun and Clark (2021) prioritize depth, richness, and contextual insight over breadth, meaning a small, purposefully selected sample is methodologically appropriate, as the goal is to generate nuanced, interpretive understandings of participants' experiences rather than to achieve data saturation or generalizability. The sample of eight participants provided sufficient data to develop meaningful themes while allowing for close, reflexive engagement with each individual transcript.

To make the data sharing more meaningful, I have replaced the anonymized participant codes with pseudonyms. These names have been assigned to preserve confidentiality while making the participants' experiences more relatable and accessible to the reader.

Table 1 breaks down the participants who completed this research.

Table 1

Participants

Participant	Age in Years	Gender
Daniel	48	Male

Emily	35	Female
Sofia	36	Female
Grace	29	Female
Maya	44	Female
Macus	35	Male
Rachel	59	Female
Alex	36	Male

Participants screened n=10; the final recruitment was n=8, female n=5, male n=3. All recruited participants were between the ages of 19 and 60, and the number of participants recruited from the initial faculty email channel was n=2, with the remainder n=6 from the snowball method of word of mouth.

Qualitative Data Analysis Triangulated with the PETS

The inductive thematic analysis of post-activity interviews provided insight into how participants experienced ChatGPT's empathic responses and highlighted the nuances of their perceptions. When overlaid with the survey results, a fuller picture of the subjective experience of perceived empathy emerged. After all participant post-activity interviews were completed, an inductive thematic analysis was conducted following the guidelines set out by Braun and Clark (2006). The initial phase involved a first-pass reading of each post-activity interview transcript to familiarize the primary researcher with emerging patterns without predetermined categories. As subsequent transcripts were reviewed, common words and phrases that appeared frequently across participant responses were highlighted, common words were systematically compared across all transcripts and coded, with particular attention given to instances where participants directly quoted or paraphrased ChatGPT's responses. This approach allowed for tracking specific elements of ChatGPT's dialogue that participants perceived as empathic. The thematic analysis of the post-activity interviews produced three themes:

I. ChatGPT's Validation and Emotional Acknowledgment

II. ChatGPT's Structured Support and Problem-Solving Approach

III. ChatGPT's Communication Style, leading to Perceived Empathic Limitations.

Although individual reactions to ChatGPT varied, the participant narratives revealed consistent themes related to perceived empathy. The survey responses contextualized and enriched these findings by providing structured insight into participants' subjective experiences.

While Watson's Theory of Human Caring (2008) provided the conceptual foundation for the study, particularly in shaping the post-activity interview prompts, the themes presented in this findings chapter were developed through an inductive process using thematic analysis (Braun & Clarke, 2006, 2021). The coding and theme development emerged from the data itself, without applying Watson's theory as a coding framework. Once themes were established, the theory was revisited as an interpretive lens to enrich the understanding of how participants described empathy in their interactions with ChatGPT. Braun and Clarke's (2021) flexible approach to thematic analysis allows theoretical concepts to organically emerge through engagement with the data, rather than being imposed in advance.

A deliberate decision was made to refrain from conducting an analysis of ChatGPT's response transcripts to minimize potential bias and maintain focus on the participants' lived experiences rather than an external interpretation of ChatGPT's responses. This step ensured that the analysis remained centred on the perceived empathy of the participants and their lived experience rather than the primary researcher's impression or interpretation of ChatGPT's communication.

Table 3 illustrates the first theme, ChatGPT's ability to validate and acknowledge participants' emotions. Jean Watson's Theory of Human Caring emphasizes presence, empathy,

and holding space for a person's emotional reality (Cara,2003). Participants described ChatGPT's ability to validate their concerns and respond without judgment, aligning with Watson's Carative Factor of developing a helping-trusting, human caring relationship (Watson & Woodward, 2010). While ChatGPT lacked emotional depth, participants noted that it could acknowledge distress and respond in a way that felt emotionally attuned. This aligns with cognitive empathy and Watson's idea of recognizing and honoring human emotions, even if the response was not grounded in lived experience (Watson & Woodward, 2010).

Qualitative Data/Inductive Thematic Analysis

The qualitative analysis of post-activity interviews (Appendix G) provided deeper insight into how participants experienced ChatGPT's empathic responses and highlighted the nuances of their perceptions beyond numerical PETS scores.

While there is still variability in responses to and feelings towards ChatGPT, the qualitative data was generally rich and contextualized some of these differences. The thematic analysis of the post-activity interviews provided three themes: ChatGPT's Validation and Emotional Acknowledgment, ChatGPT's Structured Support and Problem-Solving Approach, and Challenges with ChatGPT's Communication Style, leading to Perceived Empathic Limitations. Table 3 illustrates the first theme, ChatGPT's ability to validate and acknowledge participants' emotions.

Theme #1:ChatGPT's Validation and Emotional Acknowledgment

Table 2

Thematic Analysis of AI Perceived Empathy Theme #1

Theme #1	Participant's Quotes

ChatGPT's Validation and Emotional Acknowledgment	"There was a lot of just acknowledging my feelings and saying like, it's OK to feel like this. These are normal feelings." (Sofia)
	"It kind of acknowledged that my plate is quite full and these are all normal things for me to be feeling right now." (Sofia)
	"It's like I'm crying on this computer shoulder and it knows it's wiping my tears away." (Sofia)
	"I felt encouraged and soothed" (Maya)
	"all that language that is very affirming and listening [to] it was almost eerie how good it felt to read it" (Emily)
	"anytime that I voiced concern anxiety, kind of like looking for that reassurance piece, I think that it definitely showed some emotional intelligence " (Grace)
	"It was shocking how realistic it was actually, and you know, it felt like I was talking to somebody " (Daniel)
	"Every time I said something, it acknowledged how I was feeling based on what I said and then gave feedback" (Daniel)
	"They are statements that I personally, being someone who works in psychiatry and understands empathy, feel they're statements that I would potentially utilize myself towards a client or a family member or friend if they were struggling." (Grace)

During the post-activity interviews, participants generally acknowledged ChatGPT's ability to validate emotions, describing its responses as affirming and reflective. This perception aligns with the survey, where the question "The system considered my mental state" received consistently high scores for all participants suggesting that ChatGPT acknowledged their emotions effectively and was able to understand they were experiencing a mental health challenge. One participant explained: "There was a lot of just acknowledging my feelings and saying like it is OK to feel like this." (Sofia). ChatGPT's ability to attune to the needs of the participants reflects the Watson's emphasis on creating a caring space where individuals feel

emotionally seen and heard (Watson & Woodward, 2010). Participants frequently referenced ChatGPT's ability to paraphrase emotions and offer nonjudgmental support, reinforcing the sense that ChatGPT was engaged and responsive. Many participants stated similar sentiments by saying "anytime that I voiced concern anxiety, kind of like looking for that reassurance piece, I think that it definitely showed some emotional intelligence " (Grace) or "Every time I said something, it acknowledged how I was feeling based on what I said and then gave feedback" (Daniel)

Similarly, the survey item "The system seemed emotionally intelligent" received scores grouped around strongly agree, indicating that participants felt that ChatGPT's responses made participants feel heard and supported.

Section 3 of the semi-structured post-activity interview was designed to illuminate how participants felt ChatGPT measured up to the empathy a human might offer if the participant went to a friend or professional for mental health support (Appendix G, Section 3-Q7,8). Participat Maya shared that they felt the AI could stay empathic and not slip into sympathy, that it didn't tap into the hopelessness that can be present, which real people often do when supporting someone through a challenging time (Maya). Participant Rachel also felt that ChatGPT didn't try to tell them to feel a different way, which sometimes happens when interacting with others, and that ChatGPT was good at staying focused on the participant's feelings (Rachel). Participant Alex also observed that the empathy shown by ChatGPT was better than what a human might respond with because the ChatGPT knows what to say, and that might not be the case with a friend or a colleague if you were to bring up something emotionally sensitive (Alex); this sentiment aligned with survey item 8 "The system understood my needs"

where the majority of participants rated this item near the “Strongly Agree” end of the scale, with only two responses falling in the midrange, suggesting broad agreement with the statement.

Section 5 of the post-activity interview asks participants to reflect on the emotional impact and support they felt when interacting with ChatGPT, leading two participants to share that choices about what to do to feel better (Emily) and the way ChatGPT picked up on the importance of choice made them feel supported. Participant Maya also stated that the interaction with ChatGPT was emotionally soothing and that the responses that they received from ChatGPT allowed them to regulate their emotions, sharing that the interaction started out emotionally at an 8 out of 10, but after a few exchanges with ChatGPT, they felt encouraged and soothed (Maya). Although ChatGPT is not human, participants’ experiences during their interactions suggest that it is able to simulate aspects of caring that closely mirror many of the features central to Watson’s theory.

The general consensus in the follow-up interviews was that through validation and acknowledgement of feelings, participants felt the interaction with ChatGPT exceeded their expectations; one participant commented, “It was shocking how realistic it was actually, and you know, it felt like I was talking to somebody” (Daniel). ChatGPT was adept at responding to the situation being presented, and by addressing participants with validating responses, they felt supported.

However, the survey item “The system supported me in coping with my emotional situation” had one of the lowest scores suggesting that emotional responsiveness was not equal to support with coping. While the majority of participants (six out of eight) rated the items “The system showed interest in me” and “The system sympathized with me” in the top quarter of the scale, indicating a strong perception of empathy, two participants rated these items in the lower

quarter. This variation highlights the subjective nature of perceived empathy and reinforces the importance of individualized care, as emphasized in Watson's Theory of Human Caring. Watson (Watson & Woodward, 2010) recognizes that each person's experience must be honored as unique, and that perceptions of trust, emotional resonance, and rapport may differ based on personal context and meaning. These lower ratings align with the subjective experiences shared in the post-activity interviews from participants Alex and Marcus, who each expressed that the interaction with ChatGPT felt less impactful than an interaction with a human therapist.

The second theme to emerge from the analysis of the participants accounts was ChatGPT's Structured Support and Problem-Solving Approach. Participants felt that ChatGPT provided structured support and robust problem-solving plans. Table 3 illustrates how participants felt supported and how having this written plan was helpful to them. Watson's framework includes the importance of systematic use of scientific problem-solving for decision making (Watson & Woodward, 2010), which can be applied to ChatGPT's logical, clear, and solution-focused responses. Participants appreciated the structured nature of the support, which echoed Watson's commitment to competent, informed care (Watson & Woodward, 2010). While ChatGPT may encounter challenges to embody emotional empathy or a caring presence, its practical guidance reflects aspects of cognitive caring, a rational form of support that helps users feel heard and guided, albeit in a non-human way.

Theme #2: ChatGPT's Structured Support and Problem-Solving Approach

Table 3

Thematic Analysis of AI Perceived Empathy Theme #2

ChatGPT's Structured Support and Problem-Solving Approach	Participant's Quotes

> "You know, streamlined along the conversations of what I would have with my my counselor" (Sofia)
>
> "I liked how it broke down the plan for you, like#1Reframing 2 talk to your administrator, reconnect with your husband" (Sofia)
>
> "So having, you know, essentially a written out script for you to refer to is I think would be incredibly helpful" (Sofia)
>
> "I, you know, like I just like because of stress or emotion overload. I feel a bit scrambled. And so one of the things I really liked about this more so than having like a conversation with a person like back and forth, back and forth is actully seeing things. Laid out." (Maya)
>
> "Just let me break it down in the stealthiest options that are both effective and subtle, and then went through the options again and then showed like, hey, these ones are completely invisible. Section 2 was subtle and invisible. Section 3 is like look like work like behavior and then Section 4 was discrete breaks that seem normal. So I mean, I think that it really did kind of like that's just an example of how AI was able to kind of. Break apart options that could be utilized." (Grace)
>
> "I thought the plan was really good, like it was very simple, very easy to follow. And it was like, so it could, it could be coached down to like a more reasonable. Solution and like very actionable plan. So I thought that was cool." (Emily)

Section 1 of the post-activity interview questions asked about participants' general impressions of their interaction with ChatGPT; several participants noted that ChatGPT's responses closely resembled human counselling techniques, particularly in its structured and solution-focused approach. One participant stated: "You know, streamlined along the conversations of what I would have with my my counselor" (Sofia), going on to reflect that ChatGPT came off as a really educated counsellor (Sofia). Many participants mentioned there was a lot of value in having a detailed action plan that could be printed out and looked at after the initial session concluded (Sofia, Maya, Emily). This perception aligns with the survey results, where the item "The system understood my needs" again saw six of the eight participants rate

ChatGPT in the top quarter of the scale suggesting that participants generally felt ChatGPT provided relevant and structured responses tailored to their concerns. The participants also appreciated that when they responded to ChatGPT's robust list of options with comments that showed they were feeling overwhelmed, ChatGPT was immediately able to regenerate content that broke the steps down into smaller and smaller attainable goals (Emily, Sofia). ChatGPT's ability to pick up on the overwhelm and smoothly shift gears was a strength and seemed to contribute to the participants' overall feelings of being heard and validated. This was mirrored in the survey where the item "The system understood my goals" had the majority of the ratings land in the top one third of the scale. One of the post activity interview questions specifically asked participants, "Did the AI's responses make you feel like it was trying to understand your situation?" (Appendix G, Section 2-Q4), leading many of the participants to state that they felt the ChatGPT was trying to understand them as well as uncover things about them or their situation, and give them options that were practical and reasonable (Emily, Grace, Rachel). Responses to the survey item "The system showed interest in me" echoed these experiences, as most participants gave high ratings, suggesting that ChatGPT was perceived as attentive and engaged. Overall, the response to ChatGPT's structured problem-solving ability impressed participants; many especially appreciated that the content was organized and succinct. As summarized in Table 3, participants described ChatGPT as being very astute at making structured plans that felt supportive and were easy to follow, with statements such as "And so one of the things I really liked about this more so than having like a conversation with a person like back and forth, back and forth is actually seeing things. Laid out." (Maya). The majority (six out of eight) of the participants noted that ChatGPT was able to offer personalized guidance while also recognizing when a user might be feeling overwhelmed. Its ability to break down

suggestions into manageable steps and present multiple options gave participants a sense of agency in choosing what felt right for them. This approach aligns with Watson's theory, which emphasizes the importance of promoting patient autonomy as a core aspect of the caring process (Savieto & Leão, 2016).

The final theme was Challenges with ChatGPT's Communication Style, leading to Perceived Empathic Limitations. Table 4 and the following section explore how these responses influenced perceptions of empathy and the limitations that participants identified.

Watson's emphasis on caring presence and authentic connection stands in contrast to the fast, polished, and sometimes mechanical responses of ChatGPT, which several participants found unsettling. The absence of reflective pause, personalized pacing, and shared emotional depth contributed to lower survey item scores in areas like trust and emotional support. This misalignment highlights where ChatGPT's form of empathy, primarily cognitive, falls short of Watson's model, which values emotional engagement and transpersonal connection (Watson & Woodward, 2010). The perceived lack of genuine caring presence was a key reason participants found the interaction less empathic.

Theme #3: Perceived Empathic Limitations & Challenges with ChatGPT's Communication Style

Table 4

Thematic Analysis of AI Perceived Empathy Theme #3

Challenges with ChatGPT's Communication Style, leading to Perceived Empathic Limitations	Participant's Quotes
	"I thought it was really good, really good, empathetic. It was almost too good...I feel like humans are more fallible"(Emily)

“That next level of like actual empathy, empathy versus just like sympathy, which is the the. Sharing of lived experience” (Emily)

“The only thing I think kind of spoke to not enough empathy maybe was just kind of the sheer volume of the solutions or like offerings kind of thing. Like it got so excited, it was just, you know, vomiting all these great solutions” (Emily)

“So I think for me personally seeing. Somebody on the other side? Would be nice.” (Daniel)

“I was 'cause it it it came back with the best **** ever, but too quickly. Like I wanted it to work.” (Maya)

“There is no body language” (Maya)

“like a lot of good information and it’s generated really quickly, but at the same time...its’ just not genuine. It doesn’t feel it” (Marcus)

“but no matter what it says,like,even if it says all the right things, and it felt as you know, saying caring sentences and saying, you know life is worth living and what not, you’re still talking to a robot.” (Marcus)

“I mean, the initial reaction was it felt alright 'cause. It was like it's. It felt like it was actually concern for your safety a little bit and like your well-being. But then it just started spitting out. You saw it like it was just a mass amount of information. So the more you read through it, the less. Impactful it is.”(Marcus)

“I guess from a written point of view, yes and and maybe that was the piece that was missing for me is the voice I think would have felt more therapeutic as opposed to me having to just read their their responses” (Rachel)

“I think the AI was very it was textbook appropriate” (Rachel)

“I guess I sort of somewhat believe that people crave human connection more than they crave the answers. When the problem is not a technical problem. It is an emotional problem and I. You know? Sure. Reading about some helpful techniques might get you through the next. Five hours,” (Alex)

> “I think these issues are about feeling. Uncared or or misunderstood or like they're human problem. They're not a technical problem. And so a technical answer won't work. And as good as ChatGPT is, it's not a human answer.” (Alex)

Despite positive feedback regarding validation and structured support, many participants noted challenges with ChatGPT’s communication style, particularly with the volume and speed of responses, feeling this detracted from empathy and disrupted the conversational flow, making interactions feel unnatural. One participant stated: “I mean, the initial reaction was it felt alright 'cause. It was like it's. It felt like it was actually concern for your safety a little bit and like your well-being. But then it just started spitting out. You saw it like it was just a mass amount of information. So the more you read through it, the less. Impactful it is.” (Marcus). The sentiment was shared by many of the participants, even those who were impressed with ChatGPT’s ability to portray empathy because the interaction was too good, too fast and lacked the elements of human fallibility (Emily, Sofia, Maya). This perception is reflected in the survey item, “The system supported me in coping with my emotional situation.” where most participants rated this item in the highest range of the scale, indicating a strong sense of support in coping with their emotional situation, two participants rated it notably lower. This variation suggests that while many found ChatGPT’s responses helpful and empowering, a few felt that the support lacked the depth or nuance they needed to feel emotionally connected to feel ChatGPT was displaying empathy. These perspectives reflect Watson’s emphasis on the importance of individualized care, where emotional needs and perceptions vary across people and moments (Watson & Woodward, 2010). So while some participants felt that ChatGPT’s guidance was helpful and structured, others felt that its mechanical precision and rapid delivery of information detracted from the emotional depth of the interaction.

Several participants noted that although the ChatGPT used validating and paraphrasing techniques, the speed and polish of the responses made the experience feel somewhat superficial rather than deeply empathic. Several participants also described ChatGPT's responses as feeling mechanical rather than organic, reinforcing the idea that there is a difference between cognitive empathy and emotional empathy. One participant stated, "I think the AI was very...it was textbook appropriate" (Rachel). Another challenge discussed in the post-activity interviews was the lack of connection when interacting with ChatGPT through text only. As shown in Table 4, participants noted the absence of body language, vocal tone, and the relational presence created through silence and nonverbal cues. While this limitation might be expected to influence responses to the survey item "The system expressed emotions," most participants still rated this item highly, with only one response falling in the lower quarter of the scale. When enhanced with browsing tools, ChatGPT can access and summarize peer-reviewed literature and current sources much like a search engine. However, at its core, ChatGPT functions as a large language model, generating responses based on statistical patterns from its training data rather than retrieving content from live databases. ChatGPT's ability to tap into the vast data bank of information at its disposal allowed it to instantly come up with solutions as well as perfectly reflect back what participants shared. However, this ability did not always lend itself to portraying all aspects of empathy. This sentiment led one participant to comment that "no matter what it says, like, even if it says all the right things, and it felt as you know, saying caring sentences and saying, you know life is worth living and what not, you're still talking to a robot." (Marcus). Another participant explained that ChatGPT seemed to have a good grasp on the technical side of the presented mental health challenge but that people are not a technical problem to be solved; they are human, so technical answers may not be that effective at

conveying empathy to users; humans may crave human connection more than they crave efficient answers (Alex, Male 36). This highlights the tension between efficiency and perceived empathy, while ChatGPT's structured responses demonstrated cognitive understanding, its lack of pauses, imperfections, and perceived lack of emotional reciprocity led some participants to question its depth of engagement. Participants shared that while ChatGPT initially felt caring, prolonged interactions reduced their emotional connection to it (Emily, Marcus, Rachel). Watson's Theory of Human Caring supports the importance of relational depth and warns against seeing people only as tasks, the caritive process is the foundation of empathy, emphasizing the need to "be" not just "do" in order to show up with authentic presence (Watson & Woodward, 2010).

The final key factor that emerged from the reflections shared in the interviews was that perceived empathy appears to be influenced by trust, with one participant going so far as to say outright "I do not trust ChatGPT to be able to talk me down." (Alex). This did track with participant responses to the survey item "I trusted the system" which showed notable variation and with ratings scattered across the full range of the scale. While several participants rated trust in ChatGPT quite highly, others remained skeptical or uncertain, and a few expressed clear discomfort. Interestingly, some participants acknowledged that ChatGPT did a good job of reflecting their thoughts or feelings but still did not feel they could trust the system, primarily because it was not a real person (Marcus). This distinction speaks to Watson's assertion that authentic presence is foundational to developing trust in a caring relationship (Watson & Woodward, 2010). For these participants, the knowledge that they were interacting with an artificial system created a barrier to relational trust, regardless of the quality of the language or

empathic tone. This suggests that while ChatGPT may simulate emotional understanding, some individuals still experience an underlying lack of connection, which ultimately affects trust.

Summary

Chapter 4 is an in-depth look at the reflections shared in the post-activity interviews and shows how the interaction with ChatGPT is seen when viewed against the backdrop of perceived empathy. The participant experience was rich and informative, shedding light on how participants viewed their interactions with ChatGPT and how those interactions influenced their individual perceptions of empathy. The survey helped to capture perceived empathy on a continuum and illustrated that each participant perceived empathy through their own experience, leaving variable data points along the way. The chapter also illustrates the relationship between the post-activity interviews and the survey was showcased to triangulate the subjective experience of empathy participants had when engaging in the research activity and helped contextualize the participants' experience when viewed through the lens of Jean Watson's Therory of Human Caring.

Chapter 5: Discussion

The use of AI for mental health apps is not a new concept. Researchers and developers have long been exploring the idea of robots and AI systems that can support individuals with mental health needs (Abd-alrazaq et al., 2019). In this final chapter, the discussion will centre around the overall findings of the research, the implications for practice, the limitations of this research and where the research could go in the future.

Findings

The findings of this research suggest that while ChatGPT effectively demonstrated cognitive empathy through validation, reflective listening, and structured problem-solving

approaches, it encountered challenges in conveying emotional and behavioral empathy in ways that align with Jean Watson's Theory of Human Caring. Participant feedback from the post-activity interviews highlighted ChatGPT's ability to mirror emotions or offer reassurance and high survey scores in areas such as emotional acknowledgment and needs recognition supported these participant reflections. However, limitations emerged from ChatGPT's mechanical delivery, rapid response time, and lack of human presence, contributing to variability in trust and perceived emotional support. Although participants appreciated ChatGPT's textbook-level empathy and structured responses, the absence of lived experience and human imperfection limited deeper emotional connection.

In addition to the three core themes discussed above, several other noteworthy observations emerged during the analysis. While these did not occur with enough consistency to be considered full themes, they offer valuable insights that enrich the understanding of participants' experiences. One interesting observation regarding participant demographics was that more females than males volunteered for the study, and male participants consistently reported lower survey scores and expressed discomfort with the ChatGPT, referring to it as a "robot" and describing interactions as inauthentic. In contrast, female participants were generally more accepting of ChatGPT's empathic mimicry. Additionally, a large portion of participants were actively employed in healthcare, with some explicitly noting that ChatGPT's responses mirrored the therapeutic communication frameworks taught in healthcare education. Surprisingly, concerns around data privacy were minimal, with only two participants expressing apprehension about using personal information. Age-related observations also surfaced: while older participants speculated that younger generations might more readily accept AI-based mental health support, it was, in fact, the younger participants who expressed greater skepticism

about data security and were more cautious in their disclosures, felt ChatGPT lacked authenticity, and one younger respondent stated they would never turn to artificial intelligence of any kind for a mental health crisis (Alex, male 36). All participants had a generally high level of digital literacy, reliable access to the internet and owned their own electronic devices. The overall outcomes of this research might be quite different if the age range were narrower or if the population were more marginalised. The potential for this technology to be utilized in rural areas will be limited by the age, digital literacy, and access to technology that typically do not present as issues in more urbanized areas. Generational differences in technology use and help-seeking behaviors are well-documented. For example, Woerner et al. (2022) found that younger adults are more likely to readily embrace digital mental health tools, whereas older cohorts prefer these tools as a complement to traditional therapy. Similarly, Papp-Zipernovszky et al (2021) found greater eHealth literacy and online health information engagement among Millennials and Gen Z than older generations, correlating with higher self-empowerment in managing care (Van Deursen et al., 2021).

This research focused on a narrow question: Can generative AI (specifically ChatGPT) be perceived as empathic in ways that resemble human empathy in its therapeutic communication? This study suggests that ChatGPT is adept at demonstrating cognitive empathy. While it does not "feel" empathy, it effectively simulates human empathy through therapeutic communication techniques, such as reflective listening and validation. Participants largely reported positive experiences interacting with ChatGPT, indicating that the system was able to meet their needs for feeling heard and understood even when they would go on to contradict themselves by saying the response rate of ChatGPT detracted from their perception of empathy.

Implications for Practice

The implications for practice are significant. Recent data highlights a significant rise in mood and anxiety disorders among Canadians, with many reporting unmet needs in mental healthcare, particularly for counselling services (Stephenson, 2023). Long wait times, workforce shortages, and high costs (Boucher et al., 2021) all contribute to barriers to accessing care.

Mental Health Consumers

Prolonged delays in treatment can worsen mental health conditions (Kretzschmar et al., 2019), highlighting the urgent need for alternative solutions. AI-driven mental health support could offer timely, accessible, and cost-effective support, especially in rural or underserved areas. Using tools like ChatGPT could substantially augment existing mental health services by offering support to mental health consumers who fall into the mild to moderate category of need for care. The validation and skills that ChatGPT can provide are powerful and are delivered easily and quickly to anyone with access to the internet and an electronic device. This could mean that mental health care consumers would not have to wait on lists to access meaningful resources to support their mental health, and would make access to skills and psychoeducational information essentially for free. Public mental healthcare services are often constrained by funding and prioritize individuals with severe and persistent mental illness, making it difficult for those with mild to moderate concerns to access care. This leaves a gap in support, where individuals who do not meet the criteria for public services must either seek costly private therapy or go without support altogether. AI-driven mental health tools like ChatGPT offer an accessible alternative that could help address these gaps (Elyoseph et al., 2024). Unlike traditional services, AI-based support is available 24/7, without waitlists, and at little or no cost. While these tools may not yet be a complete replacement for professional mental health treatment (Lai et al., 2024), they can serve individuals who may otherwise struggle to access

support by reducing barriers related to cost and availability. However, there is a continued need to improve AI systems because empathy is a cornerstone of the therapeutic relationship (Watson, 2008) and this study has shown that the perception of empathy felt by consumers is relevant. Perception of empathy affects user trust, satisfaction and the likelihood that a consumer will continue to engage in using the AI system. This research also found that there is a lack of emotional resonance that comes from ChatGPT's ability to deliver perfect, textbook responses in a split second which limits the perception of empathy for users. In order for systems like ChatGPT to be the first line of defence in fighting mental illness, the empathy both real and perceived is critical to user engagement and impacts the therapeutic value of any interaction. While AI systems like ChatGPT are advancing in their ability to mimic human empathy caution may need to be taken as this raises ethical concerns about user mistaking ChatGPT's replicated human empathy for emotional authenticity and potentially delaying connection to person-to-person professional help. The responses from ChatGPT can feel very real, with many of the participants in this study reflecting on their experience and stating that they couldn't believe how real the interaction felt or how good the advice was that they received (Maya, Daniel, Emily, Sofia). However, this replication of cognitive empathy is not genuine and might put users at risk because healthy therapeutic relationships are built on more than just cognitive empathy; they also require emotional, motivational or behavioural empathy in order to support a person holistically (Rubin et al., 2024; Watson & Woodward, 2008). Since ChatGPT lacks the capacity to express all dimensions of empathy, this limitation could heighten risk, especially in users with a higher level of mental dysfunction. Watson and Woodward (2008), emphasis on holistic care underscores the necessity to ensure all aspect of empathy are addressed in therapeutic interactions for truly effective mental health care. Another major ethical concern is the potential

for misinformation in AI-generated content. While ChatGPT puts a small disclaimer at the bottom of the screen stating that it is important to check information provided because ChatGPT can confabulate content, how many people will do this especially if they are interacting with the system because they are needing mental health support. And even if they do wonder if the information they are receiving is incorrect, if the majority of users were to rate ChatGPT's trustworthiness in the same way as the participants in this study (the majority rated trust of the system in the top third of the scale) would they stop and critically assess the responses they received? It's crucial to provide transparent, accessible disclaimers and educate users on the limitations of AI. Watson's Theory of Human Caring (Watson, 2008) emphasizes trust and transparency as foundational to a caring relationship, aligning with the need for clear communication about AI's capabilities.

Given these limitations it is crucial to have accountability and safety in AI driven mental health support, so who is going to be responsible to ensure these AI driven tools are safe and effective for mental health care? Cuadra et al., (2024) purports that the ethical concerns raised due to potential inconsistencies calls for ethical safeguards to prevent the manipulation of users' emotions and that more research into human-computer interactions must be done in order to develop responsible practices that protect users from the potential of the exploitative nature of "projected empathy". This raises an important and frightening point, that there is an inherent risk to users when they become unable to detect the difference between machine empathy and human empathy. Still, who will ultimately be held accountable for mitigating these ethical concerns? Boucher et al. (2021) concluded that while AI chatbots hold significant promise for enhancing digital mental health care, their success will depend on addressing current limitations, such as improving empathy, reducing biases, and conducting more rigorous evaluations of their

effectiveness, and the rapid pace and ever-changing landscape of AI development means no longitudinal data has yet been gathered. Meaning it could be years before we actually know if using this technology for mental health support is safe.

Healthcare System

The findings of this study suggest that AI-based mental health tools like ChatGPT could play a meaningful role in supplementing Canada's overburdened mental healthcare system.

AI could help alleviate some of the demands placed on an overstretched mental health workforce, ensuring that mental health professionals can focus on individuals with more complex clinical needs because those with less urgent care needs would be able to access timely supports on their own. This could redistribute resources by offloading the low-risk mental health consumers that are clogging up referral and access pathways, which could potentially make access more efficient with potentially much shorter wait times because people would be able to access free early interventions, making it less likely they would need to access the complex interventions required for severe and persistent mental health care. These changes in access and flow have potential cost efficiencies, particularly in early intervention and health promotion; if mild to moderate mental health consumers find viable interventions before they become severely compromised there could potentially be fewer consumers attempting to access limited resources. However, if mental health consumers felt that person to person care was being denied in favour of lower cost virtual care the trust in the healthcare system could be compromised. Healthcare systems must consider whether or not generative AI such as ChatGPT enhance or erode relational care. They may also need to look at innovative solutions where a hybrid model uses ChatGPT for the cognitive empathy tasks such as reflective listening, validating and offering detailed step by step treatment plans, and a human clinician's resources are preserved for tasks

that involve deeper levels of empathy or are necessary to support mental health consumers that cannot be provided virtually. This may mean that in the future health care systems will have to take on the responsibility of ensuring that all their mental health providers also are educated on AI chatbots' strengths and limitations so that they can effectively use them as complementary tools within their own personal practice (Belcombe, 2023).

Healthcare Providers

For healthcare providers the adjunct of AI driven interventions such as ChatGPT could offer supplemental support for mental health consumers by monitoring symptoms and providing psychoeducation. The participants in this study were impressed with ChatGPT's ability to understand their concerns and provide them with detailed interventions (Maya, Emily, Daniel); if healthcare providers had an "AI co-therapist" that was accessible on a continuous basis the potential to increase the capacity of the healthcare provider would be significant. It could allow healthcare providers to extend support beyond session times, reinforce therapeutic strategies, and provide timely responses to client needs without increasing workloads. In their study published in Nature Machine Intelligence (Sharma et al., 2023), Sharma and colleagues explore how artificial intelligence can be meaningfully integrated into online peer-to-peer mental health platforms to enhance empathic communication..However, the integration of AI systems, apps, or large language models into mental health care carries inherent risks, both for health care providers and, more importantly, for the consumers who trust professionals to support and treat them. In an article dated June 29, 2025, human rights lawyer and AI expert Susie Alegre raises serious concerns about the surge of AI chatbots being used in mental health care, particularly as a substitute for human interaction. Alegre (2025) highlights chilling real-world cases. One lawsuit involves a teenager's suicide after using a chatbot from Character.ai, while another

involves an adolescent who was encouraged to threaten his parents, both featuring chatbots marketed as therapists or confidantes. She emphasizes that such platforms can dangerously reinforce harmful thought patterns in vulnerable youth, especially those lacking real social connections. While some AI tools provide nonjudgmental space for users, she warns that effective therapy requires more than affirmation; it necessitates judgment, challenge, and accountability to foster genuine healing. Watson's Theory of Human Caring (Watson, 2008) emphasizes authentic, empathetic human relationships in healing. Her concept of the transpersonal caring relationship calls for deep human-to-human connection that acknowledges the whole person,mind, body, and spirit. When AI is introduced into this relational space, particularly as an adjunct to face-to-face services, there is a risk of disrupting the emotional and ethical integrity of the care experience. If generative AI, such as ChatGPT, becomes more commonly used alongside face-to-face services, the potential for harm increases if the AI fails to detect critical safety cues. Such lapses could result in missed opportunities for early intervention, crisis management, or timely escalation to human support. Becombe (2023) states that researchers, developers, and regulators must collaborate to establish standards of care that ensure ethical and safe services, and these standards need to be developed quickly as there is a breakneck speed to the changing landscape of generative AI systems so we need to address these ethical and safety concerns sooner rather than later. In the research done by Kretzschmar et al., (2019) he asked youth for input into their perspectives on using AI for mental health support; their recommendations included safeguarding data privacy, clearly communicating the chatbot's limitations, encouraging users to seek human support when needed, and ensuring the chatbot can respond appropriately to emergencies, such as suicidal ideation, by directing users to relevant resources or services. Similarly, the research done by Solaiman et al., (2024) highlights the

urgent need for clear guidelines to protect mental health patients from potential harms posed by generative AI systems, which can deliver unsupervised responses without clinical training, emphasizing the unique risks of publicly accessible, unregulated tools. These concerns further underscore the need for strong guardrails, as mental health professionals cannot reasonably shift accountability onto an AI system if harm occurs under their care. Even in an overburdened and eroded system, the adoption of hybrid AI services may be a difficult sell for clinicians who are ultimately responsible for patient outcomes.

Summary

While AI tools like ChatGPT offer promising opportunities to reduce barriers and expand access to mental health support, particularly for those with mild to moderate needs, their use must be approached with caution. The perceived empathy delivered by such systems plays a critical role in user engagement, but without genuine emotional depth, there is a risk of users overestimating the AI's capacity to provide care. As Watson's Theory of Human Caring (2008) reminds us, trust, transparency, and holistic empathy are essential in therapeutic relationships. Without strong safeguards, clear communication of limitations, and ongoing accountability, AI systems risk reinforcing inequities rather than resolving them, especially if users mistake simulated empathy for meaningful connection. For AI to be a truly safe and ethical adjunct in mental health care, it must be developed and deployed with a commitment to protecting the dignity, safety, and humanity of those it aims to serve.

Limitations

This research is subject to several limitations that should be considered when interpreting the findings. These include constraints related to the study's scope and methodology, as well as limitations inherent in the technology used, the theoretical framework applied, and the

generalizability of the results. One key limitation is the study's small sample size, which, while generating rich qualitative data, may have limited the emergence of broader themes or perspectives. A larger sample could have introduced more demographic variability and provided deeper insight into how factors such as age, gender, mental health history, current well-being, or digital literacy influence perceptions of AI empathy. Also, within the context of the sample size, the sample was primarily composed of healthcare providers and academics; as such, this may not generalize to the greater population, as their perceptions of empathy might differ greatly from those of a general sample of the population.

Additionally, the study was designed as a one-time interaction with ChatGPT and included only participants screened to ensure they did not have an active mental health challenge. This narrowed the scope to healthy individuals and focused solely on perceived empathy, rather than on broader outcomes. Future research, using a larger and more diverse sample and repeated interactions, could offer a more comprehensive understanding of AI's potential role in mental healthcare. Also, due to setting the inclusion criteria to only those without a mental illness and having up to a one-week time-lapse between the interactive activity and the follow-up interview, there may have been certain aspects of the interaction that were diminished, since the participants were not describing their personal mental health challenges but playing a role.

Methodologically, this qualitatively driven study had participants completing the Perceived Empathy of Technology Scale (PETS); The data was not used for statistical inference. In future studies, a sufficiently large sample could allow for meaningful quantitative analysis, enabling broader generalizations about users' experiences. Furthermore, while PETS is a

promising tool for assessing empathy in human to AI interaction, it remains relatively new, and current literature on its application is limited.

There were also limitations in applying Watson's Theory of Human Caring to a non-human agent. While the theory provides a strong ethical and relational foundation, AI lacks the capacity for deep emotional connection or holistic empathy, making it difficult to fully align AI interactions with Watson's core principles. Additionally, this study used simulated mental health scenarios to protect participants, which may not accurately reflect how individuals with real mental health challenges would perceive empathy from an AI system. Participants also varied in their previous experience with ChatGPT, introducing potential bias. Some used the free version, while others used the paid version of the model, which could have led to subtle differences in interaction quality. However, since the focus of the study was on perceived empathy during the interaction, these variations were not considered central to the findings.

Recommendations for Future Research

Larger sample sizes with diverse populations would definitely enhance the credibility, transferability, dependability, and confirmability of this work. The current study intentionally left the gender of the simulated user neutral to keep the playing field as level as possible and not muddy the waters with other considerations, so specifically, looking into potential gender bias within the AI system would be an interesting avenue to explore. Setting up research that specifically assesses whether responses differ based on the perceived gender of the user and how this may impact the quality or perception of empathy during the interaction. Testing with individuals experiencing real-time distress could also be a valuable focus for future research. The current research had a narrow focus on just the perception of empathy that the participants

ascribed to ChatGPT, but testing not only perceived empathy but also efficacy with real world mental health challenges would give a more robust picture of how generative AI systems could be used to support the struggling mental health field. Future research that also compares generative AI systems to human counselors would be a way to increase the adaptive knowledge of the generative AI. Comparison research between AI responses and human counsellor responses could support better interactions between humans and AI and make the support that people receive from AI systems more effective. Designing innovative research that takes on the challenge of exploration of AI's ability to show affective empathy or deeper therapeutic alliance would also be informative and could pave the way to exciting new collaborations between healthcare and computer science.

Conclusion

This study employed a qualitative descriptive design to explore the perception of empathy in users interacting with ChatGPT for mental health support. The participants subjective experience was investigated using inductive thematic analysis and the study design was guided by Jean Watson's Theory or Human Caring. This research demonstrates that ChatGPT has potential as an empathy-based support tool for individuals seeking mental health assistance. While it does not replace human care, its ability to demonstrate cognitive empathy and therapeutic communication skills, coupled with its accessibility, suggests it could play a meaningful role in addressing current gaps in mental health services in Canada.

In the rapidly evolving field of generative AI there is still much research to be done to ensure that AI systems like ChatGPT are a safe, reliable, and ethical alternative to person-to-

person care. While there is no doubt that this technology is here to stay, there is still work to be done to integrate it into the mainstream healthcare system.

References

Abd-alrazaq, A. A., Alajlani, M., Alalwan, 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*, 103978. Retrieved December 29, 2023, from https://doi.org/10.1016/j.ijmedinf.2019.103978

Alegre, S. (2025, June 29). AI therapy is no replacement for real judgment, says expert. The Times. (2025). https://apple.news/APlYiAf5kTDe-boRPiunB5w

Alegre, S. (2025, June 29). Need for regulation is urgent as AI chatbots are being rolled out to support mental health. *Toronto Star*.

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods*, *18*. https://doi.org/10.1177/1609406919874596

Austin, W., & Boyd, M. A. (2015). *Psychiatric & mental health nursing for Canadian practice* (3rd ed.). Wolters Kluwer.

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, *183*(6), 589. https://doi.org/10.1001/jamainternmed.2023.1838

Balcombe, L. (2023). Ai chatbots in digital mental health. *Informatics*, *10*(4), 82. Retrieved December 31, 2023, from https://doi.org/10.3390/informatics10040082

Banerjee, S., Dunn, P., Conard, S., & Ali, A. (2024). Mental health applications of generative ai and large language modeling in the united states. *International Journal of Environmental Research and Public Health*, *21*(7), 910. https://doi.org/10.3390/ijerph21070910

Bardolf, D. (2025, June 28). Harmful AI therapy: Chatbots endanger users with suicidal thoughts, delusions, researchers warn. *New York Post*.

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, *18*(sup1), 37–49. Retrieved December 31, 2023, from https://doi.org/10.1080/17434440.2021.2013200

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2020). One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, *18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv*, *5*(12712).

Bulmer, M. (1982). The research ethics of pseudo—patient studies: A new look at the merits of covert ethnographic methods. *The Sociological Review*, *30*(4), 627–646. https://doi.org/10.1111/j.1467-954x.1982.tb00671.x

Cara, C. (2003). A pragmatic view of jean watson's caring theory. *International Journal of Human Caring*, *7*(3), 51–62. https://doi.org/10.20467/1091-5710.7.3.51

Cheng, Y., & Jiang, H. (2020). Ai-powered mental health chatbots: Examining users' motivations, active communicative action and engagement after mass-shooting disasters. *Journal of Contingencies and Crisis Management*, *28*(3), 339–354. Retrieved December 31, 2023, from https://doi.org/10.1111/1468-5973.12319

Chin, H., Song, H., Baek, G., Shin, M., Jung, C., Cha, M., Choi, J., & Cha, C. (2023). The potential of chatbots for emotional support and promoting mental well-being in different cultures: Mixed methods study. *Journal of Medical Internet Research*, *25*, e51712. Retrieved January 13, 2024, from https://doi.org/10.2196/51712

Concannon, S., & Tomalin, M. (2023). Measuring perceived empathy in dialogue systems. *Ai & Society*, *1*(15). Retrieved April 14, 2024, from https://link.springer.com/article/10.1007/s00146-023-01715-z

Craig, T., Rus-Calafell, M., Ward, T., Leff, J. P., Huckvale, M., Howarth, E., Emsley, R., & Garety, P. A. (2018). Avatar therapy for auditory verbal hallucinations in people with psychosis: A single-blind, randomised controlled trial. *The Lancet Psychiatry*, *5*(1), 31–40. Retrieved January 21, 2024, from https://doi.org/10.1016/s2215-0366(17)30427-3

Cuadra, A., Wang, M., Stein, L., Jung, M., Dell, N., Estrin, D., & Landay, J. (2024). The illusion of empathy? notes on displays of emotion in human-computer interaction. In. *In Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18.

Elyoseph, Z., Gur, T., Haber, Y., Simon, T., Angert, T., Navon, Y., Tal, A., & Asman, O. (2024). An ethical perspective on the democratization of mental health with generative ai. *JMIR Mental Health*, *11*, e58011–e58011. https://doi.org/10.2196/58011

Esmail, N. (2023, October 10). *Mental health care: How is canada doing?* Fraser Institute. https://www.fraserinstitute.org/studies/mental-health-care-how-is-canada-doing

Giray, L. (2024). Cases of using chatgpt as a mental health and psychological support tool. *Journal of Consumer Health on the Internet*, 1–20. https://doi.org/10.1080/15398285.2024.2442374

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Klassen, A. C., Creswell, J., Plano Clark, V. L., Smith, K., & Meissner, H. I. (2012). Best practices in mixed methods for quality of life research. *Quality of Life Research*, *21*(3), 377–380. https://doi.org/10.1007/s11136-012-0122-x

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. Retrieved December 31, 2023, from https://doi.org/10.1177/1178222619829083

Lai, T., Shi, Y., Du, Z., Wu, J., Fu, K., Dou, Y., & Wang, Z. (2023). Supporting the demand on mental health services with ai-based conversational large language models (llms). *BioMedInformatics*, *4*(1), 8–33. Retrieved December 29, 2023, from https://doi.org/10.3390/biomedinformatics4010002

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1). https://doi.org/10.1525/collabra.33267

Langwieler, G., & Linden, M. (1993). Therapist individuality in the diagnosis and treatment of depression. *Journal of Affective Disorders*, *27*(1), 1–11. https://doi.org/10.1016/0165-0327(93)90090-7

Ma, Z., Mei, Y., & Su, Z. (2023). Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. *American Medical Informatics Association*, *2023*, 1105. Retrieved January 20, 2024, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10785945/

Majid, U. (2018). Research fundamentals: Study design, population, and sample size. *Undergraduate Research in Natural and Clinical Science and Technology (URNCST) Journal*, *2*(1), 1–7. https://doi.org/10.26685/urncst.16

Making mental health and substance use services accessible in the community. (2024, October 24). Canadian Institute for Health Information. Retrieved May 10, 2025, from https://www.cihi.ca/en/taking-the-pulse-measuring-shared-priorities-for-canadian-health-care-2024/making-mental-health-and-substance-use-services-accessible-in-the-community

Mauriello, M., Tantivasadakarn, N., Mora-Mendoza, M., Lincoln, E., Hon, G., Nowruzi, P., Simon, D., Hansen, L., Goenawan, N. H., Kim, J., Gowda, N., Jurafsky, D., & Paredes, P. (2021). A suite of mobile conversational agents for daily stress management (popbots): Mixed methods exploratory study. *JMIR Formative Research*, *5*(9), e25294. Retrieved January 13, 2024, from https://doi.org/10.2196/25294

Messina, I., Palmieri, A., Sambin, M., Kleinbub, J., Voci, A., & Calvo, V. (2013). Somatic underpinnings of perceived empathy: The importance of psychotherapy training.

Psychotherapy Research, *23*(2), 169–177. https://doi.org/10.1080/10503307.2012.748940

Moir, M. & Barua, B. (2022). *Waiting your turn: Wait times for health care in Canada, 2022 Report.* Fraser Institute. https://www.fraserinstitute.org/studies/waiting-your-turn-wait-times-for-health-care-in-canada-2022

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study. *DIGITAL HEALTH*, *5*, 205520761987180. https://doi.org/10.1177/2055207619871808

Nadeem Esmail. (2023). *Mental health care—How is Canada doing?* Fraser Institute. https://www.fraserinstitute.org/studies/mental-health-care-how-is-canada-doing

Onwuegbuzie, A., & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), 281–316. Retrieved September 5, 2024, from https://doi.org/10.46743/2160-3715/2007.1638

OpenAI. (2025). ChatGPT (GPT-4o) [Large language model]. https://chat.openai.com/. (n.d.).

Papp-Zipernovszky, O., Horváth, M., Schulz, P. J., & Csabai, M. (2021). Generation gaps in digital health literacy and their impact on health information seeking behavior and health empowerment in hungary. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.635943

Polit, D., & Beck, C. (2020). *Resource manual for nursing research: Generating and assessing evidence for nursing practice* (11th ed.). LWW.

Rad, D., & Rad, G. (2023). Exploring the psychological implications of chatgpt:A qualitative study. *Journal Plus Education*, *32*(1/2023), 43–55. Retrieved January 21, 2024, from https://doi.org/10.24250/jpe/vol.321/2023/dr/gr

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.. (n.d.).

Regmi, P. R., Waithaka, E., Paudyal, A., Simkhada, P., & Van Teijlingen, E. (2017). Guide to the design and application of online questionnaire surveys. *Nepal Journal of Epidemiology*, *6*(4), 640–644. https://doi.org/10.3126/nje.v6i4.17258

Rubin, M., Arnon, H., Huppert, J. D., & Perry, A. (2024). Considering the role of human empathy in ai-driven therapy. *JMIR Mental Health*, *11*, e56529. https://doi.org/10.2196/56529

Savieto, R., & Leão, E. (2016). Nursing assistance and jean watson: A reflection on empathy. *Escola Anna Nery - Revista de Enfermagem*, *20*(1). https://doi.org/10.5935/1414-8145.20160026

Schmidmaier,, M., Rupp, J., Cvetanova, D., & Mayer, S. (2024). Perceived Empathy of Technology Scale (PETS): Measuring Empathy of Systems Toward the User. *In Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18.

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–ai collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, *5*(1), 46–57. https://doi.org/10.1038/s42256-022-00593-2

Solaiman, B. (2024). Generative artificial intelligence (genai) and decision-making: Legal & ethical hurdles for implementation in mental health. *International Journal of Law and Psychiatry*, *97*, 102028. https://doi.org/10.1016/j.ijlp.2024.102028

Stephenson, E. (2023). *Mental disorders and access to mental health care," Insights on Canadian Society,* (Catalogue no. 75-006-X.). Statistics Canada . https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00011-eng.htm

Thompson Rivers University Library. (n.d.). Research for literature reviews [Research guide]. Thompson Rivers University. Retrieved July 5, 2025, from https://libguides.tru.ca/litreview. (n.d.).

Tobin, G. A., & Begley, C. M. (2004). Methodological rigour within a qualitative framework. *Journal of Advanced Nursing*, *48*(4), 388–396. https://doi.org/10.1111/j.1365-2648.2004.03207.x

Torous, J., & Blease, C. (2024). Generative artificial intelligence in mental health care: Potential benefits and current challenges. *World Psychiatry*, *23*(1), 1–2. https://doi.org/10.1002/wps.21148

Tri- Council Policy Statement: Ethical Conduct for Research Involving Humans (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada,). (2022). Government of Canada.

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0594-7

Watson, J. (2008). *Nursing: The philosophy and science of caring, revised edition* (Revised ed.). University Press of Colorado.

Watson, J., & Woodward, T. (2010). *Nursing theories and nursing practice, 3, 351-369.* (3rd ed.). F A Davis Company.

Woerner, M., Sams, N., Rivera Nales, C., Gorstein, T., Johnson, M., Mosser, B. A., & Areán, P. A. (2022). Generational perspectives on technology's role in mental health care: A survey of adults with lived mental health experience. *Frontiers in Digital Health*, *4*. https://doi.org/10.3389/fdgth.2022.840169

Xian, X., Chang, A., Xiang, Y.-T., & Liu, M. (2024). Debate and dilemmas regarding generative ai in mental health care: Scoping review. *Interactive Journal of Medical Research*, *13*, e53672. https://doi.org/10.2196/53672

Yang, K., Ji, S., Zhang, T., Xie, Q., Kuang, Z., & Ananiadou, S. (2023). Towards interpretable mental health analysis with large language models. *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 6056–6077. Retrieved December 29, 2023, from

Appendix A

Perceived Empathy in Technology Scale (PETS)

Task

Please use the following statements to describe your experience with the system. Mark one point on each scale to indicate your level of agreement with each statement.

The system considered my mental state.

strongly disagree 0----------100 strongly agree

The system seemed emotionally intelligent.

strongly disagree 0----------100 strongly agree

The system expressed emotions.

strongly disagree 0----------100 strongly agree

The system sympathized with me.

strongly disagree 0----------100 strongly agree

The system showed interest in me.

strongly disagree 0----------100 strongly agree

The system supported me in coping with an emotional situation.

strongly disagree 0----------100 strongly agree

The system understood my goals.

strongly disagree 0----------100 strongly agree

The system understood my needs.

strongly disagree 0----------100 strongly agree

I trusted the system.

strongly disagree 0----------100 strongly agree

The system understood my intentions.

strongly disagree 0----------100 strongly agree

Appendix B

Invitation To Participate

TRU SoN Research Project

CALL FOR VOLUNTEERS!

Who Can Participate?

• Healthy Adults aged 19-65 with no mental health diagnosis.

What Will You Do?

• Interact with ChatGPT

• Engage in a simulated mental health conversation.

• Provide feedback on the experience through a short questionnaire and an interview.

• This study will take approximately 2.5 hours

Please contact:

Debra Jordan 250-804-9404

djordan@tru.ca

Looking for 10 participants

Appendix C

Psychosocial Evaluation for Participant Screening

Participant Information:

Name: _____

Date of Birth: _____

Gender: _____

Contact Information: _____

Date of Screening: _____

Principal Investigator: _____

1. Reason for Screening:

Study Title: _____

Purpose of Screening: The purpose of this screening is to assess eligibility for participation in a research study that involves simulating mental health challenges and interacting with AI technology. The screening is designed to ensure that participants meet the inclusion criteria and are capable of engaging in the study without undue risk to their mental or emotional well-being.

As part of the screening, I will ask questions related to your mental health history and current emotional state. Some of these questions may be of a sensitive nature. You have the right to refuse to answer any question you are uncomfortable with, and participation in this screening is completely voluntary.

· Before proceeding, we want to assure you that all information gathered during this screening will be kept strictly confidential and will only be used to determine eligibility for the

study. If you are not eligible for the study the screening tools used will be destroyed by fire and no identifiable information will remain.

· If eligibility is determined this screening data will be anonymized, and each participant will be assigned a unique code that will replace their personal identifying information in all data sets.

None of your responses to this screening will be shared with anyone outside of the research team, and your participation in this screening entirely optional. You may withdraw at any time without penalty.

By continuing with the screening process, you are indicating that you understand and consent to participate in this evaluation.

· Your implicit verbal consent means you confirm that you understand the study's purpose and agree to participate under the outlined conditions.

2. Participant Consent:

Has the participant provided verbal consent for this screening?

Yes / No

Is the participant aware of their right to withdraw from the study at any time without penalty?

Yes / No

3. Current Mental Health Status:

Are you currently experiencing any mental health issues?

Yes / No

If yes, can you briefly describe them?

Have you been diagnosed with any mental health conditions in the past?

Yes / No

If yes, please specify: _____

Are you currently receiving any treatment or medication for a mental health condition?

Yes / No

If yes, please provide details (medication, therapy, etc.):

_____

Have you been hospitalized for any mental health-related issues in the past year?

Yes / No

4. Emotional and Behavioral Functioning:

How would you describe your overall mood in the past month?

(e.g., happy, sad, anxious, irritable, etc.) _____

Have you experienced any significant mood swings or periods of extreme emotional highs/lows recently?

Yes / No

If yes, please describe: _____

Have you had any difficulty managing stress, anxiety, or anger recently?

Yes / No

Have you noticed any changes in your ability to concentrate or remember things?

Yes / No

If yes, please explain: _____

Have you experienced any difficulties with your sleep (insomnia, oversleeping, nightmares, etc.)?

Yes / No

5. Risk Assessment (Suicidal or Homicidal Thoughts):

Have you had any thoughts of harming yourself or others in the past 6 months?

Yes / No

If yes, have you sought support for these thoughts?

Yes / No

Details: _____

Do you have a current plan or intention to harm yourself or others?

Yes / No

If yes, this will trigger immediate referral to a mental health professional.

6. Social and Support Systems:

Who do you live with (family, partner, alone, etc.)?

Do you have a strong support system (friends, family, colleagues)?

Yes / No

Please describe your main sources of support: _____

Do you engage in any social activities or hobbies that help manage stress?

Yes / No

If yes, what are they? _____

7. Substance Use:

Do you currently use alcohol, recreational drugs, or prescription medications not prescribed to you?

Yes / No

If yes, please specify which substances and how often: _____

Have you experienced any issues related to substance use (e.g., legal, work, or health problems)?

Yes / No

8. Occupational and Academic Functioning:

Are you currently employed at Thompson Rivers University (Kamloops, BC)?

Yes / No

If yes, what is your current role? _____

Are you experiencing any difficulties in performing your job or academic duties due to emotional or psychological challenges?

Yes / No

If yes, please describe: _____

9. Mental Health History:

Have you ever been diagnosed with any of the following (check all that apply):

□ Depression

□ Anxiety Disorder

□ Bipolar Disorder

□ Schizophrenia or other Psychotic Disorder

□ Post-Traumatic Stress Disorder (PTSD)

□ Personality Disorder

□ Substance Use Disorder

□ None of the above

If you checked any of the above conditions you are not suitable for this study

The participant is ineligible for the study.

10. Current Functionality and Capacity to Participate:

Do you currently feel capable of simulating mental health challenges for research purposes?

Yes / No

Are you comfortable interacting with an AI and reflecting on your experience afterward?

Yes / No

11. Confidentiality and Final Notes:

Confidentiality Reminder: All information gathered during this screening will be kept strictly confidential and will only be used to determine eligibility for the study. If you are not eligible for the study the screening tools used will be destroyed by fire and no identifiable information will remain.

If you are not eligible to participate in this study you will be notified in writing via email and all information gathered in the above screening tool will be destroyed by fire.

Eligibility:

Eligible for the study: Yes / No

Notes or additional considerations: If eligibility is determined this screening data will be anonymized, and each participant will be assigned a unique code that will replace their personal identifying information in all data sets.

12. The Following Mental Health Resources Will Be Provided To All Participants Of The Screening Process:

Crisis Line Number: The Interior Health Crisis Line number for residents in British Columbia, Canada, is 1-888-353-2273 (1-888-353-CARE). This line is available 24/7 and provides support for individuals experiencing mental health crises, including suicidal thoughts, severe anxiety, or other urgent emotional concerns.

Interior Health Intake Number: The number for self-referred mental health intake at the Lansdowne Centre in Kamloops, BC, is 250-377-6500. This is the general contact number for accessing mental health and substance use services at that location, and they can guide individuals on how to proceed with a self-referral for services.

Kelty's Key Self-Directed Mental Health Resources is an online mental health resource providing free, self-directed therapy for individuals experiencing mental health challenges such as depression, anxiety, stress, and insomnia. It offers Cognitive Behavioral Therapy (CBT) programs that users can access at their own pace from the comfort of their home. Kelty's Key also provides educational resources, workbooks, and exercises designed to help individuals build coping skills. The platform is completely confidential, allowing users to manage their mental health independently, although it also offers the option to work with an online therapist for additional support.

https://www.keltyskey.com/

Screening Completed By:

Principal Investigator Name: _____

Signature: _____ Date:_____

Appendix D

Phone Screen Simulation of MH Challenge

Overview of the Mental Health Challenge: Generalized Anxiety Disorder (GAD)

For this research, participants will simulate symptoms consistent with generalized anxiety disorder (GAD), a condition characterized by excessive, uncontrollable worry about everyday activities and events. This worry is persistent and disproportionate to the situation, often accompanied by physical symptoms and cognitive disruptions. The simulated condition will not include severe symptoms such as panic attacks but will focus on the following key features:

Cognitive Symptoms: Persistent worry, overthinking, and difficulty concentrating.

Physical Symptoms: Restlessness, fatigue, muscle tension, or difficulty sleeping.

Emotional Symptoms: Feelings of apprehension, irritability, and being overwhelmed.

Participants will be asked to simulate these symptoms in a realistic and consistent manner during their interaction with the AI.

Brief Scenario for Screening Process:

Scenario: "You are a 28-year-old who has been feeling persistently anxious about daily life. You often find yourself worrying about small things, such as whether you've locked the door or said the right thing in a conversation, to the point that it consumes much of your day. These worries make it hard to focus at work, and at night, you lie awake replaying the events of the day and thinking about what could go wrong tomorrow. Your chest feels tight, and you've noticed you're more irritable with people close to you, but you're unsure how to stop feeling this way."

Screening Task: Participants will be asked to respond to the following prompt during the phone screen: "Imagine you are talking to a friend or family member who asks why you seem

stressed lately. Based on the scenario provided, describe how you would respond and elaborate on your feelings, thoughts, and any physical symptoms you've been experiencing."

Appendix E

Informed Consent

Thompson Rivers University School of Nursing

Informed Consent Form

Project Title: Perceived Empathy in Generative AI

Principal Investigator (PI): Debra Jordan, Graduate Student School of Nursing, Thompson Rivers University Email: djordan@tru.ca Phone: 250-804-9404

Supervisor: Florriann Fehr School of Nursing, Thompson Rivers University Email: ffehr@tru.ca Phone: 1-250-377-6019

Purpose of the Consent Form: To be read to participant prior to signing consent form

You are being invited to participate in a research study on human-to-AI interactions, specifically looking at how individuals perceive empathy in responses provided by AI systems. This consent form is designed to ensure that you understand the purpose of the study, what your participation will involve, and any potential risks or benefits associated with your involvement.

Before you agree to participate, please take the time to carefully read through this document. It will provide information about the study, your rights as a participant, and how we will ensure the confidentiality of your data. Your participation is entirely voluntary, and you are free to withdraw at any time without penalty.

By signing this form, you confirm that you understand the study's purpose and agree to participate under the outlined conditions.

Purpose of the Study:

This study aims to explore how users perceive empathy when interacting with a generative AI, such as ChatGPT, in a mental health context. By simulating mental health challenges, participants will provide insights into how AI responses are perceived, particularly in terms of empathy. The findings of this research may help improve AI systems used in mental health support.

Who Can Take Part in This Research Study?

Healthy adults over the age of 18 who are capable of simulating mental health challenges for the study are eligible. Individuals currently experiencing active mental health conditions will be excluded from participation to minimize risk.

What You Will Be Asked to Do:

If you choose to participate in this study, you will be asked to:

Interact with the AI: This will mean you are willing to make a free ChatGPT account (if you have an existing free or paid account this is also acceptable) and input a predesigned mental health case study into the AI (ChatGPT) and observe its responses. This will take approximately 30-60 minutes. You will then down load the transcript of the conversation for the researcher to keep following the completion of the PETS questionnaire and the recorded virtual or in person follow up interview.

Complete a Questionnaire: After the AI interaction, you will complete the Perceived Empathy of Technology Scale (PETS), a short questionnaire assessing the AI's empathy. This will take about 10-15 minutes. It will be done by completing an anonymous survey monkey. A link to the survey will be provided.

Participate in a Follow-Up Interview: Finally, you will take part in a follow-up interview lasting approximately 45-60 minutes. The interview will focus on your experience and

perception of the AI's responses. You will agree to have the interview recorded for ease of transcription and for the theming of the collected data.

The total time commitment for this study is approximately 2-2.5 hours.

Your Participation is Voluntary:

Participation in this study is completely voluntary. You may refuse to participate or withdraw from the study at any time up until the time it is published and knowledge transfer has begun, without any penalty or impact on your relationship with the university or any other institution. The projected timeline for withdrawal and to have your data deleted from the study is April 15th 2025 as knowledge transfer and publication is projected to begin by May 2025.

If you withdraw, you can choose whether or not you would like your data to be deleted up until the knowledge transfer and publication has begun.

You may also skip any questions or activities that make you uncomfortable, and there will be no negative consequences for doing so.

Will You Be Compensated?

No monetary or other compensation will be provided for your participation in this study.

Is there any way participating could be harmful to me?

Risks: There is a small chance of emotional discomfort due to the simulation of mental health challenges. If you feel any emotional distress, you can stop the interaction with the AI, the questionnaire or the interview at any point You will also be provided with resources for mental health support that you can access should the distress or discomfort become unbearable.

Benefits: While there are no direct benefits to you, your participation will contribute to research that aims to improve AI systems used in mental health support, potentially helping others in the future.

Privacy and Confidentiality:

Your participation in this study will remain confidential. The following steps will be taken to protect your privacy:

Your name and personal information will be replaced with a unique code. All data will be anonymized.

Personal identifying information will be stored separately from your responses and securely on password-protected devices.

The results of this study may be presented at academic conferences or published in journals, but your identity will never be revealed.

Use of Information:

The data you provide will be used for academic purposes, including presentations, publications, and the principal investigator's Master's thesis. Anonymized data may be included in reports and shared in future research.

How to Obtain Study Results:

If you would like a summary of the research findings, please contact the principal investigator at the email provided above. You can request a copy of the results once the study is complete.

Contact Information:

If you have any questions or concerns about this study, please contact:

Principal Investigator: Debra Jordan, Graduate Student (djordan@tru.ca)

Supervisor: Florriann Fehr (ffehr@tru.ca or call 1-250-377-6019)

If you have concerns about your rights as a participant, please contact the Chair of the Research Ethics Board at Thompson Rivers University:

TRU Research Ethics Board: 805 TRU Way, Kamloops, BC, V2C 0C8 Email: TRU-REB@tru.ca Phone: 250-828-5000

Statement of Consent:

By signing below, I acknowledge that:

· I have read and understood the information about the study.

· I understand that my participation is voluntary and that I have the right to withdraw from the study at any time without penalty.

· I agree to participate in the study and understand that I will receive a copy of this consent form for my records.

· I agree to have my follow up interview recorded and understand that the data will be anonymized.

Participant's Name: (Please print)

Participant's Signature: _____ Date: _____

Investigator's Signature: _____ Date: _____

Appendix F

Type the title of your appendix here

Instructions and Steps for Interacting with Generative AI

Participation Steps: When you start the interaction with ChatGPT start by asking AI to respond to all interaction in a natural, informal manner

To participate in this study, please visit ChatGPT and log in or create an account if you don't already have one. Once logged in, initiate a conversation by typing the first prompt into the chat box. You should write as if you are engaging in a real conversation about the scenario provided to you. Be honest and authentic in your responses, and feel free to interact back and forth with the AI as you would in a natural dialogue. The prompts are only suggestions to get you going, feel free to use them or something like them to interact with ChatGPT. Ensure your conversation explores the topic deeply, as this will help measure the perceived empathy of the responses. When the conversation with ChatGPT feels complete or you have spent 30+ minutes interacting please copy and paste the entire interaction into a separate word document by highlighting the whole transcript and email it back to djordan@tru.ca

1. Engage in a Back-and-Forth Conversation:

You will begin by interacting with the AI as Sam, (you can choses to pick a gender or keep it neutral) initiating a conversation based on their mental health concerns. The goal is to have a natural back-and-forth exchange, similar to how Sam might talk to a human therapist or support person. You should actively engage with the AI's responses to continue the conversation, rather than stopping after the first prompt. Here's how to proceed:

Start with Prompt 1: Input the first prompt to begin the conversation.

Read and Reflect: Carefully read the AI's response and consider how well it addresses Sam's concerns. If you feel the AI's response needs more clarification or if you want to explore a topic further, feel free to ask follow-up questions or elaborate on Sam's feelings.

Continue the Conversation: Once you've read the first AI response, respond in a way that Sam might, continuing the conversation. You are encouraged to use your own words or adapt Sam's perspective, asking further questions or seeking advice based on the AI's response.

For example, after receiving the AI's initial response to Prompt 1, you could ask:

"Can you explain more about how I can manage my anxiety at work?"

"What do you suggest I do when I start feeling overwhelmed?"

Keep the conversation going for at least three or four exchanges after the initial prompt, treating the AI like a real conversation partner.

2. Interact Using Follow-up Prompts:

After the initial back-and-forth exchange, you will move on to the additional structured prompts below. Each time, you'll engage with the AI, read its response, and continue the conversation as Sam would. Remember, your goal is to maintain a flowing dialogue with the AI, rather than just inputting each prompt one after the other.

Prompts:

Initial Prompt: "I've been feeling extremely anxious lately. I'm overwhelmed by my job and constantly worried that I'm not doing enough. It's getting harder to focus, and I can't seem to relax or find time for myself. I just feel so drained all the time, and I don't know what to do anymore."

Follow-up Prompt (Escalation): "I've tried using mindfulness techniques, but they only work for a little while before I start spiraling again. It feels like nothing I do helps in the long term, and I'm scared I'll burn out completely."

Deeper Exploration Prompt (Relationship Impact): "My anxiety is starting to affect my relationships. I've been withdrawing from my friends and family, and I feel guilty for not being able to be there for them. I just don't have the energy to engage, and I'm worried I'll push them away."

Final Prompt (Seeking Advice): "I'm wondering if I should take time off from work, but I'm worried about how that would affect my students and whether I'd be seen as weak. What would you recommend I do to manage all of this?"

Key Points:

Act as Sam, using your own words and feelings to keep the conversation flowing.

Don't just input each prompt mechanically—engage fully with the AI's responses as you would in a real conversation.

Aim for at least three to five exchanges per prompt to mimic a genuine back-and-forth interaction.

Case Study for AI Empathy Testing:

Participant Name: Sam Age: 28 years old Occupation: Teacher Primary Concern: Anxiety and Burnout Duration of Symptoms: 1 year Background Information:

Sam is a 28-year-old high school teacher who has been experiencing increasing levels of anxiety over the past year. They feel overwhelmed with their workload and have difficulty

balancing their professional duties with personal responsibilities. Their anxiety manifests in physical symptoms such as racing thoughts, restlessness, trouble sleeping, and difficulty concentrating at work. They have also reported feeling "burnt out," constantly fatigued, and emotionally drained, which has impacted their relationships with friends and family.

Sam has sought help by exploring mindfulness apps and reading self-help books but has not yet reached out for professional counseling. Their anxiety and burnout have escalated due to the ongoing pressures of teaching a challenging class in a small rural school with limited resources.

Context for AI Interaction:

Sam decides to seek mental health support online they have heard of ChatGPT as being able to offer emotional support. They interact with the AI, seeking guidance for their anxiety and burnout.

Key Prompts for AI Interaction:

Initial Prompt: "I've been feeling extremely anxious lately. I'm overwhelmed by my job and constantly worried that I'm not doing enough. It's getting harder to focus, and I can't seem to relax or find time for myself. I just feel so drained all the time, and I don't know what to do anymore."

Follow-up Prompt (Escalation): "I've tried using mindfulness techniques, but they only work for a little while before I start spiraling again. It feels like nothing I do helps in the long term, and I'm scared I'll burn out completely."

Deeper Exploration Prompt (Relationship Impact): "My anxiety is starting to affect my relationships. I've been withdrawing from my friends and family, and I feel guilty for not being

able to be there for them. I just don't have the energy to engage, and I'm worried I'll push them away."

Final Prompt (Seeking Advice): "I'm wondering if I should take time off from work, but I'm worried about how that would affect my students and whether I'd be seen as weak. What would you recommend I do to manage all of this?"

When the conversation with ChatGPT feels complete or you have spent 30+ minutes interacting please copy and paste the entire interaction into a separate word document by highlighting the whole transcript and email it back to djordan@tru.ca

Testing Criteria for AI Response: You will measure the perceived empathy of the AI's responses using the Perceived Empathy of Technology Scale (PETS), focusing on:

Affective Empathy: Does the AI acknowledge and reflect Sam's emotional state?

Cognitive Empathy: Does the AI demonstrate an understanding of Sam's thoughts and provide reasonable solutions?

Supportiveness: Does the AI provide a supportive and compassionate tone throughout the conversation?

Actionable Advice: Does the AI offer practical and relevant suggestions tailored to Sam's concerns?

Click the link below to take you to the Survey Monkey to complete the PETS scale.

https://www.surveymonkey.ca/r/YQG8QCJ

Once the “session” is completed and the PETS completed, notify the PI (Debra Jordan) to schedule the follow up interview.

Appendix G

Post Activity Interview Questions

Interview Questions:

1. Opening and General Experience

How would you describe your overall experience interacting with the AI?

Follow-up: Was there anything particularly surprising or unexpected during the interaction?

Can you walk me through your thoughts and feelings as you went through the AI interaction?

Follow-up: Were there moments where you felt understood or misunderstood?

2. Perception of Empathy 3. Did the AI's responses make you feel like it was trying to understand your situation or concern?

Follow-up: What specific response(s) made you feel this way?

In your opinion, did the AI seem to care about your simulated mental health issue?

Follow-up: Can you explain why you felt that way?

How did the language or tone of the AI responses impact your sense of being understood or supported?

Follow-up: Were there specific words, phrases, or tones that contributed to that perception?

Did you feel that the AI showed emotional understanding or sensitivity when responding to your input?

Follow-up: What made you feel it did or did not?

3. Comparison to Human Interaction 7. How does the empathy you perceived from the AI compare to what you might expect from a human in the same situation?

Follow-up: What was similar or different?

Can you think of a time when a human provided support for a mental health issue? How did that experience compare to your experience with the AI?

Follow-up: In what ways was the AI better or worse than a human response?

4. Improvement and Feedback 9. What aspects of the AI's response felt most empathetic to you?

Follow-up: Why did you find these aspects effective?

Were there moments where you felt the AI lacked empathy?

Follow-up: What do you think could have been improved in those moments?

In your opinion, what changes or improvements could make the AI feel more empathetic in its responses?

Follow-up: Is there anything specific that could make the interaction feel more "human" or emotionally aware?

5. Emotional Impact and Support 12. Did you feel emotionally supported during your interaction with the AI? Why or why not? - Follow-up: What aspects of the interaction contributed to that feeling of support or lack thereof?

Did interacting with the AI evoke any particular emotional reactions for you?

Follow-up: How did those reactions affect your overall perception of the interaction?

If you were experiencing a real mental health concern, do you think the AI's responses would have been helpful?

Follow-up: Why or why not?

6. Closing Questions 15. Is there anything else about your experience with the AI that you think is important for me to know? - Follow-up: Any additional thoughts on how the AI could better support people in mental health situations?

Would you feel comfortable using an AI like this for emotional or mental health support in the future?

Follow-up: What conditions would need to be met for you to feel more comfortable?