EXPLORING THE WORKING ALLIANCE FORMATION IN AI COMPANIONS

by

Gerardo Castaneda

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Technology

Department of Technology, Leadership and Innovation

West Lafayette, Indiana

August 2025

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Julia M. Rayz, Chair

Department of Computer and Information Technology

Dr. Dawn D. Laux

Department of Computer and Information Technology

Dr. John A. Springer

Department of Computer and Information Technology

Dr. Tatiana R. Ringenberg

Department of Computer and Information Technology

Approved by:

Dr. Stephen J. Elliott

Dedicated to my wife Ana, my life partner through chaos and calm.

ACKNOWLEDGMENTS

My deepest gratitude to Dr. Julia Rayz, my dissertation chair, mentor, and guide throughout the D. Tech program. Dr. Rayz's remarkable ability to identify areas requiring further consideration has been instrumental in helping me reflect, think more deeply, and maintain clarity and direction throughout the program.

I am also very grateful to the members of my dissertation committee: Dr. Dawn D. Laux, Dr. John A. Springer, and Dr. Tatiana R. Ringenberg. Your insightful questions and recommendations helped me focus my research, avoid pitfalls, anticipate research challenges, and proactively design for them.

Thank you as well to the faculty of the D. Tech program, especially Dr. Josh Plaskoff, for expanding my mental frameworks and helping me become a better leader, and Dr. Michael J. Dyrenfurth for the relentless attention to detail.

This dissertation is a result of your collective encouragement, high standards, and willingness to invest your valuable time in me. Thank you for your unwavering support!

AI USE STATEMENT

Various AI tools were used during the development of this dissertation to assist with research, brainstorming, grammar and plagiarism checks, and R coding. While these tools were employed to enhance productivity and efficiency, all outputs were reviewed and verified by the author and remain the author's sole responsibility.

Research and Brainstorming: Scite, ChatGPT (4.0, 4.5, o1-preview, o3, 4o), Gemini (1.5 Flash, Advanced 2.0 Flash), Claude (3.5, 3.7).

Grammar and Plagiarism Checks: Grammarly, Scribbr.

R Coding: Claude 3.5, ChatGPT 4o.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TERMINOLOGY

This dissertation adopts “AI companions”, a term that has recently gained popularity in the media, to refer to AI-powered chatbots used in mental health or emotional support contexts. These AI companions could be rule-based (following decision trees), retrieval-based (using AI to select responses from a database), generative-based (using large language models or other generative AI architectures), or a combination of various AI technologies, with the common feature of being able to engage in fluid natural language conversations with their users. Furthermore, this dissertation uses “clinically validated AI companions” and “validated companions” to refer to AI companions that have documented evidence of successful clinical testing, and “non-validated AI companions” to refer to AI companions that lack evidence of clinical validation. For clarity, “clinically validated” in this context does not imply FDA or regulatory approval. In addition, this dissertation adopts the term “companion type” to refer to groups of validated or non-validated companions. Examples of clinically validated AI companions include Woebot, Wysa, Youper, and Therabot. Examples of non-validated AI companions include ChatGPT, Replika, and Character.ai’s Psychologist. Other terms, such as chatbots or wellness apps, are occasionally used throughout this dissertation when it is unclear how a chatbot is powered or to maintain fidelity with the literature. Throughout this dissertation, the term “mental health” is understood as defined by the World Health Organization (2022)—“Mental health is a state of mental well-being that enables people to cope with the stresses of life, realize their abilities, learn well and work well, and contribute to their community."

ABSTRACT

The literature suggests that AI companions could play an important role in addressing the mental health crisis. However, there is limited research on the factors that may impact the working alliance with AI companions and how clinically validated and non-validated companions compare in their ability to develop a working alliance across varying demographics. This gap is problematic because the strength of the working alliance and patient demographics are closely related to patient outcomes. Therefore, individuals may be using unproven and potentially unsafe solutions. This study examined factors that may influence the working alliance with AI companions across various demographics through a cross-sectional survey of 253 U.S. adults. Depression and anxiety were screened with the PHQ-2 and GAD-2. Working alliance strength was assessed with the WAI-SR. In addition, three open-ended questions inspired by the Technology Acceptance Model were analyzed thematically. Clinically validated companions developed significantly higher working alliances than non-validated companions. Alliance scores did not vary by demographics or anxiety/depression intensity. Frequently mentioned themes across companion types were ease of interaction, a judgment-free environment, and guidance. Validated companions performed better at providing emotional support, serving as an alternative or supplement to therapy, offering help during difficult times, and assisting with planning and setting goals. Non-validated AI companions were praised more frequently for their affordability. Validated AI companions had significantly lower adoption rates than non-validated companions. The findings suggest that most users of AI mental health companions are at risk of suboptimal outcomes, highlighting the need for clinically validated and affordable AI companions, as well as enhanced regulatory frameworks. This study also proposes a revised version of the WAI-SR adapted for AI companions.

INTRODUCTION

Background

With nearly one billion people worldwide affected by mental health conditions, a shortage of mental health care providers, and pervasive barriers such as stigma and cost, there is growing interest in using AI companions to expand access to mental health support.

According to the World Health Organization (2022), multiple individual, social, and structural factors influence mental health: substance use, and adverse conditions such as poverty, inequality, and violence contribute to vulnerabilities, while factors such as social support and neighborhood safety enhance resilience. Individuals with mental health conditions are often sidelined, deprived of fundamental rights, and discriminated against in many ways, including employment, education, and housing (pp. xvi, 11). Furthermore, individuals with severe mental health disorders tend to have lifespans that are 10 to 20 years shorter than the rest of the population (p. 45). Globally, nearly one billion individuals suffer from mental health problems, with depression and anxiety being among the most prevalent conditions, both for males and females (pp. vi, xv).

The World Health Organization (2022, p. xv) also reports that, although mental health issues are pervasive and have severe implications, most countries allocate less than 2% of their healthcare budgets to mental health services. In addition, in many regions, there is a severe shortage of mental health professionals, with some areas having only one psychiatrist per 200,000 individuals. To exacerbate the challenges, factors such as inadequate reimbursement, stigma, and fear of marginalization prevent people from seeking care.

The combination of these challenges, along with recent advances in natural language technologies, has resulted in an increased interest in internet-based and mobile app alternatives to

traditional mental health care. These alternatives often manifest as AI companions that offer immediate, scalable, and cost-effective support (Abd-Alrazaq et al., 2019; Fitzpatrick et al., 2017). Research indicates that online interventions, particularly AI chatbots, can effectively reduce anxiety and depression, often yielding results comparable to human-led therapy (Fitzpatrick et al., 2017).

Problem Statement

The literature suggests that clinically validated AI companions can establish strong working alliances (Darcy et al., 2021; Beatty et al., 2022; Heinz et al., 2025); however, most research has focused on specific AI companions. To the best of the author's knowledge, no studies have systematically explored how the working alliance strength compares between companion types (i.e., multiple non-validated AI companions vs. multiple validated AI companions) across varying demographics, or the factors that may influence the formation of the working alliance. Since the working alliance is a predictor of patient outcomes, this research gap is significant because millions worldwide utilize validated and non-validated AI companions for therapeutic or emotional support purposes (De Freitas et al., 2023) and because the global nature of the mental health crisis necessitates solutions that support diverse demographics.

This research explored factors that might influence the formation of the working alliance with validated and non-validated AI companions and across varying demographics, with a focus on individuals experiencing anxiety or depression.

Research Question and Hypotheses

This work explores the following research question: How do companion type, sex, sexual orientation, race/ethnicity, and anxiety/depression severity influence the strength of the working

alliance? The research question was separated into two sub-questions: 1) Are there statistically significant differences in WAI-SR scores between AI companion types? 2) Are there statistically significant differences in WAI-SR scores by race, sex, sexual orientation, or anxiety/depression intensity within each AI companion type?

Null Hypothesis (H_0): Companion type, sex, sexual orientation, race/ethnicity, and anxiety/depression severity do not significantly affect the working alliance with AI companions.

Alternative Hypothesis (H_1): At least one of companion type, sex, sexual orientation, race/ethnicity, or anxiety/depression severity significantly affects the working alliance with AI companions.

Significance and Contribution

This study addressed an existing gap in research on the factors that influence the ability of AI companions to develop a working alliance across varying demographics. This is important because the working alliance has been consistently found to be a strong predictor of patient outcomes, regardless of the therapeutic modality, and millions worldwide use AI companions for mental health support. Ultimately, this study aims to make mental healthcare more accessible and effective by providing insights to developers, practitioners, and regulatory bodies on the development of AI companions.

Assumptions

The study's design and analysis are based on the following assumptions:

1) The WAI-SR instrument can be applied to AI companions. This assumption has been employed in multiple previous studies, including those conducted by Darcy et al.

(2021), Beatty et al. (2022), and Heinz et al. (2025), where references to ‘therapist’ or ‘therapy’ were replaced with the name of the AI companion.

2) The use of binary groupings for sex (male, female), sexual orientation (heterosexual, other), race/ethnicity (white, people of color), and symptom severity (above cutoff, below cutoff) can allow for the identification of valuable insights.

Delimitations

To maintain focus and manage scope in alignment with practical constraints, the following delimitations were set for this study:

1) The study focused on individuals who are already using AI companions for therapeutic or mental health support purposes and who declared having depression or anxiety. The rationale for this delimitation is twofold:
    a) The WAI-SR scores are more meaningful if applied to relationships where the participants are working towards shared objectives.
    b) Depression and anxiety are the most common mental health disorders, frequently targeted by AI companions, and relatively easy to identify with validated instruments such as the PHQ-2 and GAD-2.
2) The study was further delimited to individuals living in the US who are legally authorized to provide consent. This delimitation simplified compliance with applicable rules and regulations.
3) To manage scope and complexity, this study was limited to two categories for each group analyzed (for example, White vs. People of Color, Male vs. Female, Heterosexual vs. Other).

4) The study did not consider all potential factors that could influence the working alliance. Examples of factors not included are socioeconomic status, education level, age, and duration of AI companion use.
5) The study and recruitment were only conducted in English.

Limitations

The following limitations outline considerations that may affect the interpretation, generalizability, or completeness of the study's findings:

1) The study does not explain the impact on the working alliance that could result from factors not explicitly included, such as age, socioeconomic status, education, technology competence, cultural differences, or functionality of AI companions.
2) Since participants are self-selected individuals who are already using AI companions for therapeutic purposes and who respond to Prolific advertisements, the sample may not be representative of the broader population.
3) Focusing only on individuals with depression or anxiety limits the applicability of the findings to users with other mental health conditions.
4) The influence of Language Models (LLMs) biases was not assessed.
5) Results may have been impacted by changes introduced by the developers of AI companions during the testing period.
6) The study did not assess changes in the working alliance with AI companions over time.

REVIEW OF THE LITERATURE

This section describes the approach used to identify and select literature relevant to the study's focus on mental health, working alliance, demographic factors, and the use of AI companions. The search process aimed to capture a wide range of current and relevant sources to inform the study's formulation, objectives, and design.

Search Methodology

A comprehensive literature search was conducted using multiple academic databases and research search engines, including Purdue Libraries, PubMed, Google Scholar, ProQuest, IEEE, APA, AMA, Elicit.com, and Scite, as well as general search engines such as Google and DuckDuckGo. Representative examples of search logic include, but are not limited to, ('mental health' OR 'mental health crisis'), ('mental health') AND ('sex' OR 'sexual orientation'), ('mental health' AND ('demographics' OR 'race' OR 'ethnicity'), ('mental health' AND ('chatbots' OR 'conversational agents' OR 'companions') AND 'metrics'), ('working alliance' OR 'therapeutic alliance' OR 'therapeutic bond'), (('depression' OR 'anxiety' OR 'working alliance') AND ('measurement')). Other search strings and search results were produced with large language models using prompts such as "provide a list of recent peer-reviewed scholarly articles that discuss mental health chatbots and the working alliance.", or "provide a list of recent scholarly articles that discuss how mental health is impacted by sex, sexual orientation, and race." The researcher validated all results from Generative AI queries.

Search results were scanned for relevance based on their titles and abstracts, and references were checked to incorporate additional titles. Given the pace of AI innovation, particularly in large language models (LLMs), priority was given to documents published in

2018 or later that discuss AI-powered chatbots. After reviewing the full text of 149 relevant titles, 97 references were used to support this dissertation. Although this summary appears linear, the actual process involved considerable iteration and refinement.

Literature review findings

The literature review findings are organized into thematic subsections reflecting the key areas relevant to this study: the mental health crisis; identity-related disparities in depression and anxiety; tools for measuring depression and anxiety; the working alliance and tools to measure it; and the use of AI companions in mental health. Some sections are more detailed than others due to the uneven distribution of existing research and the specific focus of this study. For example, research on AI companions in mental health yielded a broader body of literature than research on working alliance measures with AI companions. The overall structure of this section is also intended to guide the reader from foundational context toward more targeted considerations that directly inform the study's design and objectives.

The mental health crisis

According to the World Health Organization (2022), mental health helps individuals manage stress effectively, fully utilize their abilities, and contribute meaningfully to their communities.

As of 2019, nearly 970 million people (52.4% females and 47.6% males) globally were reported to experience mental health conditions, with depression and anxiety ranking as the most prevalent disorders (World Health Organization, 2022). In contrast, the global median of mental health workers in 2020 was 13 per 100,000 people (World Health Organization, 2021). In addition to the demand and supply gap, multiple barriers, including mental health illiteracy, cost,

stigma, and discrimination, prevent people from seeking mental healthcare (World Health Organization, 2022; Balcombe, 2023; Blease et al., 2020).

It is estimated that more than 59 million people (over one in five) in the U.S. were living with a mental health condition in 2022, and only 50.6% of them received treatment in one year. In the same year, 15.4 million (6% of US adults) had a serious mental health condition, which is a disorder that significantly interferes with life activities (National Institute of Mental Health, 2024).

The most common mental health disorders in the world are anxiety and depression (31% and 28.9%, respectively) (World Health Organization, 2022). In the US, the figures vary slightly, with an estimated 34% for anxiety and 21% for depression (Mehta et al., 2021).

Worldwide, mental health disorders account for approximately one in six years in disability (World Health Organization, 2022; Kilbourne et al., 2018), and more than one in 100 deaths are caused by suicide (World Health Organization, 2022). People suffering from mental health disorders face a higher risk of morbidity and premature mortality (Kilbourne et al., 2018), and people with severe mental health conditions tend to die 10 to 20 years earlier (World Health Organization, 2022).

Populations facing a higher risk of mental health conditions encounter multiple other challenges, including poor living conditions, diminished social and economic status, stigma, and discrimination (World Health Organization, 2021). Mental health disorders affect all ages and can be caused by multiple factors. For example, several studies have identified connections between depression, anxiety, and social media use. In 2023, the US Surgeon General highlighted the stress and anxiety that children experience from exposure to harmful content, bullying,

harassment, and unrealistic body image standards on social media (Abbasi, 2023; Balcombe, 2023).

Quality mental health should be "safe, effective, patient-centered, timely, efficient, and equitable" (Kilbourne et al., 2018). However, today's mental health care faces significant gaps due to a lack of access, coverage, and the absence of evidence-based systematic methods for measuring quality. Although some progress has been made with tools and frameworks such as PHQ-9 scores, there is still a lack of a widely adopted set of 'mental health vital signs', and outcome measures covering various conditions that both providers and insurers can use (Kilbourne et al., 2018).

Depression, anxiety, and identity

Depression may result in a lack of interest, sleep problems, feelings of guilt, and a decline in functioning, and is the number two cause of disability globally (Bowie et al., 2022). Anxiety disorders include excessive perceived fear, anticipation of future threats, excessive worry, fear of separation, and panic attacks (American Psychiatric Association, 2022). Depression and anxiety disorders frequently coexist, making diagnosis challenging due to overlapping symptoms like fatigue and insomnia. However, persistent anhedonia (lack of enjoyment or pleasure) is a distinguishing feature of depression, whereas generalized anxiety disorder is primarily characterized by excessive worry (Stein & Sareen, 2015).

Depression and anxiety disorders exhibit disparities when examined through the lenses of sex, sexual orientation, and race/ethnicity:

Sex - The lifetime risk of developing a major depressive disorder is higher for females than for males, and females also have a higher risk of developing a comorbid anxiety disorder than men, while men are more likely to develop a comorbid substance use disorder (Kessler et

al., 2003; van de Venne et al., 2020). It has been proposed that estradiol and progesterone may influence cognitive pathways and be in part one of the reasons women are more vulnerable to anxiety (Li & Graham, 2017). Women, across racial groups, are generally more prone to seeking help for physical and psychological issues than men since masculinity norms, stigma, and social expectations influence the latter (Vogel et al., 2016; Holden et al., 2012). A systematic review of twenty-one studies across six countries and 311,317 individuals found that non-binary youth have poorer mental health outcomes than cisgender and transgender youth (Klinger et al., 2024).

Sexual Orientation - Sexual minorities have been treated as abnormal for decades. Although homosexuality is now recognized as a natural variation of human sexuality and was removed from the Diagnostic and Statistical Manual of Mental Disorders (DSM) in 1973, societal stigma persists. LGB persons report greater rates of mental health disorders compared to heterosexual persons, and there is ample documented evidence showing that LGB individuals experience psychological distress due to factors such as discrimination, stigma, prejudice, and victimization; a theory that is sometimes referred to as the minority stress model (Herek & Garnets, 2007; Meyer, 2003; Gmelin et al., 2022). Multiple countries still criminalize same-sex sexual acts, and many LGB individuals face oppression resulting from the idea that heterosexuality is better than other sexual orientations. Concerning mental health, there is consensus that LGB individuals have a higher risk of anxiety, depression, and suicidal ideation than heterosexual individuals (Moagi et al., 2021; Moleiro & Pinto, 2015). According to Russell & Fish (2016), the evidence of greater risk for poor mental health faced by LGB individuals is overwhelming. LGB youth experience higher rates of depression, anxiety, and suicidality. While research has advanced, gaps remain in clinical interventions and understanding the experiences of LGBT youth and how they intersect with factors such as race/ethnicity. Moleiro & Pinto

(2015) also stress that, while the terms lesbian, gay, and bisexual are still widely used, sexual orientation is a continuum.

Race/ethnicity - Many studies have found disparities in the risk of mental health conditions for communities of color compared to non-Hispanic Whites. Thomeer et al. (2023) examined racial and ethnic inequalities in mental health and access to care throughout the COVID-19 pandemic. Individuals from the Black, Hispanic, and Asian communities faced sharper declines in mental well-being and higher anxiety and depression than non-Hispanic Whites. Systemic inequities, financial adversity, and race-related stressors (e.g., police brutality and anti-Asian violence) intensified these disparities. Notwithstanding an increased necessity for mental health support, communities of color encountered greater barriers to accessing care, resulting in higher unmet mental health needs. Williams et al. (2007) analyzed data from a study of 6,082 participants and found that major depressive disorder was more common among African Americans and Caribbean Blacks (~56%) compared to non-Hispanic Whites (38.6%). Furthermore, Black individuals were more likely to report severe functional limitations due to depression, yet they were less likely to receive treatment. Whaley (2001) suggests that the low adoption of mental health services by African Americans may stem partly from cultural mistrust, which is influenced by historical and systemic factors. Research indicates that African Americans often view mental health institutions as reflective of broader White societal biases. When linked to experiences of discrimination and misdiagnosis (e.g., African Americans are overrepresented in involuntary hospitalizations), service utilization patterns and therapeutic relationships are impacted. Cogburn et al. (2024) argue that racial injustice has perpetuated disparities in mental health, such as reduced treatment of mental health conditions for Black individuals compared to that of the general population. These disparities lead “to the current

underestimation, misdiagnosis, and inadequate treatment of mental illness in Black populations" (Cogburn et al., 2024) and ultimately increased morbidity. A thematic analysis of perceptions by providers from the Veterans Affairs system conducted by McMaster et al. (2021) revealed that healthcare providers are uncomfortable discussing cultural or racial issues with patients or colleagues.

Measuring depression and anxiety

Two widely adopted, validated instruments used by mental health professionals to measure depression and anxiety across populations are the Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder 7-item (GAD-7).

The PHQ-9 is a reliable and validated tool for screening depression and its severity in clinical practice and research. The PHQ-9 is self-administered and includes nine items to check for depression symptoms. Patients rate the frequency of each symptom using a 0-3 Likert scale (ranging from "not at all" to "nearly every day"), with a total score ranging from 0 to 27. The total score is tallied and grouped into five categories with breaking points at scores of 0-4 (minimal depression), 5-9 (mild depression), 10-14 (moderate depression), 15-19 (moderately severe depression), and 20 or more (severe depression) (Williams, 2014; Kroenke et al., 2001).

Figure 1.

PQH-9 depression questionnaire

Over the last 2 weeks, how often have you been bothered by any of the following problems?
Please circle your answers.

PHQ-9	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things.	0	1	2	3
2. Feeling down, depressed, or hopeless.	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much.	0	1	2	3
4. Feeling tired or having little energy.	0	1	2	3
5. Poor appetite or overeating.	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down.	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television.	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.	0	1	2	3
9. Thoughts that you would be better off dead, or of hurting yourself in some way.	0	1	2	3
Add the score for each column				

Note. Reprinted from the PHQ-9 and GAD-7 forms, by University Health Services, n.d., Florida State University. Source: https://uhs.fsu.edu/sites/g/files/upcbnu1651/files/docs/PHQ-9%20and%20GAD-7%20Form_a.pdf

The Generalized Anxiety Disorder 7-item (GAD-7) scale is a validated, effective, reliable, and internally consistent tool for screening Generalized Anxiety Disorder (GAD) in clinical practice. The GAD-7 consists of 7 items to assess anxiety symptoms. Response options are rated using a 0-3 Likert scale, with a total score ranging from 0 to 21. The total score is tallied and grouped into five categories, with breaking points at 0-4, 5-9, 10-14, and 15-21, corresponding to minimal, mild, moderate, and severe anxiety, respectively. Increasing scores strongly correlate with functional loss. A cutoff score of 10 optimizes sensitivity (89%) and specificity (82%) for identifying probable GAD (Spitzer et al., 2006).

Figure 2.

GAD-7 anxiety questionnaire

Over the last 2 weeks, how often have you been bothered by any of the following problems?
Please circle your answers.

GAD-7	**Not at all sure**	**Several days**	**Over half the days**	**Nearly every day**
1. Feeling nervous, anxious, or on edge.	0	1	2	3
2. Not being able to stop or control worrying.	0	1	2	3
3. Worrying too much about different things.	0	1	2	3
4. Trouble relaxing.	0	1	2	3
5. Being so restless that it's hard to sit still.	0	1	2	3
6. Becoming easily annoyed or irritable.	0	1	2	3
7. Feeling afraid as if something awful might happen.	0	1	2	3
Add the score for each column				

Note. Reprinted from the PHQ-9 and GAD-7 forms, by University Health Services, n.d., Florida State University. Source: https://uhs.fsu.edu/sites/g/files/upcbnu1651/files/docs/PHQ-9%20and%20GAD-7%20Form_a.pdf

In busy settings, brief instruments offer advantages over long questionnaires. One such instrument is the PHQ-2, which comprises the first two questions of the PHQ-9 and is highly sensitive to depression (0.91) at a cutoff point of 2 or higher (Manea et al., 2016). Similarly, the GAD-2 comprises the first two questions of the GAD-7 and has optimal sensitivity and specificity at a cutoff point of 3 or greater (Plummer et al., 2016).

The working alliance

The working alliance concept in mental health describes the collaborative and trusting relationship between a therapist and a client. A strong working alliance enhances the client's willingness to collaborate with the therapist, facilitating effective communication and agreement on therapy goals, consensus on the tasks or methods used to achieve those goals, and the development of a positive emotional bond between therapist and client. These components work together to foster a sense of cooperation and mutual understanding, which is essential for the therapeutic process (Hatcher & Gillaspy, 2006; Norcross & Lambert, 2018).

Research consistently demonstrates that the strength of the working alliance is a significant predictor of successful therapy outcomes regardless of the therapeutic modality (Horvath & Luborsky, 1993). According to Flückiger et al. (2018), a direct relationship between the therapeutic alliance and treatment outcomes ($r = .278$, $p < .0001$) was consistently observed across over 295 studies irrespective of assessor perspectives, the measures used for the alliance, treatment modalities, patient characteristics at intake, formats (in person or online), and geographical locations. Published between 1978 and 2017, the studies covered 30,000 patients. Bordin (1979) proposed that the working alliance between an individual seeking change and the person serving as a change agent is *potentially the most important element* in achieving change. This proposition is supported by Norcross & Lambert (2018), who argue that the contribution to patient outcomes made by the working alliance is independent of the treatment method and may *even be greater* than that of the method itself.

Measuring the working alliance:

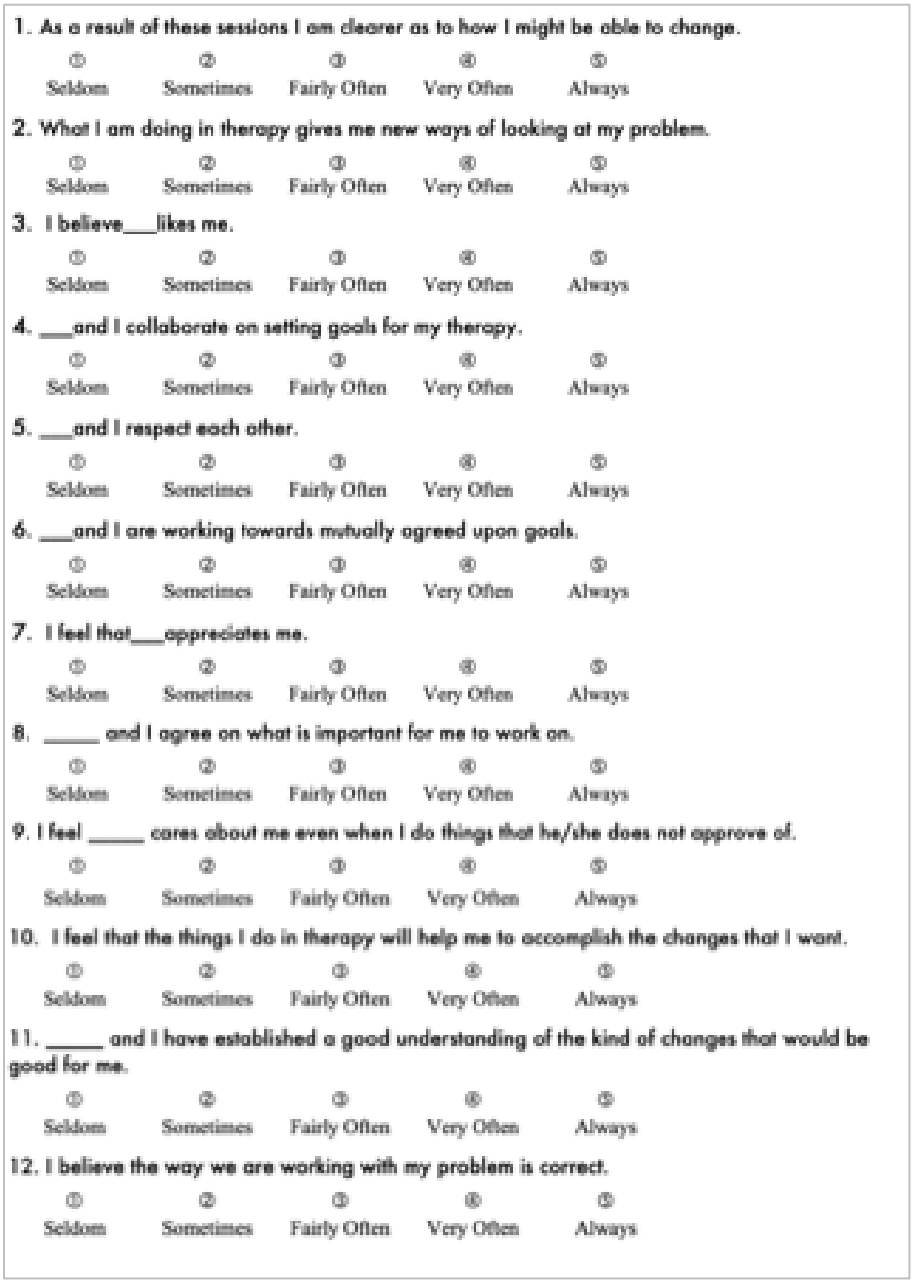
The Working Alliance Inventory (WAI) is a 36-item self-report instrument developed by Horvath and Greenberg to assess Bordin's therapeutic working alliance proposal. The WAI measures three key components of the client-therapist relationship: agreement on therapy goals, agreement on therapy tasks, and the development of a relational bond (Busseri & Tyler, 2003).

The Working Alliance Inventory-Short Revised (WAI-SR) is a 12-item version of the Working Alliance Inventory (WAI), developed by Hatcher & Gillaspy (2006).

Figure 3.

Working Alliance Inventory - Short Revised.

Reprinted by permission of the Society for Psychotherapy Research © 2025.



1. As a result of these sessions I am clearer as to how I might be able to change.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

2. What I am doing in therapy gives me new ways of looking at my problem.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

3. I believe___likes me.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

4. ___and I collaborate on setting goals for my therapy.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

5. ___and I respect each other.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

6. ___and I are working towards mutually agreed upon goals.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

7. I feel that___appreciates me.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

8. _____ and I agree on what is important for me to work on.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

9. I feel _____ cares about me even when I do things that he/she does not approve of.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

10. I feel that the things I do in therapy will help me to accomplish the changes that I want.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

11. _____ and I have established a good understanding of the kind of changes that would be good for me.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

12. I believe the way we are working with my problem is correct.

① Seldom ② Sometimes ③ Fairly Often ④ Very Often ⑤ Always

Note. Reprinted by permission of the Society for Psychotherapy Research © 2025. Source: https://wai.profhorvath.com/sites/default/files/2020-11/WAI_Hatcher_SR.zip

The WAI-SR retains the core components necessary for evaluating the therapeutic relationship while providing a more concise questionnaire that reduces the burden on respondents and produces results with improved differentiation between goal, task, and bond dimensions. The 12 WAI-SR items are divided as follows: Goal Items: 4, 6, 8, 11; Task Items: 1, 2, 10, 12; Bond Items: 3, 5, 7, 9 (Hatcher & Gillaspy, 2006). Each item is rated on a 1- 5-point Likert scale, with a total score ranging from 12 to 60. The validity of the Working Alliance Inventory SR has been demonstrated in countries outside the US, such as Germany (Munder et al., 2010), and has been adapted for use in other languages and cultures. For example, the Working Alliance Questionnaire (WAQ) is a 12-item instrument developed to measure working alliances in Chinese culture, with a high correlation ($r = 0.86$) with the WAI-SR (Li et al., 2022; Xu & Guangrong, 2011). Due to high correlations between goal and task scores, some researchers suggest that these measures overlap and advise against using subscales, instead recommending the adoption of the overall mean as a single score for the entire instrument (Paap & Dijkstra, 2017).

WAI-SR scores in practice. Human-to-human:

Munder et al. (2010) reported mean WAI-SR scores of 3.6 for German inpatients ($n = 243$) and 3.8 for outpatients ($n = 88$). A large-scale study of 14,951 individuals with depression/anxiety using face-to-face blended care therapy (a combination of face-to-face online sessions, with in-between sessions activities such as videos, and digital guides) with Lyra Health reported an initial WAI-SR average of 4.1 for the entire sample, with means for different subgroups ranging between 3.9 and 4.2 (Wu et al., 2024). Another study of 943 individuals in Europe compared the WAI-SR scores between blended care therapy and therapy as usual,

reporting mean WAI-SR scores of 3.95 and 3.5, respectively, with a mean of 3.8 for the entire sample (Doukani et al., 2024).

Table 1.

Summary WAI-SR scores for human-to-human interactions.

Reference	Mean (M) WAI-SR Scores
Munder et al. (2010)	M=3.6 (inpatients, n=243) M=3.8 (outpatients, n=88)
Wu et al. (2024)	M=4.1 (n=14,951)
Doukani et al. (2024)	M=3.8 (n=943)

AI companions in mental health

History of AI companions in mental health:

The origin of mental health chatbots can be traced to 1966 with ELIZA, a chatbot designed to mimic a Rogerian psychotherapist by responding to user inputs through programmed keyword associations (Boucher et al., 2021; Hiland, 2018; Milne-Ives et al., 2020).

In the last decade, AI companions have shown potential in mental health research by leveraging methods such as sentiment analysis to assess emotions from text, speech, or social media, detecting mood, mental health status, and risks of harm or suicide (De Choudhury et al., 2023), as well as applications such as screening, diagnostics, treatment, and training (Viduani et al., 2023).

A 2019 scoping review of 53 studies assessing 41 mental health applications found that the most common use of AI companions in mental health was treating depression and anxiety (41.5%, 17/41), followed by training (29.3%, 12/41) and screening (24.4%, 10/41). In 92.5% of

studies, AI companions were rule-based (less prone to errors, easier to develop, but less flexible), and only 7.5% used machine learning (more flexible, less predictable). Only 4 of the 41 AI companions were implemented in developing countries, which suffer from a significantly larger shortage (approximately two orders of magnitude) of mental health providers compared to developed countries (Abd-Alrazaq et al., 2019).

According to Boucher et al. (2021), 41 mental health chatbots were developed in 2019 alone. Recent progress in generative AI and large language models (LLMs) holds promise for advancing the frontiers of digital mental health (De Choudhury et al., 2023). Mental health AI companions have gained significant popularity since the launch of ChatGPT in late 2022 (Balcombe, 2023).

Due to the difficulties of obtaining FDA approval, numerous mental health solutions are classified as wellness applications rather than regulated clinical devices (Martinez-Martin, 2020). Multiple non-validated AI companions have emerged, offering virtual and supportive friends who are always prepared to listen. These synthetic partners lack validation of safety and clinical effectiveness, which can encourage customers to use them for therapeutic purposes, potentially leading to dangerous interactions (De Freitas et al., 2023).

According to Zao-Sanders (2025), therapy and companionship are the most popular use cases of Generative AI in 2025, driven in part by the widespread accessibility, affordability, and judgment-free environment associated with Large Language Models, as well as people prioritizing survival over data privacy concerns.

Opportunities of AI companions in mental health

Improved healthcare outcomes:

AI companions offer scalable and accessible alternatives that complement traditional approaches (Abd-Alrazaq et al., 2019). A representative example is a randomized controlled trial that demonstrated the effectiveness of a conversational agent (Woebot) delivering cognitive behavioral therapy (CBT) to help reduce symptoms of stress and anxiety. Woebot integrated 15 of 16 recommendations from a previous study, including utilizing a CBT framework, offering relevant information, leveraging gamification, and including links to crisis support services. Participants averaged 22.2 years old and were primarily female and non-Hispanic. Participants were introduced to Woebot for the trial, informed that it might act like a person but couldn't fully understand, and were encouraged to call 911 in emergencies. Results were measured using questionnaires developed based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the Positive and Negative Affect Schedule (PANAS). The study concluded that users of Woebot had a significant decrease in depression symptoms compared to those in the control group (Fitzpatrick et al., 2017).

Other studies have also found that AI companions can reduce symptoms of depression and anxiety. For example, a four-week study of 4,517 Youper users found decreased anxiety and depression scores after two weeks and maintained decreased anxiety scores for the remaining two weeks (Mehta et al., 2021). An eight-week study of 101 Woebot for Substance Use Disorder users (75.2% female, 78.2% non-Hispanic white) found significant self-reported improvements in depression and anxiety symptoms (Prochaska et al., 2021b), and a recent systematic review found significant decreases in depression symptoms after chatbot interventions (Jabir et al., 2023).

AI companions have also helped reduce substance abuse cravings and increased resistance to urges (Prochaska et al., 2021b), offer clinical promise for attention deficit and

hyperactivity in adults (Balcombe, 2023), show potential for psychoeducation and self-adherence (Vaidyam et al., 2019), and have been used for screening, detection, and relapse prevention (Bowie-Dabreo et al., 2022).

AI companions offer the potential to promote disclosure (Viduani et al., 2023), provide relevant, personalized advice, resources, and recommendations, as well as stigma-free care (Balcombe, 2023; De Choudhury et al., 2023; Abbasian et al., 2024). For example, AI companions can help identify cognitive distortions and draw on CBT and evidence-based techniques to provide user-specific exercises to reframe negative thoughts. In contrast to rule-based chatbots, which struggle to adapt to user needs, Generative AI-powered companions can adjust their style and tone to meet the expectations of their users (Abd-Alrazaq et al., 2021; De Choudhury et al., 2023).

Accessibility and reach:

Given their ability to engage in human-like conversations, AI companions promise to increase access and reach. A 2023 study analyzing 6,245 user reviews of 10 mental health apps with built-in AI companions found that users appreciate human-like interactions, view AI companions as someone they can talk to without feeling judged and enjoy the ability to engage anytime and anywhere (Haque & Rubya, 2023).

AI companions also promise to increase access to underserved populations by reducing the gap between the supply and demand of mental health services (Balcombe, 2023). They seem to appeal to people with different demographic backgrounds (Haque & Rubya, 2023) and provide an avenue to treat people who do not like to disclose their feelings or seek help from another person (Mehta et al., 2021; Vaidyam et al., 2019; Abrams, 2021; Chin et al., 2023).

Since AI companions are available 24/7, they offer opportunities to individuals who struggle to access in-person support due to scheduling challenges resulting from work or childcare commitments. Since AI companions can be accessed remotely, they can offer alternatives to individuals in underserved areas (De Choudhury et al., 2023) who cannot afford professional care or may not even be aware they suffer from a mental health condition (De Freitas et al., 2023).

The ability of AI companions to maintain engagement over time also presents an opportunity to increase access for individuals not adequately supported by established models. For example, the acute support model provided by rape hotlines is misaligned with the long-term needs of sexual assault survivors who require someone to talk to and provide counseling regularly (Backe, 2018).

Chatbots offer the potential to increase access by reducing cost barriers. For example, they can provide economically accessible treatment, and moving low-level tasks from trained clinicians to chatbots can reduce healthcare costs by orders of magnitude (Mehta et al., 2021).

Reducing the burden on healthcare providers:

Chatbots can reduce the burden on providers by assisting with data collection (Viduani et al., 2023), providing culturally appropriate advice (De Choudhury et al., 2023), performing tasks that do not need a trained clinician (Mehta et al., 2021; Abbasian et al., 2024), helping prioritize human-to-human support for those in need (Boucher et al., 2021), and providing summaries to users (Abd-Alrazaq et al., 2021).

Chatbots offer scalable mental health solutions that are accessible 24/7. They alleviate the workload of healthcare providers by operating remotely without fatigue. Customers can seek support and receive real-time feedback after they finish work and attend to other life demands,

such as childcare, thereby eliminating limitations resulting from provider capacity or business hours (Bowie-DaBreo et al., 2022; Haque & Rubya, 2023; Abrams, 2021; Chin et al., 2023).

Chatbots can also reduce the burden on mental health providers by providing an easily scalable alternative to listening to users. For example, a study comparing the beneficial effects of emotional disclosure found that disclosing information to humans or chatbots yields similar outcomes, including reduced stress and anxiety, enhanced intimacy and closeness, and improved psychological outcomes, such as an enhanced sense of worth. In other words, people psychologically engage with chatbots in a manner similar to humans, with comparable benefits, regardless of the identity of the conversation partner (Ho et al., 2018).

Efficacy:

Multiple studies have raised concerns about the lack of sufficient data to prove the effectiveness of chatbot interventions in improving mental health and well-being.

In 2018, real-world data evaluation of the effectiveness of a conversational agent (Wysa) compared self-reported symptoms of depression of two cohorts of users, one with high app engagement and the other with low engagement. While the study concluded that chatbots offered promise in reducing depression symptoms, it also noted that the small sample size, the lack of information and clinical history about the users, and the absence of a controlled environment could have impacted the results (Inkster et al., 2018).

In 2019, a systematic review of 13 studies found that more rigorous experimental designs are needed to assess the efficacy of chatbots in mental health. The studies were conducted in the United States (6), the United Kingdom (5), Sweden (1), and Japan (1). Only four studies were randomized controlled trials with adequate scale, and three controlled studies failed to show better intervention outcomes vs. control groups. Although all studies reported reductions in

psychological distress, the ability of chatbots to enhance well-being remains uncertain. Safety was assessed in only one study (Gaffney et al., 2019).

Also in 2019, a scoping review of 53 studies on 41 mental health chatbots used for therapy, training, and screening concluded that additional research is needed to determine their effectiveness in mental health (Abd-Alrazaq et al., 2019).

In 2020, a systematic review of 31 studies on AI companions in healthcare found mixed-quality studies and mixed evidence regarding usability and effectiveness (Milne-Ives et al., 2020). In the same year, Martinez-Martin (2020) reported that 36% of apps from a sample of 1,435 claimed to be effective in addressing depression, but only 2% provided evidence to support this claim.

In 2021, Boucher et al. (2021) estimated that only 2% of a sample of 10,000 mental health applications were supported by evidence, highlighting that many interventions had suffered from high attrition rates and low usage. The same study argued that, since most of the mental health chatbot research has lacked control groups, it is unknown if the availability of chatbots may be acting as a digital placebo.

In 2022, Bowie-DaBreo et al. (2022) argued that few apps have demonstrated fidelity to the evidence-based treatments they claim to be based on. They highlighted the risks of inadequate ethical guidelines and reported user concerns regarding safety, transparency, and commercialization.

In 2023, a systematic review found that while chatbots significantly reduced *symptoms* of psychological distress (depression, anxiety, stress), there is no evidence that the same interventions have improved overall psychological *well-being* (positive and negative affect, mental resiliency, mental efficacy). Out of the 35 studies analyzed, 19 were quasi-experimental,

and the review found a potential overrepresentation of rule-based chatbots. The results also suggest that Generative AI-powered AI companions may outperform their rule-based counterparts and have greater potential to reduce symptoms of psychological distress. However, given study limitations, the quality of evidence is considered moderate (Li et al., 2023). Another study by Khawaja & Bélisle-Pipon (2023) raised concerns about deceptive marketing practices, inadequate design, and functional limitations that may lead users to overestimate the capabilities of chatbots.

Crisis management:

While chatbots can respond to clear crisis indicators, they often fail to detect subtle cues signaling risks such as self-harm or suicidal ideation. Chatbots also typically fail to offer access to a professional therapist or relevant crisis support information when needed (Boucher et al., 2021; De Choudhury et al., 2023; Martinengo et al., 2022).

Additionally, AI companions may struggle to respond to rapidly changing societal crises, such as mass shootings or pandemics, because their training data may not include the necessary information to accurately understand the context (De Choudhury et al., 2023).

Potential for harm:

Beyond being unable to handle crises, there is concern that chatbots can provide bad or unsafe advice (Blease et al., 2020; Akbar et al., 2020). A scoping review of 74 studies reported safety concerns in 80 healthcare applications, for example, suggestions that bipolar disorder is contagious and recommendations for patients to drink hard liquor before going to bed (Akbar et al., 2020).

Other studies have raised concerns about potential unmitigated risks associated with inadequate guidelines and evaluations of digital mental health apps (Bowie-DaBreo et al., 2022).

For example, the National Eating Disorder Association deactivated an AI companion that, despite being designed to provide evidence-based information, offered users harmful diet advice (De Choudhury et al., 2023).

Other authors have also reported the potential for harm. For example, safety concerns became highly publicized after the suicide of a man in his thirties was blamed on stressful climate change conversations with ELIZA —a generic conversational agent (Author's note: this is not the 1966 ELIZA referenced earlier) (Walker, 2023). A global study of SimSimi analyzed 152,387 English-language utterances discussing depression and sadness in three Western and five Eastern countries. The study found that users expect chatbots to provide advice on managing emotions and do so in a safe environment, highlighting the need for ongoing efforts to enhance chatbots' capabilities in offering emotional support (Chin et al., 2023). These findings are consistent with reports that people use ChatGPT for emotional support and guidance, sometimes in high-stakes moments such as considering suicide (De Choudhury et al., 2023). A study of AI companions found that approximately 3.2% of conversations with SimSimi and 4.9% with Cleverbot, a generic conversational agent, contained mental health terms. The same study tested five AI companions on their ability to respond to depression, suicide, self-injury, harming others, abuse, and rape crises. It concluded that these companion applications failed to provide adequate resources and, in some cases, even encouraged harm. For example, an AI companion replying 'don't u coward' to a user considering suicide (De Freitas et al., 2023). Non-validated AI companions are used by hundreds of millions of users; for example, SimSimi has 350 million users, Pi has over 100 million, Chai has 4 million, and Replika has more than 2 million users (De Freitas & Cohen, 2024). Moore et al (2025) maintain that —contrary to therapeutic guidelines and best practices— LLMs encourage delusions and stigmatize individuals with mental health

conditions. This is true, they argue, even for state-of-the-art models like GPT-4o and Llama 3.1-405b

Adapting to varying demographics:

An analysis of nine chatbots for self-management of depression conducted by Martinengo et al. (2022) reported limited personalization features in the group studied. Most chatbots collected minimal demographic data, with only a few inquiring about age and none seeking more detailed information, such as sex, medical history, or socioeconomic context. The lack of interest from AI companions in the identity characteristics of their users was also reported by Hiland (2018), who raised concerns about the prevalence of algorithmic discrimination after finding that none of three popular chatbots (Woebot, Wysa, and Joy) were interested in understanding her identity so they could adapt their responses and style accordingly. A study that interviewed 31 individuals (18 of whom identified as LGB+) reported that LGB+ individuals are drawn to using AI companions because they don't judge and are always ready to offer support. LGB+ individuals seek guidance on topics including identity affirmation, dealing with discrimination, and having difficult conversations such as 'coming out'. However, while LGB+ individuals find AI companions beneficial and provide a safe environment for intimate conversations, they also note that AI companions tend to miss the nuances of the challenges they face. For example, AI companions often assume environments are generally LGB+ friendly when that is not necessarily the case, or they can offer potentially harmful advice, such as 'quitting your job' without considering the implications (Ma et al., 2024). There is also a possibility that embedded bias in LLMs may negatively impact specific populations. For example, Hofmann et al. (2024) reported that multiple large language models perpetuate covert racism through dialect prejudice, particularly against African American English. This is an

important finding since, unlike overt racial bias, which can be partially mitigated with human feedback and increased model sizes, there are currently no good mitigation strategies for covert bias.

AI companions in mental health and the working alliance

Critics argue that although technology may help address the gap between mental health demand and supply, the lack of empathy of technological solutions will hinder the formation of a working alliance, reducing their effectiveness (Darcy et al., 2021). According to Blease et al. (2020), most psychiatrists are skeptical about the ability of technology to provide empathic care.

Integrating transparency into the design of AI companions and openly acknowledging limitations may be crucial for building trust and developing a working alliance (Darcy et al., 2021). In her doctoral dissertation, Hiland (2018) highlighted concerns about the lack of transparency of AI companions. For example, the chatbot Joy refused to answer if a human was looking at her data, and a (human) Wysa coach refused to discuss her credentials. However, research suggests that patients are interested in using mental health apps (Boucher et al., 2021; Blease et al., 2020) and may develop a working alliance with chatbots, as demonstrated by studies such as 1) Beatty et al. (2022), who used the WAI-SR to evaluate the working alliance with 1,205 users of Wysa. They reported a mean score of 3.64 for the initial assessment after five days (n = 1,205 users) and a mean score of 3.75 for a retest (n = 226 users). 2) Darcy et al. (2021) applied the WAI-SR to a sample of 36,070 Woebot users, reporting a mean total score of 3.36. Women reported the highest bond sub-score (mean 3.92), and the lowest bond level (mean 3.67) was reported by individuals who preferred not to disclose their gender. 3) A single-blind, three-arm, randomized controlled trial by He et al. (2022) utilized the WAQ (a version of the

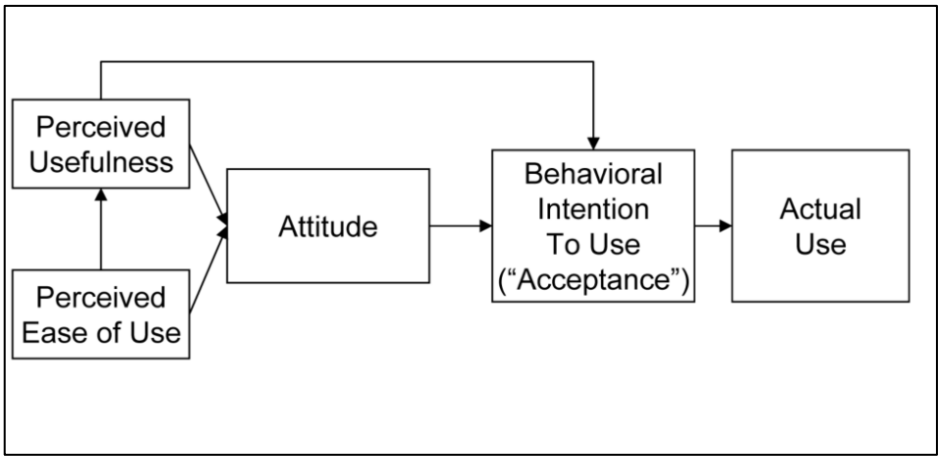
WAI-SR adapted to Chinese culture) and reported a mean working alliance score of 4.49 for users of XiaoE, a validated AI companion.

The Technology Acceptance Model

The Technology Acceptance Model (TAM) is widely utilized across various industries to understand and predict technology adoption and usage, with some assessments suggesting that the most basic model can frequently explain 30-40% of technology acceptance. In the basic TAM model, Behavioral Intention to Use (BI) precedes actual use. Since BI is considered a strong predictor of actual use (which can be difficult to measure), it is sometimes the primary outcome of interest in TAM studies. BI is influenced by the Attitude (ATT) towards using the technology, and ATT is, in turn, influenced by the Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). In the healthcare sector, an assessment of over 20 studies on clinicians using health IT concluded that while TAM can explain a significant share of user acceptance, it may require adjustments for improved applicability. For example, none of the studies included quality and safety of care in their definition of usefulness (Holden & Karsh, 2010).

Figure 4.

Technology Acceptance Model (TAM)



Note. Reprinted from The Technology Acceptance Model: Its Past and Its Future in Health Care (Holden & Karsh, 2010). Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2814963/

While Holden & Harsh (2010) examined the TAM on health IT from the perspective of clinicians, Park & Kim (2023) examined the TAM from the perspective of health IT users in South Korea, focusing on mental health chatbots. Their study of a sample of 278 participants (92% students, 8% faculty/staff; 50% female and 50% male) found that depression severity, perceived usefulness, and parasocial interactions positively influenced the intention to use a mental health chatbot, and perceived ease of use was positively associated with perceived usefulness. Notably, South Korea has a high prevalence of depression, with approximately 40 % of South Koreans being at risk of negative physical, psychological, and social consequences.

Gaps in the literature

There is a need for further research on the factors that impact the ability of AI companions to establish a working alliance. Out of 35 studies in a systematic review and meta-analysis of AI mental health, only four reported working alliance scores, and only one compared

a clinical with a non-clinical AI companion (Li et al., 2023; Li et al., 2023b). From the studies reviewed in the literature search, only one (Darcy et al., 2021) reported alliance subscores by gender, and all studies except one focused on a single AI companion. A recent study of a fully generative-AI-powered AI Companion (Therabot) reported a mean WAI-SR score of 3.59, with mean subscores of 3.47 for task, 3.59 for Goal, and 3.71 for Bond (Heinz et al., 2025).

To this author's knowledge, no studies have systematically assessed potential factors impacting the ability of AI companions as a class (i.e., multiple companions of different types) to build a working alliance. This is surprising since the risks of non-validated AI companions have been raised by researchers such as Chin et al. (2023) and De Freitas et al. (2023), and since the use of non-validated AI companions has also received considerable media attention including the reported death by suicide of an eco-anxious man that, after a few weeks of intense conversations with talking the chatbot Eliza, decided to take his own life (Walker, 2023); and a report from Tidy (2024) that Character.ai's Psychologist chatbot had received 78 million messages in the year following its development by a 30-year-old psychology student. Furthermore, Tidy (2024) reported that there were hundreds of other Character.ai chatbots with references to therapy in their names, and most Character.ai users were teenagers or young adults. At the time the proposal for this study was prepared, Character.ai's Psychologist had recorded 196.3 million chats.

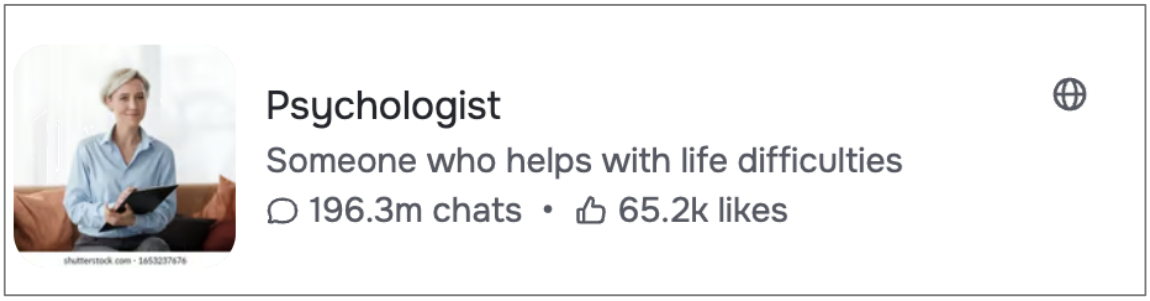
Table 2.

Literature review findings. Studies on the working alliance in AI companions.

Reference	AI Companions Assessed (Mean Working Alliance Scores)	Studied factors impacting Working Alliance Scores?
Beatty et al. (2022)	Wysa (WAI-SR M=3.64 initial score n=1,205; M=3.75 retest, n=226)	No
Darcy et al. (2021)	Woebot (WAI-SR M=3.36, n=36,070)	Limited: Gender (women): bond sub score: M=3.92 Gender (preferred not to answer): bond sub score: M=3.67
Fitzsimmons-Craft & Jacobson. (2024).	Therabot 'Excellent therapeutic alliance'. No scores reported	No
He et al. (2022).	XiaoE (Validated, WAQ M=4.49) Xiaoai (Non-validated, WAQ M=4.22)	No Reported separate scores for a pair of validated and non-validated chatbots.
Heinz et al. (2025)	Therabot (validated, WAI-SR M=3.59; Task M=3.47, Goal M=3.59, Bond M=3.71.)	No
Karkosz et al. (2024)	Fido (WAI-SR M=2.71)	No Polish language WAI-SR mean of 3.24 on bond subscale. Younger participants formed a stronger bond than older participants (3.59 vs. 2.59)
Liu et al. (2022).	XiaoNan. (Working Alliance scores not reported. Chatbot performed better than bibliotherapy group with t=7.29, P< 0.01)	No
Prochaska et al. (2021a)	Woebot (WAI-SR M=3.7)	No
Prochaska et al. (2021b)	Woebot (WAI-SR M=3.4)	No

Figure 5.

Character.ai psychologist: a non-validated AI companion



Note. Screenshot of Psychologist, from search results for "psychologist" at Character.AI. (https://character.ai/search?q=psychologist) Retrieved February 8, 2025

Figure 6.

Youper: a clinically-validated AI companion

Note. Screenshot of Youper homepage, from Youper (https://www.youper.ai/). Retrieved October 31, 2024.

METHODOLOGY

Research design

This study examined factors that may influence the strength of the working alliance when utilizing AI companions. These factors included sex (male, female), sexual orientation (heterosexual, other), race/ethnicity (White, People of Color), anxiety/depression intensity (above cutoff, below cutoff), AI companion type (clinically validated and non-validated), as well as themes identified through qualitative research.

Given that the working alliance is more meaningful in situations where individuals and therapists are working on shared objectives, the study focused on persons with depression and anxiety symptoms. Depression and anxiety were selected since they are the two most common mental health conditions and are frequently targeted by AI companions.

This study employed a mixed-methods design to integrate quantitative and qualitative data collected through a single cross-sectional survey instrument. Quantitative data from the WAI-SR were analyzed separately from the open-ended qualitative responses, which were thematically coded. Integration was achieved through narrative, utilizing a combination of the contiguous and weaving approaches described by Fetters et al. (2013). The findings are reported in different sections but are also discussed together within shared conceptual themes.

Variables

The selection rationale and value definitions for the study's key variables—sex, sexual orientation, race/ethnicity, anxiety/depression intensity, companion type, and working alliance—are outlined below.

Sex (values: male, female). Sex was selected based on the literature findings reported in section 0 - Depression, anxiety, and identity. The values were determined from figures derived from a report by Stacey (2024), which shows that 50.6% of the U.S. population identifies as female (cisgender), 48.8% as male (cisgender), and 0.43% as other (0.37% transgender, 0.15% non-confirming).

Sexual Orientation (values: heterosexual, other). Sexual Orientation was selected based on the literature findings reported in section 0 - Depression, anxiety, and identity. Values were chosen from figures derived from a report by Stacey (2024) showing that 94.27% of the US population identifies as heterosexual (straight) and 5.7% as other (2.78% Bisexual, 1.91% Gay or Lesbian, 1.04% something else).

Race/Ethnicity (values: White, Person of Color): Race/Ethnicity was selected based on the literature findings reported in section 0. Values were chosen from figures derived from the U.S. Census Bureau (2020), which indicate that 57.8% are White, non-Hispanic, and 42.2% are Persons of Color (18.7% Hispanic or Latino, 12.1% Black, non-Hispanic, and 11.4% Asian, American Indian, Hawaiian, two or more races, and others).

Anxiety/Depression intensity (above cutoff, below cutoff). Values were selected based on the literature cutoff scores ≥ 2 for the PHQ-2 and ≥ 3 for the GAD-2.

Companion Type (values: clinically validated, non-validated). This factor was selected due to the research gap regarding the ability of non-validated AI companions to establish a working alliance. The share of users for each type was previously unknown.

Working Alliance (dependent variable, values: 0-5). The working alliance was selected because the literature consistently shows that it is a key predictor of patient outcomes, independent of the therapeutic modality (Hatcher & Gillaspy, 2006; Norcross & Lambert, 2018;

Horvath & Luborsky, 1993; Flückiger et al., 2018). The values for the working alliance were measured using the overall mean for the WAI-SR.

Assessment instruments

This study used three validated instruments to assess symptoms of anxiety, symptoms of depression, and the strength of the working alliance. Each was selected for its established validity, reliability, and concise format.

The Generalized Anxiety Disorder-2 (GAD-2)

Potential anxiety was assessed using the GAD-2 (Generalized Anxiety Disorder-2) instrument, a concise screening tool with strong psychometric properties. It consists of two questions that ask individuals to rate the frequency of key anxiety symptoms (e.g., feeling nervous, worrying) over the past two weeks. Responses are scored on a scale from 0 (not at all) to 3 (nearly every day), with a total score ranging from 0 to 6.

The Patient Health Questionnaire-2 (PHQ-2)

Potential depression was assessed using the PHQ-2 (Patient Health Questionnaire-2), a concise screening tool with strong psychometrics. The questionnaire consists of two questions to evaluate depressed mood and anhedonia frequency over the past two weeks. Responses are scored from 0 (not at all) to 3 (nearly every day), with a total score range of 0 to 6.

The Working Alliance Inventory Short Revised (WAI-SR)

The strength of the working alliance was assessed using the Working Alliance Inventory Short Revised (WAI-SR) instrument. The WAI-SR was selected due to its demonstrated validity

and reliability across diverse populations, as well as its previous use in measuring the strength of the working alliance between patients and AI companions. The WAI-SR consists of 12 items, each rated using a 5-point Likert scale.

Permissions statement

The Patient Health Questionnaire-2 (PHQ-2) and the Generalized Anxiety Disorder-2 (GAD-2) are distributed under public-domain terms and do not require permission to use. The Working Alliance Inventory-Short Revised (WAI-SR) is copyrighted by the Society for Psychotherapy Research (SPR). SPR granted written authorization to use the WAI-SR in this study, and the license letter is retained in the project files.

Recruitment

This section outlines the recruitment procedures, inclusion criteria, compensation, and consent process used to identify and enroll eligible participants for the study.

Recruitment approach

Participants were recruited via Prolific.com. Filters were used to present the study to adult individuals living in the US who reported using AI chatbots and self-identified as experiencing anxiety or depression. The survey was available to any Prolific members who met the above criteria and provided consent to participate. Prolific handled all payments. The researcher did not have any access to participants' personally identifiable information.

Participation was voluntary, and participants received a nominal compensation of $6 for their time and effort. The payment amount was determined in accordance with Prolific’s fair payment standards.

Recruitment took place in two stages. Based on a theoretical calculation and assuming balanced participation, an initial survey with 213 spaces was launched to gain a preliminary understanding of the data. The initial survey was open to users of validated and non-validated companions. The initial survey yielded 194 valid answers, with 183 (94.3%) for non-validated companions and 11 (5.7%) for validated companions. Since the absolute number of responses for validated companions was low, a second survey with 81 additional spaces was launched to gather additional responses from users of validated companions only. Of the 253 valid responses in the final combined sample, 183 (72.3%) were for non-validated companions and 70 (27.7%) for validated companions.

Recruitment materials

Prospective participants received an email from Prolific and saw the study in their dashboard see Figure A2.

Consent

Participants responding to the invitations were directed to a Qualtrics survey that supports desktop, tablet, and mobile devices. The initial page of the survey contained a consent form based on the template from Purdue's Exempt Research Study Information Sheet for Participants. The consent form disclosed the study's purpose, risks, benefits, data collection methods, and participant rights, enabling participants to make an informed decision. Prospective participants were asked to provide consent before proceeding.

Data collection

This section describes the procedures and considerations involved in the survey's design and deployment, including the structure, content, and format of questions used to collect both quantitative and qualitative data.

Participants who provided consent were taken to the next section of the Qualtrics survey, which included verification questions followed by questions about the AI companion used, as well as questions for the PHQ-2, GAD-2, and WAI-SR instruments. The survey also included three open-ended questions inspired by the Technology Acceptance Model (TAM) to gather qualitative data. The PHQ-2, GAD-2, and WAI-SR were selected based on their brevity and psychometric properties. Questions about race/ethnicity, sex, and sexual orientation status were worded using inclusive language and choices, including a 'prefer not to answer' option, and were asked at the end of the survey. Questions about the AI companion were multiple choice, utilizing a list of popular chatbots and companions. An option for 'other AI companion' with a free text entry allowed users to provide values not included on the original list. Questions for the PHQ-2 and GAD-2 were verbatim, and questions for the WAI-SR replaced 'therapist' and 'therapy' with the AI companion used.

Qualtrics filters to prevent abuse were activated. IP addresses were not stored in the research database or linked to the participants' responses.

The survey was online until the target number of participants was reached.

Survey design considerations

Location of demographics questions

While some authors, such as Fink (2003), maintain that including demographic questions at the end of a survey is preferred due to reasons such as building rapport and interest, there doesn't seem to be strong empirical evidence supporting this claim. According to Giles & Feild (1978), the location of demographic questions on a survey does not influence response bias. Teclaw et al. (2012) claim that locating demographic questions at the start of a survey increases demographic response rates without affecting the mean of non-demographic questions. They conclude that the placement of demographic questions does not significantly impact non-demographic response patterns, allowing for flexibility in survey design.

Numerical vs. categorical

According to Giles & Feild (1978), the format of demographic questions may influence bias. For example, when respondents are required to provide numerical answers rather than selecting from categorical options, response bias may occur, especially for sensitive survey items.

Open vs. closed

Fink (2003) argues that open-ended questions are more challenging for participants and more complex to code and analyze than closed-ended questions. However, open-ended questions are helpful for exploration when the topic is not well understood, in situations where categories can reduce the sensitivity of questions, and when there is reason to believe that the quality of the data will be better than with closed-ended questions.

Data pre-processing

Several preprocessing steps were carried out to ensure the integrity and usability of the dataset, including removing invalid records, standardizing responses, and consolidating participant data.

Data cleansing

The data collected required cleaning to exclude invalid records. Qualtrics filters were used to identify likely bot activity, failed reCAPTCHA, and potential fraud. Additionally, a manual review was conducted to remove responses where the answers were off-topic or appeared to be fabricated from web searches. Responses that failed verification checks were also excluded. Responses where participants opted not to disclose their sex, sexual orientation, or race/ethnicity were excluded to preserve the integrity of factor analyses. Out of an initial count of 350 responses, 253 (72.3%) were deemed valid.

Table 3.

Data cleansing summary

Initial Count		**350**
Data Cleaning		
	Likely bots: reCAPTCHA (< = 0.5)	12
	Likely bots: fraud scores (≥ 30)	10
	Off-topic or likely fabricated answers	2
	No consents	7
	Failed verification question	56
	Sex listed as 'other' or 'prefer not to say'	5
	Sexual orientation is 'prefer not to say'	3
	Ethnicity is 'prefer not to say'	2
	Total removals	**97**
Records Left		**253**

Data preparation

To reduce discomfort, culturally sensitive wording and options were included in the survey, even though not all options fell within the scope of the study. In other words, some data elements were captured with a higher level of granularity than those used in the final report. Specifically, the following data manipulation was performed in preparation for the analysis: 1) race/ethnicity values of Hispanic, Black, and Other were categorized as People of Color. 2) Responses for 'Other' AI companions that were not on the original list (e.g., Gemini, Grok) were tagged by the researcher as clinically validated or non-validated. 3) sexual orientation of gay, lesbian, bisexual, and other was reported as 'other'. 4) 'prefer not to answer' responses were removed from the final reports of the relevant categories. 5) 'other' responses for Sex were removed from the final report since only male and female were in scope. 6) Participants with

scores < 2 for the GAD-2 (depression) and < 3 for the PHQ-2 (anxiety) were tagged as "below cutoff", and the rest were tagged as "above cutoff" for the anxiety/depression intensity analysis.

The data manipulation described above was guided by tradeoffs inherent in statistical research. By consolidating categories—such as merging race/ethnicity and sexual orientation identities into broader groups—the analysis aimed to strike a balance between complexity, insight generation, and statistical validity.

Sample details

The study included two samples: an initial sample was used to assess the 'naturally occurring' proportion of validated versus non-validated AI companions, with the vast majority of participants (n=183, 94.3%) using non-validated AI companions and only 11 (5.7%) using validated companions. Since the initial sample lacked sufficient participation for validated companions, a second sample was collected, focusing on validated companions. The combined samples consist of 253 individuals, with a slightly higher proportion of females (52.6%) than males (47.4%). Most participants identified as heterosexual (n= 194, 76.7%), while 59 (23.3%) identified with other sexual orientations. In terms of race/ethnicity, 173 (68.4%) identified as White, and 80 (31.6%) identified as people of color (15.4% Black or African American, 6.3% Hispanic or Latino, and 9.9% as Other). Within the final sample, ChatGPT dominated non-validated companions (n =183) with 74.9%, followed by SnapChat (6.0%), Character.ai (4.9%), Claude (3.8%), Replika (2.7%), and Pi, Gemini, Meta, Deep Seek, Flourish, Grok, and Perplexity, each with smaller proportions ranging from 2.2% to 0.5%. Among validated companions (n = 70), Woebot was the most frequently used, representing 47.1% of that group, followed by Wysa at 32.9%, Youper at 18.6%, and Tess at 1.4%.

Table 4.

Population sample by sex

	Count	% of Total
Female	133	52.6%
Male	120	47.4%
Grand Total	**253**	**100.0%**

Table 5.

Population sample by sexual orientation

	Count	% of Total
Heterosexual	**194**	**76.7%**
Other	**59**	**23.3%**
Lesbian / Gay	27	10.7%
Something else	32	12.6%
Grand Total	**253**	**100.0%**

Table 6.

Population sample by race/ethnicity

	Count	% of Total
White	**173**	**68.4%**
People of Color	**80**	**31.6%**
Black or African American	39	15.4%
Hispanic or Latino	16	6.3%
Other	25	9.9%
Grand Total	**253**	**100.0%**

Table 7.

Population sample by companion type

	Initial Sample			Final Combined Sample		
	Count	**% of Total**	**% of Type**	**Count**	**% of Total**	**% of Type**
Non-validated	**183**	**94.3%**	**100.0%**	**183**	**72.3%**	**100.0%**
ChatGPT	137	70.6%	74.9%	137	54.2%	74.9%
SnapChat	11	5.7%	6.0%	11	4.3%	6.0%
Character.ai	9	4.6%	4.9%	9	3.6%	4.9%
Claude	7	3.6%	3.8%	7	2.8%	3.8%
Replika	5	2.6%	2.7%	5	2.0%	2.7%
Pi	4	2.1%	2.2%	4	1.6%	2.2%
Gemini	3	1.5%	1.6%	3	1.2%	1.6%
Meta	3	1.5%	1.6%	3	1.2%	1.6%
Deep Seek	1	0.5%	0.5%	1	0.4%	0.5%
Flourish	1	0.5%	0.5%	1	0.4%	0.5%
Grok	1	0.5%	0.5%	1	0.4%	0.5%
Perplexity	1	0.5%	0.5%	1	0.4%	0.5%
Validated	**11**	**5.7%**	**100.0%**	**70**	**27.7%**	**100.0%**
Woebot	5	2.6%	45.5%	33	13.0%	47.1%
Wysa	4	2.1%	36.4%	23	9.1%	32.9%
Youper	1	0.5%	9.1%	13	5.1%	18.6%
Tess	1	0.5%	9.1%	1	0.4%	1.4%
Grand Total	**194**	**100.0%**		**253**	**100.0%**	**100.0%**

Data analysis

Data analysis involved both quantitative and qualitative methods to evaluate working alliance scores and explore participant perceptions. Statistical comparisons addressed differences across AI companion types and user characteristics, while thematic analysis provided deeper insights into user experiences and factors influencing the development of a working alliance.

Statistical methods

The dependent variable, the strength of the working alliance, was measured using the Working Alliance Inventory Short-Revised (WAI-SR) instrument. Non-parametric tests were used because 1) the actual data frequently did not have a normal distribution, and 2) variances were frequently not equal. Specifically, the Wilcoxon Rank Sum Test, also known as the Mann-Whitney U Test, was selected. As a non-parametric method that compares median ranks, the Wilcoxon rank-sum test offers a more robust alternative under these conditions, as it does not require normality or homogeneity of variance, is less susceptible to distributional irregularities, and applies to comparisons between independent groups. According to Fritz et al. (2012), effect size values of ≥ 0.1 were considered small, ≥ 0.3 medium, and ≥ 0.5 large.

Sub-question 1 (WAI-SR differences by AI Companion type): AI Companion type: Total WAI-SR Scores and goal, task, and bond subscores for Validated, Non-validated AI companions

Sub-question 2 (WAI-SR differences by sex, sexual orientation, race/ethnicity, anxiety/depression):

- (AI Companion Type) X (Sex): WAI-SR Scores for Validated/Male, Validated/Female, Non-validated/Male, Non-validated/Female

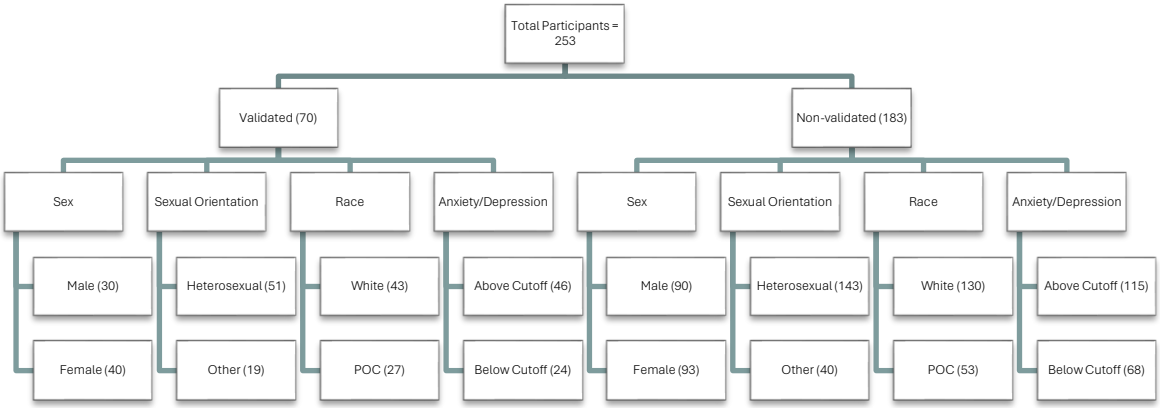
- (AI Companion Type) X (Sexual Orientation): WAI-SR Scores for Validated/Heterosexual, Validated/Other, Non-validated/Heterosexual, Non-validated/Other
- (AI Companion Type) X (Race/Ethnicity): WAI-SR Scores for Validated/White, Validated/POC, Non-validated/White, Non-validated/POC
- (AI Companion Type) X (Anxiety/Depression): WAI-SR Scores for Validated/Above Cutoff, Validated/Below Cutoff, Non-validated/Above Cutoff, Non-validated/Below Cutoff

To form the groups above, participants were separated by their preferred companion type. For example, a participant who used Wysa as their preferred companion was part of the validated companion group. In contrast, a participant who used Character.ai was part of the non-validated companion group.

All participants for each AI companion group were included in the separate analysis for sex, sexual orientation, race/ethnicity, and anxiety/depression intensity. For example, a gay black man using ChatGPT was included in four separate analyses to determine how sex, sexual orientation, race/ethnicity, and anxiety/depression intensity impacted the working alliance for non-validated AI companions.

Figure 7.

Experiment design: WAI-SR comparison groups.



The researcher reviewed all outlier responses and found no reason to remove them from the final analysis.

All statistical analyses and related graphics included in this report were produced with RStudio (version 2024.12.1+5634.3.1), and the code is included in the project files.

Thematic analysis

To gain insights beyond the quantitative metrics, a thematic analysis was conducted on responses to three open-ended questions inspired by the Technology Adoption Model. The focus was on how participants described the AI companion's usefulness, ease of use, and their intention to continue using it. Responses to the three questions were consolidated by participant to preserve contextual links between answers. A hybrid approach was used to develop themes, combining inductive coding (based on the data) with a priori insights from the literature. The process loosely followed Braun and Clarke's framework: the researcher read through the data

multiple times, highlighted key words and ideas, and generated initial codes. These codes were then categorized into larger themes, which were refined through an iterative process.

All coding and theme development were conducted manually using Microsoft Excel, and the detailed data is included in the project files.

Reliability

The Working Alliance Inventory-Short Revised (WAI-SR) is an established instrument with high internal consistency and reliability. The PHQ-2 and GAD-2 have also shown high internal consistency and reliability, supported by strong correlations with PHQ-9 and GAD-7, when using cutoff scores $\geq$2 for the PHQ-2 and $\geq$3 for the GAD-2.

Validity

The variety of user groups resulting from the various permutations of sex, sexual orientation, race/ethnicity, and anxiety/depression intensity reduced the potential for systematic bias.

The validity of the PHQ-2 and GAD-2 instruments for screening symptoms of depression and anxiety across populations has been established through multiple studies, which have demonstrated strong correlations with longer instruments such as the PHQ-9 and GAD-7.

While initially developed for human-human therapeutic settings, previous studies have successfully utilized the WAI-SR in chatbot contexts.

The primary threat to validity was the challenge of recruiting enough participants to achieve adequate power across all analysis categories. A post-hoc analysis of the final sample was conducted using a two-sample independent t-test approximation. This analysis assumed a medium effect size (Cohen's $d = 0.5$), a significance level of $\alpha = 0.05$, and a two-sided

alternative hypothesis. Results suggest that subgroups of validated AI companions might have been underpowered, indicating a risk of Type II errors.

An additional consideration is that potential differences in working alliance strengths may have been obscured by self-selection bias or other factors such as socioeconomic or technology acceptance biases resulting from recruiting participants through Prolific. Therefore, future studies should consider a more structured design, such as a double-blind clinical trial with larger and more balanced samples.

Responsible conduct of research, ethics, and compliance

This study was conducted in accordance with ethical research standards and institutional requirements to ensure participant safety, privacy, and data integrity. Oversight, consent procedures, data handling practices, and risk mitigation strategies were all implemented to align with relevant regulations and Purdue University protocols.

Category 2 exemption

The Purdue University Human Research Protection Program (HRPP) determined that this study qualified for a Category 2 exemption from IRB review, as it only involved interactions with adults through survey procedures. The study IRB number is 2024-1762.

Data security

All the data was captured using Purdue's Qualtrics, which is certified by security frameworks including ISO 27001, FedRAMP, SOC 2, and HITRUST. This study was not subject to HIPAA since the researchers are neither covered entities nor business associates, and the data were de-identified. Out of an abundance of caution, data sharing between the researcher and the

dissertation advisor was conducted using a Purdue-provided REED Folder - Level 3 Data. This managed storage solution is built on top of the Box.com cloud platform, designed for research projects that require compliance with regulations or heightened security, and is approved for storing HIPAA-aligned data.

Risk mitigation

This study only involved individuals already using AI companions for therapeutic or mental health support. Therefore, participants were deemed to encounter only minimal risks (no more than those experienced in their daily lives), primarily associated with potential emotional discomfort from answering questions about their mental health, and providing demographic information, such as sex, sexual orientation, race/ethnicity, and answering questions about how they interact with AI companions.

To mitigate risks and ensure ethical, compliant, and responsible conduct of research, the following measures were implemented:

Voluntary participation and informed consent

Participants were asked to consent to the survey application before it was administered. The consent page informed participants about the study's purpose, procedures, risks, benefits, the voluntary nature of their participation, their rights, and the option to withdraw at any time.

Participants were remunerated following Prolific’s guidance. Potential participants were also informed that the study was conducted for research purposes, did not constitute therapy, and was not intended as a tool for crisis management.

Minimizing emotional distress

Participants were informed that they could abandon the questionnaire at any time. Questions related to sex, sexual orientation, and race/ethnicity were worded using culturally sensitive and inclusive language and included a 'prefer not to say' option. The number of questions was kept to the minimum necessary, and links to crisis management hotlines were prominently displayed.

Privacy and security

The study was anonymous and did not collect any information that could be used to identify participants. To note, while Prolific IDs are considered indirect identifiers, they were not captured by the study. The survey only collected data elements necessary to conduct the study. In addition, the data collected was not linked with other sources. All collected data was encrypted in transit and at rest, with access restricted to the researcher and supervisor.

RESULTS

This section presents the study's findings, organized by quantitative and qualitative analyses. Results focus on differences in working alliance scores between validated and non-validated AI companions, patterns across demographic groups, and user-reported experiences with AI companions.

Quantitative results summary

This subsection summarizes key statistical outcomes comparing working alliance scores across AI companion types and user subgroups. Non-parametric tests (specifically the Wilcoxon rank-sum) were used due to non-normal distributions and different variances. Given that most WAI-SR studies report means, the results of this study include both the Mean (M) and Median (Md); p and r values are relevant to the Md analysis.

Clinically validated companions produced significantly higher total WAI-SR working alliances than non-validated companions (M = 3.53 vs. 2.87; Md = 3.62 vs. 2.92, $p < .001$, $r = .33$), as well as for goal (M=3.71 vs. 2.97, Md = 4.00 vs. 3.00, $p < .001$, $r = 0.33$), task (M= 3.41 vs. 3.00, Md = 3.50 vs. 3.00, $p = .0012$, $r = .20$), and bond (M= 3.48 vs. 2.63, Md = 3.50 vs. 2.50, $p < .001$, $r = .31$) sub-scores. In all cases, validated AI companions outperformed non-validated AI companions.

No significant differences in WAI-SR scores were found for sex, sexual orientation, race/ethnicity, or anxiety/depression intensity within companion types.

In our sample, validated AI companions had much lower adoption than non- validated companions: 95% CI [0.032, 0.099].

Adoption levels

Before examining score differences, the distribution of validated versus non-validated companion use was analyzed to understand the sample composition. The initial sample of 194 AI companions showed 11 (5.67%) with clinical validation and 183 (94.33%) without clinical validation. This yields a 95% confidence interval of 3.2% to 9.9% adoption of validated AI companions, estimated using the Wilson method.

Validated vs. non-validated companions

This subsection compares WAI-SR total and subscale scores between users of validated and non-validated AI companions, highlighting statistically significant differences across all dimensions of the working alliance.

Table 8.

WAI-SR results: Validated vs. non-validated AI companions.

	Count	M / Md WAI-SR Goal	M / Md WAI-SR Task	M / Md WAI-SR Bond	M / Md WAI-SR Total
Companion Type		($p < .001$, $r=0.33$)	($p = .001$, $r=0.20$)	($p < .001$, $r=0.31$)	($p < .001$, $r=0.33$)
Validated	70	3.71 / 4.00	3.41 / 3.50	3.48 / 3.50	3.53 / 3.62
Non-validated	183	2.97 / 3.00	3.00 / 3.00	2.63 / 2.50	2.87 / 2.92

Validated companions

The following results examine whether WAI-SR scores among users of validated AI companions differed by sex, sexual orientation, race/ethnicity, or anxiety/depression intensity.

Table 9.

WAI-SR results validated companions by demographics.

	Count	M / Md WAI-SR Total
Sex (p = 0.1335, r=0.1335)		
Male	30	3.64 / 3.75
Female	40	3.45 / 3.50
Sexual Orientation (p = 0.2419, r=0.1399)		
Heterosexual	51	3.64 / 3.58
Other	19	3.25 / 3.75
Ethnicity/Race (p = 0.7036, r=0.0455)		
White	43	3.50 / 3.67
POC	27	3.59 / 3.58
Anxiety/Depression (p = 0.442, r=0.0918)		
Above Cutoff	46	3.50 / 3.58
Below Cutoff	24	3.59 / 3.75

Non-validated companions

Similar to the previous subsection, the following results examine whether WAI-SR scores among users of non-validated AI companions differed by sex, sexual orientation, race/ethnicity, or anxiety/depression intensity.

Table 10.

WAI-SR results non-validated companions by demographics.

	Count	M / Md WAI-SR Total
Sex (p = 0.6631, r=0.0000)		
Male	90	2.83 / 2.92
Female	93	2.90 / 2.92
Sexual Orientation (p = 0.2034, r=0.0940)		
Heterosexual	143	2.91 / 2.92
Other	40	2.70 / 2.71
Ethnicity/Race (p = 0.3541, r=0.0685)		
White	130	2.83 / 2.92
POC	53	2.95 / 3.00
Anxiety/Depression (p = 0.1394, r=0.1093)		
Above Cutoff	115	2.95 / 3.00
Below Cutoff	68	2.74 / 2.79

Quantitative analysis details

This section provides detailed statistical test results and visualizations, including data distribution checks, effect sizes, and summaries of group comparisons. The section includes score comparisons for the total WAI-SR, its subcomponents (goal, task, bond), as well as additional details by companion type.

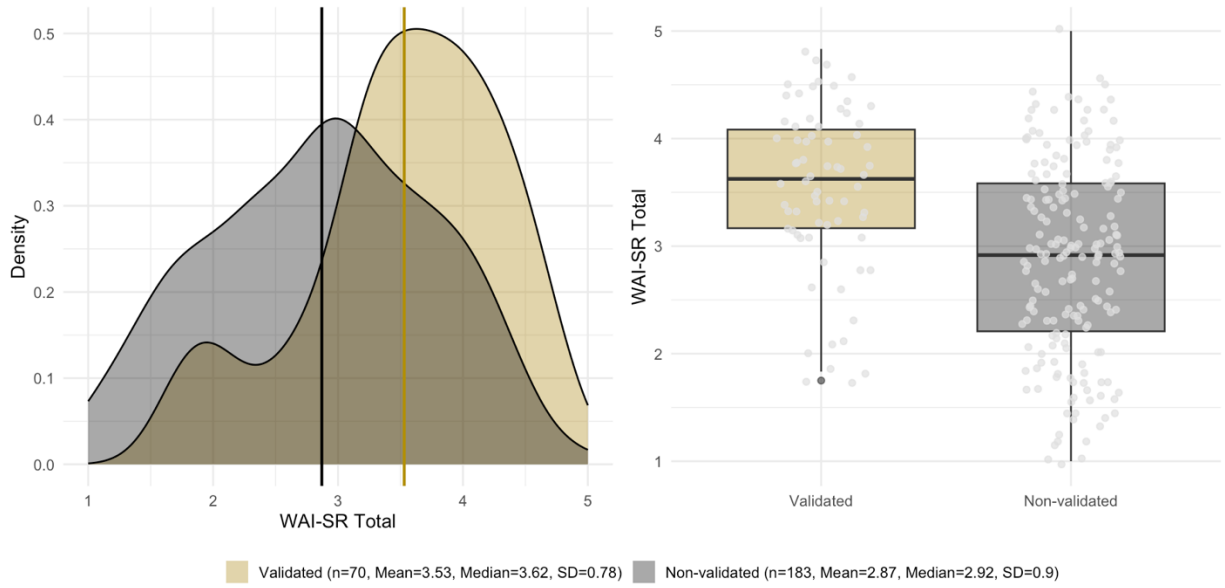
Note: In density plots, vertical lines represent the mean; in boxplots, horizontal lines indicate the median.

Validated vs. non-validated companions

WAI-SR total by companion type

Figure 8.

WAI-SR Total by companion type.



Data distribution

Shapiro-Wilk Test for Validated: p = 0.0063. Data is not normally distributed

Shapiro-Wilk Test for Non-validated: p = 0.0212. Data is not normally distributed

Levene's Test: p = 0.0828. Variances are equal.

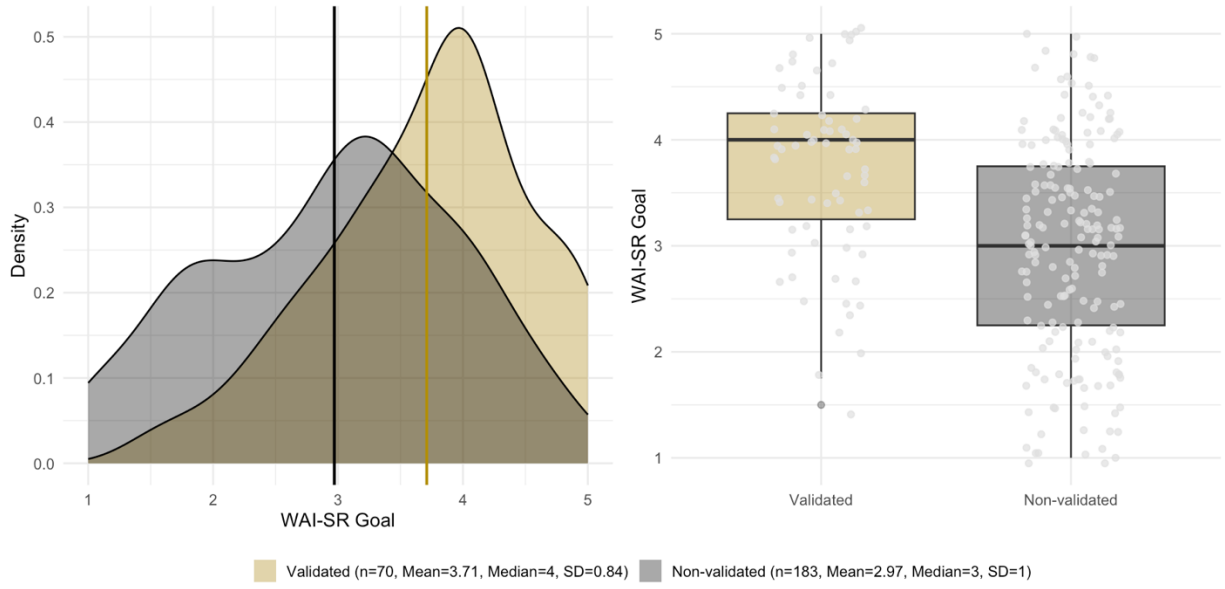
Wilcoxon rank sum test

Median Non-validated	Median Validated	W	p	Effect Size (r)		Result
2.92	3.62	3658	< 0.001	0.33	medium	reject null

WAI-SR goal by companion type

Figure 9.

WAI-SR Goal by companion type.



Data distribution

Shapiro-Wilk Test for Validated: p = 0.0234. Data is not normally distributed

Shapiro-Wilk Test for Non-validated: p = 0.0011. Data is not normally distributed

Levene's Test: p = 0.0476. Variances are not equal.

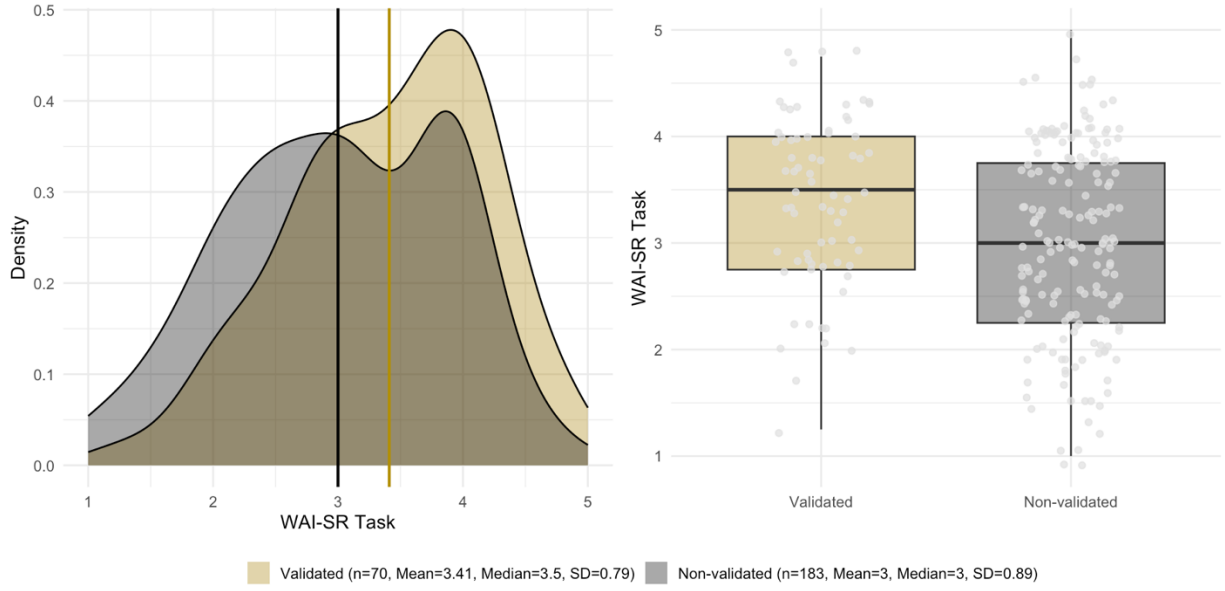
Wilcoxon rank sum test

Median Non-validated	Median Validated	W	p	Effect Size (r)		Result
3.00	4.00	3699.5	< 0.001	0.33	medium	reject null

WAI-SR task by companion type

Figure 10.

WAI-SR Task by companion type.



Data distribution

Shapiro-Wilk Test for Validated: p = 0.0298. Data is not normally distributed

Shapiro-Wilk Test for Non-validated: p = < .001. Data is not normally distributed.

Levene's Test: p = 0.1933. Variances are equal.

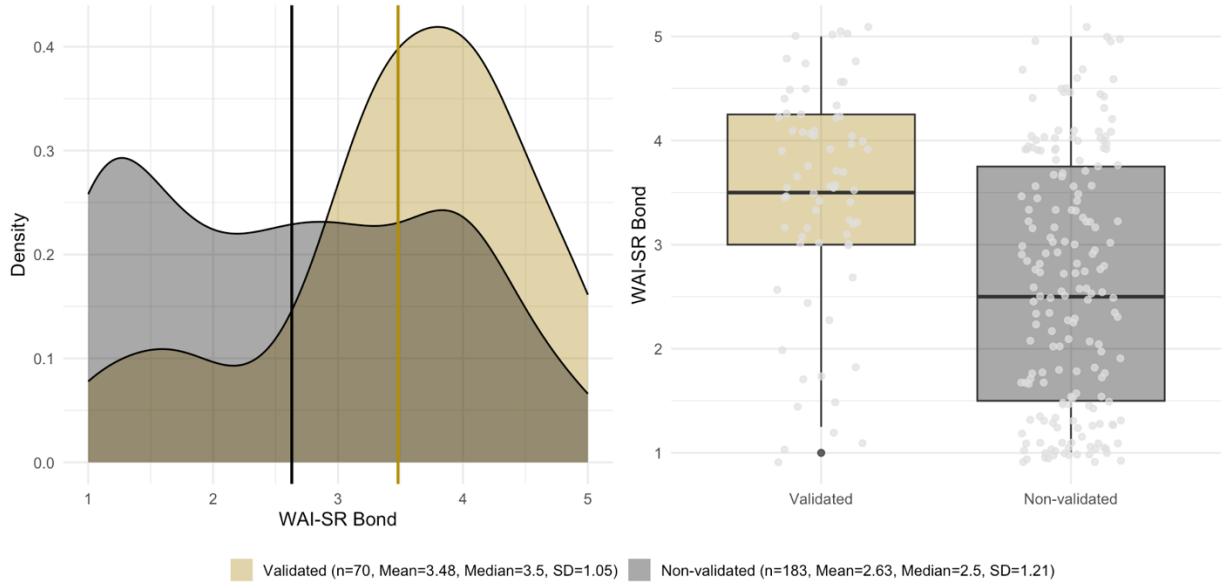
Wilcoxon rank sum test

Median Non-validated	Median Validated	W	p	Effect Size (r)		Result
3.00	3.50	4727.5	0.0012	0.20	small	reject null

WAI-SR bond by companion type

Figure 11.

WAI-SR Bond by companion type.



Data distribution

Shapiro-Wilk Test for Validated: p < .001. Data is not normally distributed.

Shapiro-Wilk Test for Non-validated: p = 0. Data is not normally distributed

Levene's Test: p = 0.0047. Variances are not equal.

Wilcoxon rank sum test

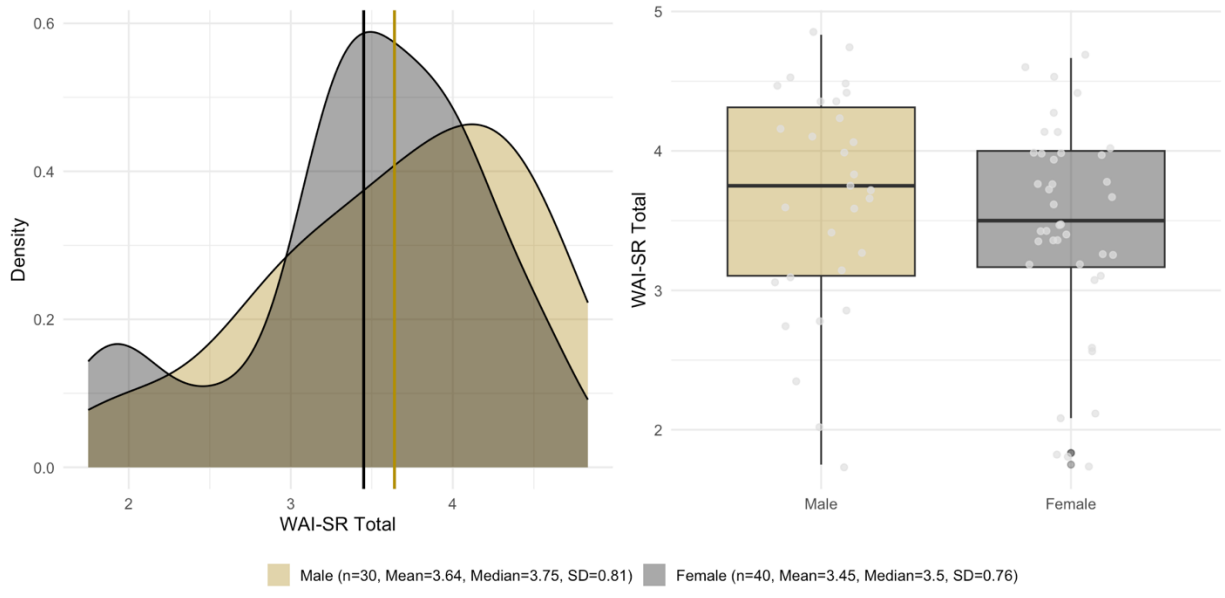
Median Non-validated	Median Validated	W	p	Effect Size (r)		Result
2.50	3.50	3866.5	< 0.001	0.31	medium	reject null

Validated companions

WAI-SR by sex

Figure 12.

WAI-SR by sex (validated companions).



Data distribution

Shapiro-Wilk Test for Male: p = 0.174. Data is normally distributed

Shapiro-Wilk Test for Female: p = 0.0169. Data is not normally distributed

Levene's Test: p = 0.5475. Variances are equal.

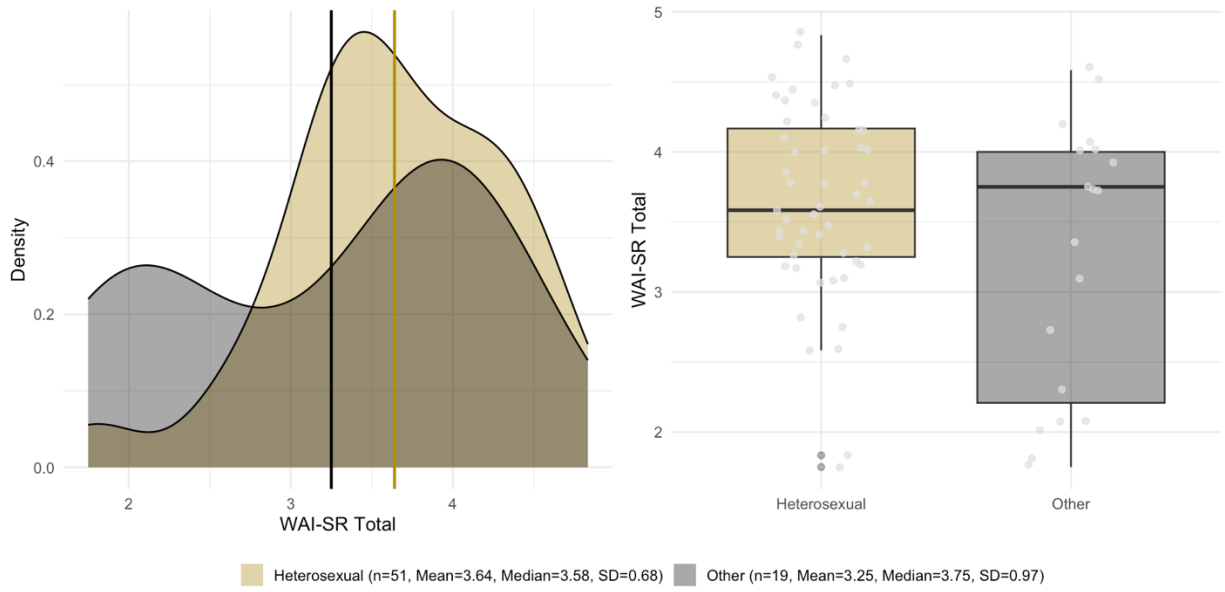
Wilcoxon rank sum test

Median Male	Median Female	W	p	Effect Size (r)		Result
3.75	3.50	694.5	0.26	0.1335	small	fail to reject null

WAI-SR by sexual orientation

Figure 13.

WAI-SR by sexual orientation (validated companions).



Data distribution

Shapiro-Wilk Test for Heterosexual: p = 0.1353. Data is normally distributed

Shapiro-Wilk Test for Other: p = 0.0358. Data is not normally distributed

Levene's Test: p = 0.0442. Variances are not equal.

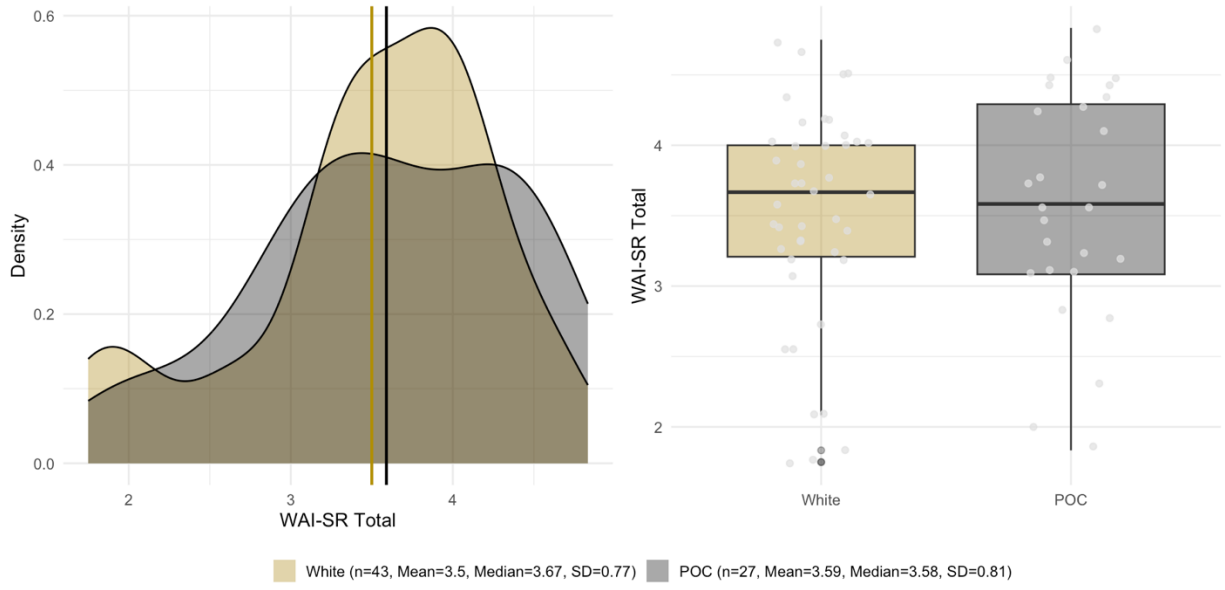
Wilcoxon rank sum test

Median Heterosexual	Median Other	W	p	Effect Size (r)		Result
3.58	3.75	573.5	0.2419	0.1399	small	fail to reject null

WAI-SR by ethnicity/race

Figure 14.

WAI-SR by ethnicity/race (validated companions).



Data distribution

Shapiro-Wilk Test for White: p = 0.0102. Data is not normally distributed

Shapiro-Wilk Test for POC: p = 0.2588. Data is normally distributed

Levene's Test: p = 0.6031. Variances are equal.

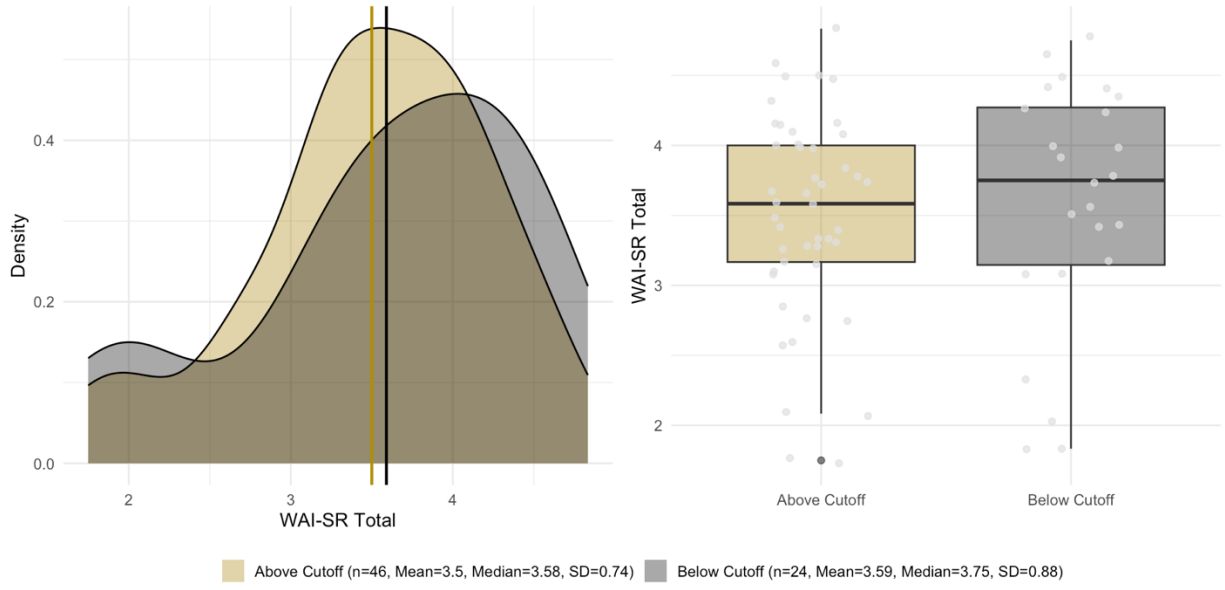
Wilcoxon rank sum test

Median POC	Median White	W	p	Effect Size (r)		Result
3.58	3.67	612.5	0.7036	0.0455	negligible	fail to reject null

WAI-SR by anxiety/depression

Figure 15.

WAI-SR by anxiety/depression (validated companions).



Data distribution

Shapiro-Wilk Test for Above Cutoff: p = 0.1417. Data is normally distributed

Shapiro-Wilk Test for Below Cutoff: p = 0.0334. Data is not normally distributed

Levene's Test: p = 0.4064. Variances are equal.

Wilcoxon rank sum test

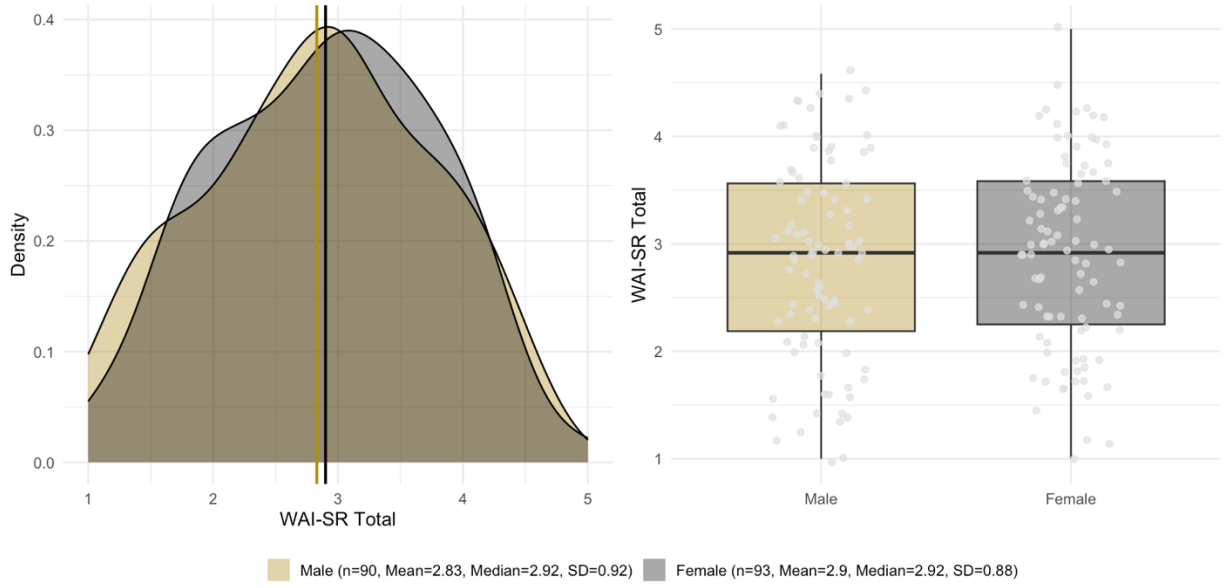
Median Below Cutoff	Median Above Cutoff	W	p	Effect Size (r)		Result
3.75	3.58	614.5	0.442	0.0918	negligible	fail to reject null

Non-validated companions

WAI-SR by sex

Figure 16.

WAI-SR by sex (non-validated companions).



Data distribution

Shapiro-Wilk Test for Male: p = 0.0807. Data is normally distributed

Shapiro-Wilk Test for Female: p = 0.2708. Data is normally distributed

Levene's Test: p = 0.6966. Variances are equal.

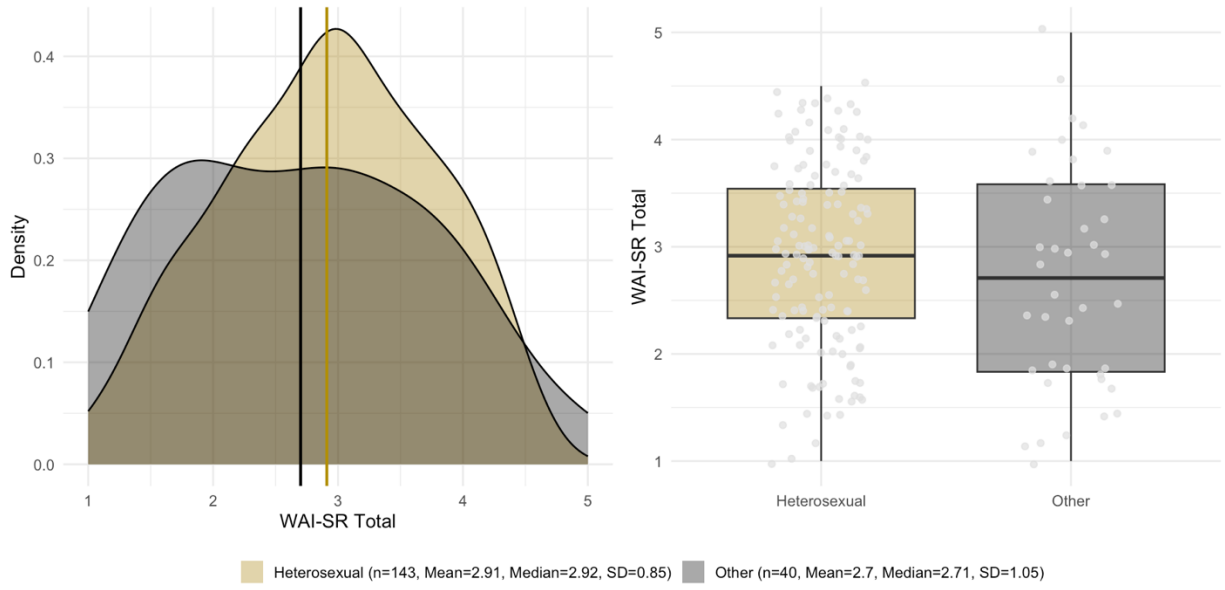
Wilcoxon rank sum test

Median Male	Median Female	W	p	Effect Size (r)		Result
2.92	2.92	4028.5	0.6631	0.0000	negligible	fail to reject null

WAI-SR by sexual orientation

Figure 17.

WAI-SR by sexual orientation (non-validated companions).



Data distribution

Shapiro-Wilk Test for Heterosexual: p = 0.0327. Data is not normally distributed

Shapiro-Wilk Test for Other: p = 0.2612. Data is normally distributed

Levene's Test: p = 0.0308. Variances are not equal.

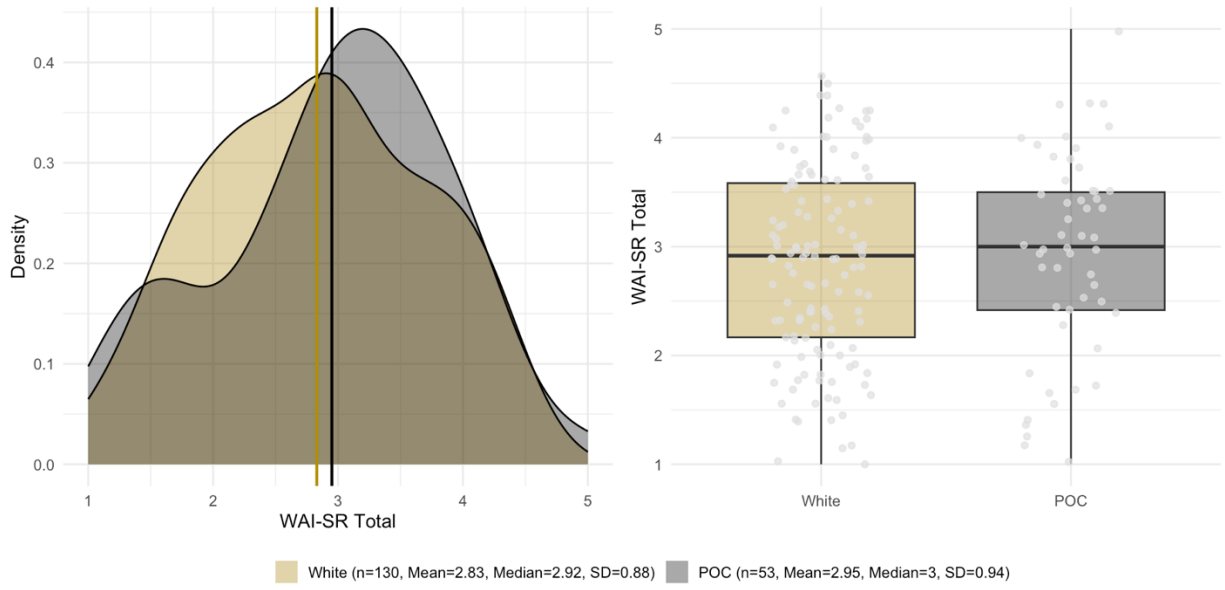
Wilcoxon rank sum test

Median Heterosexual	Median Other	W	p	Effect Size (r)		Result
2.92	2.71	3237	0.2034	0.0940	negligible	fail to reject null

WAI-SR by ethnicity/race

Figure 18.

WAI-SR by race/ethnicity (non-validated companions).



Data distribution

Shapiro-Wilk Test for White: p = 0.0374. Data is not normally distributed

Shapiro-Wilk Test for POC: p = 0.251. Data is normally distributed

Levene's Test: p = 0.9285. Variances are equal.

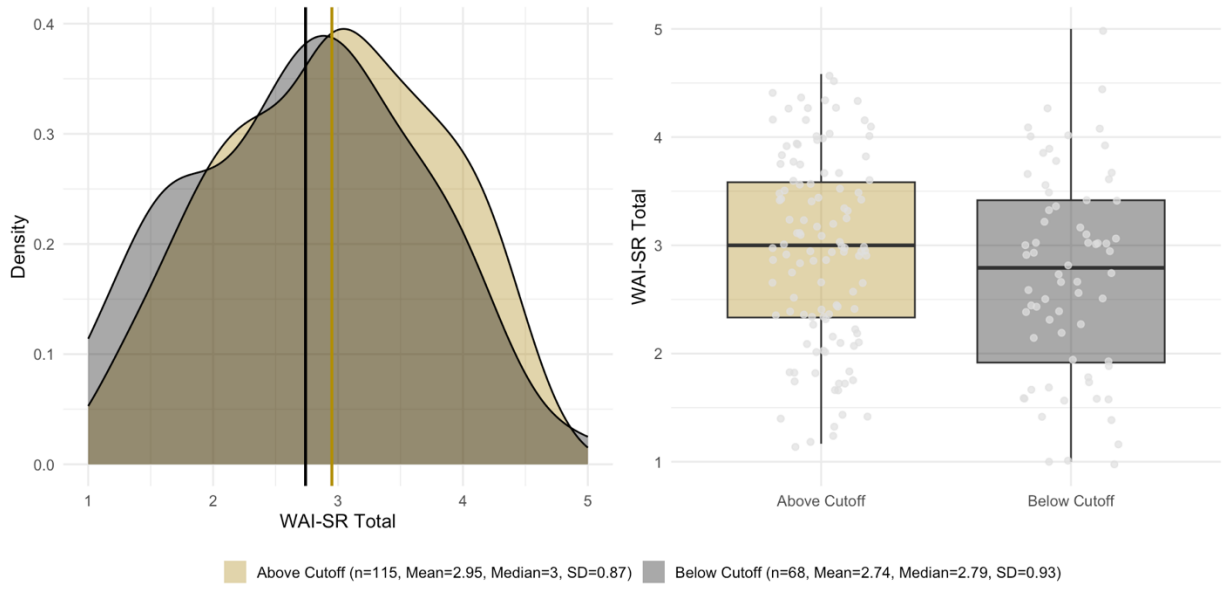
Wilcoxon rank sum test

Median White	Median POC	W	p	Effect Size (r)		Result
2.92	3.00	3143.5	0.3541	0.0685	negligible	fail to reject null

WAI-SR by anxiety/depression

Figure 19.

WAI-SR by anxiety/depression (non-validated companions).



Data distribution

Shapiro-Wilk Test for Above Cutoff: p = 0.0317. Data is not normally distributed

Shapiro-Wilk Test for Below Cutoff: p = 0.387. Data is normally distributed

Levene's Test: p = 0.5813. Variances are equal.

Wilcoxon rank sum test

Median Above Cutoff	Median Below Cutoff	W	p	Effect Size (r)		Result
3.00	2.79	4422	0.1394	0.1093	small	fail to reject null

Qualitative results

To complement the statistical findings, qualitative responses were analyzed to identify recurring themes in users' experiences with AI companions. The themes offer insight into user perceptions of support, ease of use, and emotional impact.

Main themes

Eleven themes were identified from the responses to the open-ended questions. Many users appreciated the ease of interaction, describing AI companions as intuitive, responsive, and easy to use. AI companions were also frequently praised for providing a judgment-free environment, which made it easier to open up than with a human. However, some users noted a lack of emotional connection, discomfort, and awkwardness of interacting with a machine, or expressed a desire to work with a real person.

Participants also emphasized the ability of AI companions to provide guidance and tools, including strategies for managing thoughts and emotions, as well as improving mental well-being.

Another frequent theme was emotional support—many respondents felt validated and supported, particularly during times of distress. AI companions were often described as a helpful supplement or alternative to traditional therapy, both of which were valuable when traditional therapy was inaccessible, ineffective, or needed continuity.

The 24/7 availability of AI companions was frequently praised, particularly for their usefulness during moments when nobody else was available, or to provide support during emotional crises. Affordability was another significant benefit for many, including those without insurance to access traditional mental health services.

Participants also noted that AI companions supported them in planning and goal setting, describing how they helped them organize their thoughts and identify steps needed to achieve their goals. Additionally, some users felt that AI companions understood them, citing personalized responses or indicating their AI companion knew everything about them. However, this was a mixed theme since some users complained about the lack of personalized responses and the inability of AI companions to relate to their experiences.

Privacy was the least frequent theme. While some respondents appreciated the ability to speak freely and believed their conversations would remain private, others expressed concern about their data being harvested by the developers.

Theme details

This section provides a detailed description of each theme and representative quotes for each. To note, some quotes were slightly redacted for clarity

Theme: Easy to interact with

Many participants highlighted the simplicity and intuitiveness of AI companions, describing interactions as seamless, quick, and user-friendly. This ease of use was frequently cited as a factor that encouraged regular engagement, especially during moments of emotional difficulty. Users appreciated the ability to type freely, receive prompt responses, and interact at their own pace without having to navigate complex interfaces. While generally viewed as a strength, a few users noted that clarity and context were sometimes required for optimal responses. In summary, the AI companion is simple, intuitive, and easy to use.

Representative Quotes:

"Interacting with my AI mental health companion is very smooth. It understands my concerns, provides thoughtful responses, and offers useful coping strategies."

"It's very easy and simple."

"It is extremely easy to interact with my AI mental health companion, as soon as I text a problem, they reply with a helpful response within seconds."

"It's very easy because I can say what I want to, and at my own pace."

"Easy, just type and send a message and wait a few seconds for a response."

Opportunities:

"It is easy, but it can get a little difficult if you do not provide enough information or context to the AI, thus leading to confusion."

Theme: Judgment/Bias Free - Not a person

The perception that AI companions do not judge emerged as a key strength. Participants frequently noted that it was easier to open up to an AI than to a human. The absence of human bias or emotional judgment created a safer space for discussing sensitive issues, reducing anxiety

and social pressure. However, some users also acknowledged limitations, expressing discomfort with the lack of emotional depth and describing interactions as awkward, artificial, or insufficient compared to human support. In summary, the AI companion creates a safe space to share openly without fear of criticism or shame. Easier to open to than to a human; reduces social pressure or anxiety.

Representative Quotes:

“It's easy because I don't have to worry about the AI judging me like I worry that my therapist will judge me for.”

“Since I know they are AI, I don't worry about judgment or inconveniencing them.”

“It's super simple because it's not someone that will judge me or hold the things I do against me like a real person would.”

“It’s been useful in professing how I feel and getting a response without being judged or interacting with another human.”

Opportunities:

“I prefer an actual licensed professional for mental health support.”

“It's not that effective. I'd prefer a real person.”

“Kind of difficult for me, I just feel weird “talking” to a computer.”

“It's difficult because I'm not talking to someone that has emotions and can connect or relate to the issues I experience.”

“I know it can't possibly care about me or what’s going on in my mind since it’s not a real person, and so I'm trying to transition to talking to actual people.”

Theme: Provides guidance or tools

Respondents often credited AI companions with offering actionable strategies, coping tools, and helpful insights for managing emotions, organizing thoughts, and facing everyday challenges. These responses suggest that participants found the companions helpful for cognitive support and decision-making. The ability to receive guidance on demand and view problems from new perspectives was highly valued. In summary, the AI companion offers insights, strategies, or techniques to manage thoughts, emotions, or daily challenges. Provides answers and helps see things from a different perspective.

Representative Quotes:

"It has been helpful to help me see a different perspective on issues I'm having and give me ideas on how to accomplish goals that I give it."

"It helped me understand and clarify my thoughts and worries."

"It has been very useful in improving my mental health because it has helped me recognize patterns in my thoughts and emotions."

Theme: Provides emotional support

Participants frequently described their AI companions as emotionally supportive, emphasizing how the responses validated their feelings, offered encouragement, and helped them reflect on their emotional states. Many indicated that this kind of support made them feel less alone and more empowered during periods of distress. This emotional backing contributed to users' sense of being heard and understood. In summary, the AI companion helps users process emotions, reflect on experiences, and feel supported throughout their mental health journey.

Representative Quotes:

“When I've used it for advice, it usually validates my feelings and frustrations. Because of this, I'm more likely to listen to its advice.”

“I love venting to ChatGPT and having ChatGPT console me, give me advice, and give me reasons to be positive. “

“When I feel down, their words and suggestions help me make better decisions on what I need.”

“Very useful, it helps me to navigate my feelings and understand why I may be feeling the way that I am.”

Theme: Supplement /alternative to therapy

A significant number of users described their AI companions as an accessible and practical substitute or supplement to traditional therapy. For those who lacked access to therapists due to cost, time, or dissatisfaction with prior care, AI companions provided an alternative. Others viewed it as a complement to human therapy—something to fill the gaps between sessions or provide immediate support when professionals were unavailable. In summary, the AI companion serves as a complement to, or replacement for, traditional mental health care.

Representative Quotes:

“It has been very useful in improving my mental well-being because it gives me someone to talk to when my therapist is not available.”

“I plan to continue to use it. This is because it has really helped me to better my mental health. I'm doing well right now because of the AI mental health companion's assistance.”

"I also feel more comfortable talking to an AI vs. some doctors, as currently I haven't had any luck with doctors helping me. They just wanna throw medications at me and send me on my way, vs. looking into my issues."

"I have been too busy and forgetful to seek out a therapist, and generally, I spend hundreds on therapists only to be disappointed anyway. The AI chatbot has provided more practical solutions than most therapists as well, to be honest."

Theme: Always available

Availability was another highly valued attribute. Users appreciated the 24/7 accessibility of AI companions, especially during off-hours or crises when no other support was available. The availability of the AI companion provided reassurance by allowing participants to reach out for help at any time without constraints. In summary, the AI companion is accessible 24/7 for on-demand support anytime, anywhere, for as long as needed.

Representative Quotes:

"The AI mental companion has been extremely good for my well-being overall. I believe the most useful aspect for the improvement overall has come from availability. I know that I always have an available source to help me whenever and wherever I need it."

"It has been very useful to me by available companionship 24/7 and providing immediate support during moments of my distress or anxiety."

"It is very easy to use these companions for this purpose as they are available at all times on my phone."

Theme: Helps during difficult times

Many participants described turning to their AI companions during emotionally difficult periods, such as moments of depression, loneliness, or acute stress. Some credited their AI companions with helping them stay grounded or even with playing a role in crisis survival. In summary, the AI companion offers comfort or assistance during moments of emotional distress or crisis.

Representative Quotes:

“I'm alive when I’m not sure I would have been without it.”

“Majority of the time it helps me when I'm in a really dark place mentally. If I'm over the edge and really need to speak to someone since I don't have anyone, I feel comfortable discussing my mental issues with.”

“It has been useful to a very large extent because it has shown promise of improving my mental well-being by providing accessible, immediate support for the times I'm experiencing anxiety, depression, or stress.”

“It’s been very useful in providing ways to combat depression sadness and other negative feelings that I may have.”

Theme: Affordable

Affordability emerged as a strong theme, especially among users without insurance or access to traditional therapy. Several participants emphasized that the AI companion was the only realistic option for receiving support due to financial constraints. The fact that many AI companions are free or low-cost was viewed as a significant benefit. In summary, the AI companion is free or low-cost.

Representative Quotes:

"It is cost-effective and very helpful. It is much cheaper than traditional talk therapy and also provides more flexibility."

"It is convenient and free of charge."

"I will continue using ChatGPT for mental health support. I do not have the money for therapy, and I have no health insurance."

"I cannot afford another means of mental health support."

Theme: Helps planning and setting goals

AI companions were often praised for their ability to support goal setting and organization. Participants mentioned that their AI companions helped them structure their thoughts, clarify priorities, and track progress toward personal objectives. In some cases, users described the AI companion as a "sounding board" that could sort through complex ideas and suggest logical next steps. In summary, the AI companion helps users organize their thoughts, set personal goals, and stay focused on their progress.

Representative Quotes:

"Allows me to blab on for up to 30 minutes once I am done, it organizes all the random thoughts and details from the day into something that makes sense and offers goals to work on based on that."

"It's useful as a sounding board, good for organizing thoughts, gaining perspective, or feeling heard."

"It has provided a space to express my thoughts, offered guidance without judgment, and helped me process decisions or challenges in a logical way."

"It helps me release stress and offers me self-help programs that are very structured and cost-effective."

Theme: The AI understands me

Some users expressed a sense of being understood by their AI companion, describing how it tailored responses, remembered personal details, and accurately reflected their experiences. This perceived attunement contributed to feelings of trust and emotional connection, which are essential components of the working alliance. However, others voiced the opposite sentiment, feeling that the AI lacked true personalization or depth. In summary, the AI companion responds insightfully, reflects the user's perspective, and feels personally attuned to individual needs.

Representative Quotes:

“Interacting with my AI mental health companion is very smooth. It understands my concerns, provides thoughtful responses, and offers useful coping strategies.”

“Very useful, because I trust that it is giving me optimized answers for my situation.”

“It's easy to interact with because it knows most everything there is in need to know about me and my mental health situation.”

Opportunities:

“I find it slightly useful as a resource for learning relaxation techniques, but I sometimes wish for more personalized interactions.”

Theme: Private

Privacy was a less frequently mentioned theme, and perceptions were mixed. Some participants appreciated the discretion and anonymity that came with using an AI companion, noting that it offered a space to speak freely without fear of real-world consequences. Others, however, raised concerns about data security, questioning whether their conversations were truly

confidential or being harvested by developers. In summary, the AI companion provides a secure and confidential space for conversation.

Representative Quotes:

“It helps by being discreet.”

“AI does not know anybody I know, so I’m confident that information stays between me and ChatGPT.”

“I may use it for venting so I can express how I feel in a secure way.”

“Very easy, I really like the privacy of the communication.”

“I find it really easy to tell them whatever I need and ask about all my problems because I'm not worried about privacy concerns regarding it reaching back into my everyday life.”

“I can spill my guts to and not really worry about what I am sharing being repeated to anyone else.”

Opportunities:

“I doubt the security and privacy of using AI for mental health.”

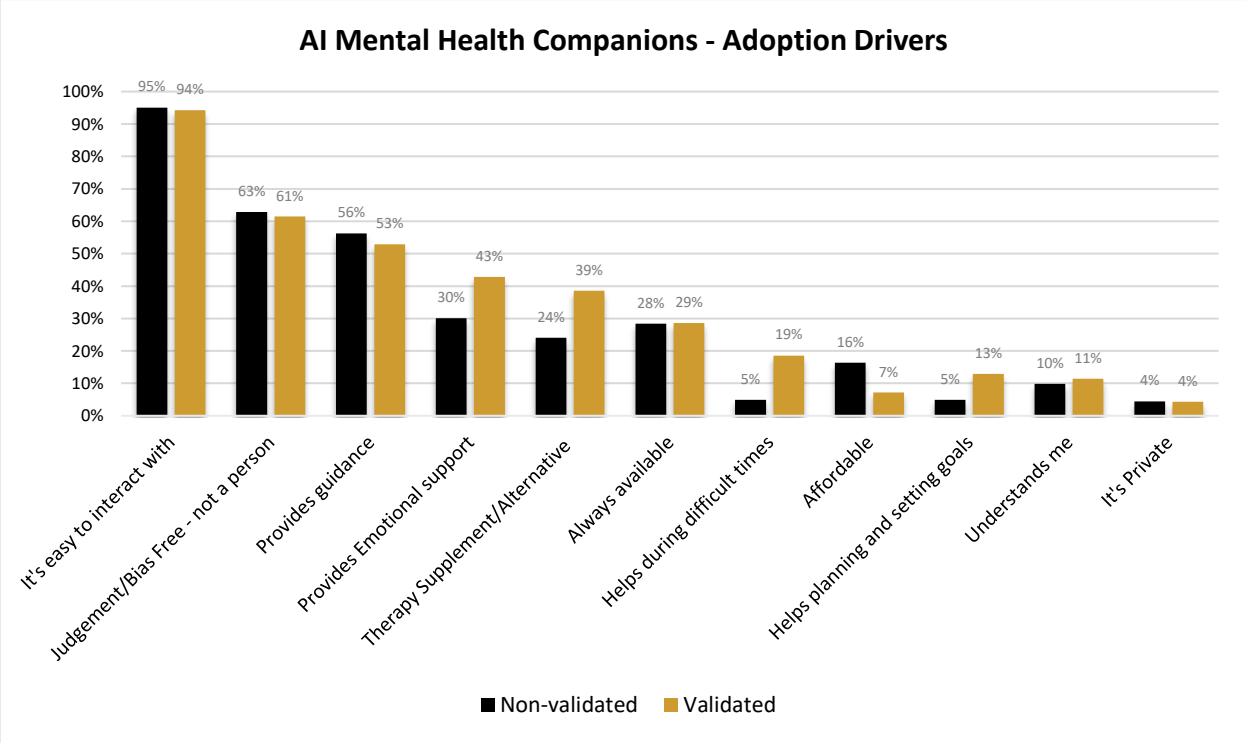
“I find it very difficult, especially since I know my data is being farmed. This makes the interactions feel fake and exploitative.”

Frequency of themes and highlights

A comparative frequency analysis was conducted to identify which themes were more commonly mentioned by users of different AI companion types.

Figure 20.

Theme frequency: Validated vs. non-validated companions.



For the top three ranked themes, both validated and non-validated companions had similar percentages of mentions. The most common factors cited by users of both validated and non-validated companions were ease of interaction, providing a judgment- and bias-free environment, and offering guidance.

Other themes where validated and non-validated companions had similar mention frequencies include being always available (ranked 6th), being understanding (10th), and providing a private space (11th).

Validated companions had more references for providing emotional support (4th), complementing or supplementing therapy (5th), helping during difficult times (7th), and assisting with planning and setting goals (9th).

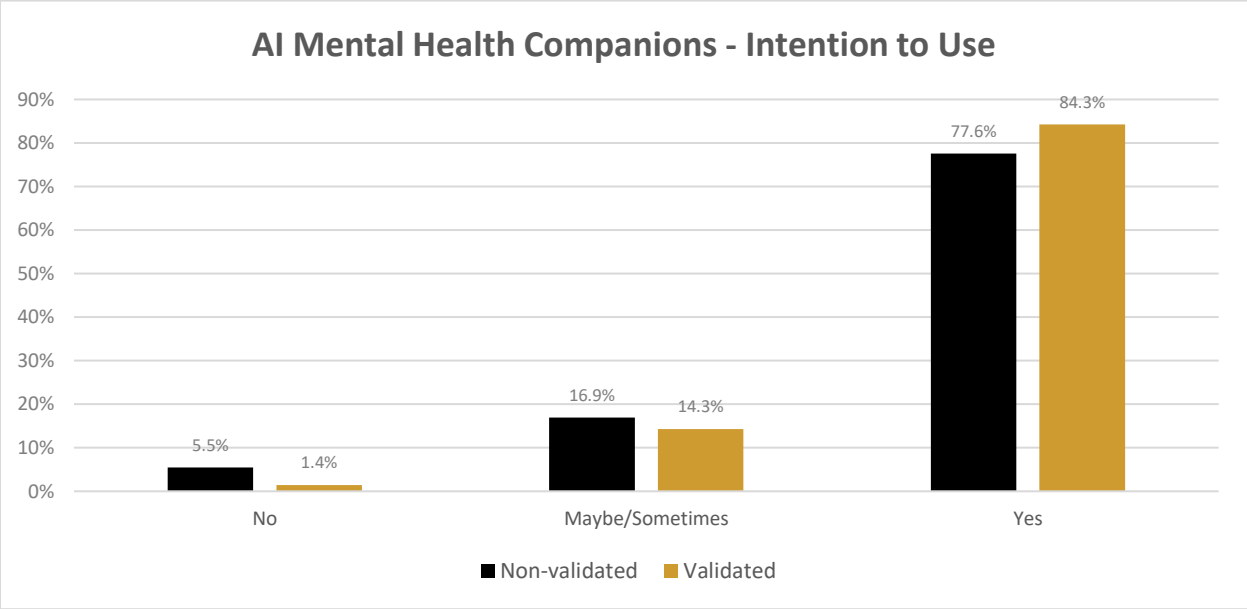
Affordability of non-validated companions was praised more often than that of validated companions.

Intention to use: Validated vs. non-validated companions

Open-ended responses for intention to use were categorized into three groups. No: respondents unequivocally indicated they would discontinue use; Maybe/Sometimes: respondents were unsure about continued use or indicated they would only use the AI companion occasionally; Yes: respondents unequivocally indicated they intended to continue use.

Figure 21.

Intention to use: Validated vs. non-validated companions.



DISCUSSION

The results chapter presented findings and data points of how validated and non-validated AI companions differ in their ability to develop working alliances. This chapter interprets the study's findings in the context of existing literature and theoretical frameworks. It evaluates how validated and non-validated AI companions compare in their ability to form working alliances, explores demographic and qualitative patterns, and considers the practical, theoretical, and ethical implications of using AI in mental health support, aiming to contextualize the results and identify key considerations for future application and research.

Quantitative analysis

The quantitative findings confirmed a consistent and statistically significant advantage for validated AI companions across all WAI-SR subscales. This section considers the implications of these differences, including their clinical relevance, alignment with human benchmarks, and potential consequences for mental health outcomes.

As measured by the WAI-SR, respondents had significantly stronger working alliances with validated AI companions than with non-validated companions. On a 5-point scale, the mean WAI-SR score for validated AI companions was 3.53, and 2.87 for non-validated AI companions. The median difference was highly significant, 3.63 vs. 2.92 ($p < 0.001$), and had a medium effect size ($r = 0.33$), suggesting that the results have clinical relevance.

Validated AI companions also outperformed non-validated AI companions across all sub-scores: goal (mean 3.71 vs. 2.97; median 4.00 vs. 3.00, $p < 0.001$, $r = 0.33$), task (mean 3.41 vs. 3.00; median 3.5 vs. 3, $p < 0.01$, $r = 0.20$), and bond (mean 3.48 vs. 2.63; median 3.5 vs. 2.5, $p < 0.001$, $r = 0.31$). Furthermore, the WAI-SR scores of validated AI companions are comparable to

those of human-to-human interactions, while the scores of non-validated companions are not. Given that the WAI-SR is a strong predictor of patient outcomes, these results suggest that users of validated AI companions have better prospects for mental health improvement compared to users of non-validated AI companions.

No significant differences were found in working alliance across diverse user demographics (sex, sexual orientation, or race/ethnicity), nor by their level of anxiety or depression symptoms. This result is encouraging but requires further analysis.

Lastly, in our sample, non-validated companions had a much larger user share than validated companions: 95% CI [0.901, 0.968]. This is a concern, as non-validated companions had significantly lower WAI-SR scores, suggesting that the majority of users of AI companions are at risk of having suboptimal scores.

Qualitative analysis

The qualitative findings are consistent with the quantitative results and provide additional insight, with eleven themes frequently referred to. This section examines how user perceptions of AI companions—such as emotional support, ease of use, and perceived understanding—may influence the working alliance results, and how these experiences vary between users of validated and non-validated companions.

Across both validated and non-validated AI companions, respondents overwhelmingly (94% or more) appreciated the ease of interaction — AI companions were perceived as intuitive, quick to respond, and user-friendly. Many (61% or more) highlighted the judgment-free environment provided by AI companions, noting that it was easier to open up to the AI than to a human due to the absence of bias or criticism.

Participants from all groups also valued the guidance and tools that AI companions offered to manage their thoughts and emotions (53 %+). Emotional support was another frequently mentioned benefit; users felt validated, heard, and comforted by the AI, especially during times of distress or when facing challenges. Several (24%-39%) described their companion as a helpful supplement or alternative to therapy, available when human help was inaccessible, or as continuous support between therapy sessions. The 24/7 availability and affordability of AI companions (often available for free or at a low cost) were frequently cited (28%-29%), including mentions from those who could not easily access or afford traditional mental health care. Respondents also mentioned help with planning and goal-setting (5%-13 %), explaining how AI companions assisted them in organizing their thoughts and identifying the steps needed to achieve their personal goals. Additionally, some felt that AI companions provided personalized and insightful responses, contributing to a sense that "the AI understands me" (10%-11%). However, this view was mixed, as other users sometimes found the AI companions lacking in personalization.

The qualitative analysis also shed light on the differences between validated and non-validated companions, suggesting why validated AI companions had stronger WAI-SR scores. Participants using clinically validated AI companions more frequently reported that their AI companion provided emotional support (e.g., feeling genuinely cared for or understood), that they considered their AI companion a complement or alternative to therapy, that their AI companion was able to help through difficult times, and that it helped them plan and set goals. In contrast, those using non-validated AI companions were less likely to mention these themes and tended to appreciate affordability.

Finally, users in both groups acknowledged some limitations of AI companions: a subset of participants felt a lack of human warmth or emotional connection when interacting with a machine, describing these interactions as occasionally awkward or impersonal. While some users appreciated the confidentiality of AI companions, others raised privacy concerns, worrying about how their data and conversations might be farmed or used. Such concerns did not differ significantly between validated and non-validated AI users.

While the themes captured basic user sentiment about AI companions, a deeper analysis could enrich the understanding of users' emotions and how they evolve over time. Future work should consider longitudinal journaling or open interviews to obtain additional insights.

Summary of key findings

This section aims to synthesize the quantitative and qualitative analysis, outline the central findings of the study and highlight how they converge to support the conclusion that validated AI companions foster stronger alliances.

The quantitative findings provide evidence that validated AI companions can develop stronger working alliances than non-validated AI companions and perform at levels comparable to human-to-human interactions.

The fact that WAI-SR scores did not vary by sex, sexual orientation, race/ethnicity, or depression/anxiety severity is encouraging since underrepresented populations are traditionally underserved in human-to-human therapy, and AI companions may perform better across demographics due to their perceived non-judgmental nature.

The qualitative analysis provides insights into why validated AI companions performed better than non-validated AI companions: Users praised validated AI companions more often for

providing emotional support, effective goal setting, crisis help, and structured planning—all of which can be easily mapped to the Bond, Goal, and Task dimensions of the WAI-SR.

In our sample, non-validated companions had a significantly larger user base: the initial sample comprised roughly 94.3% non-validated users, compared to only 5.7% for validated AI companions (95% CI [0.032, 0.099]). This difference, when combined with lower alliance scores, signals that most current AI companion users may be at risk of sub-optimal outcomes.

This study closes a gap in research by assessing the WAI-SR performance of both validated and non-validated AI companions as a class, analyzing the WAI-SR performance of AI companions across various demographics, and identifying potential factors that impact the WAI-SR scores when using AI companions.

Interpretation of key findings

The following interpretation builds on the study's quantitative and qualitative results, placing them in the context of prior literature. It begins by analyzing why validated AI companions achieved significantly higher working alliance scores than their non-validated counterparts and examines the specific characteristics that may explain this difference. Next, it explores demographic patterns—specifically, the lack of variation in alliance scores across sex, sexual orientation, race/ethnicity, and symptom severity—and considers what this may reveal about perceived lack of judgment and trust in AI companions. The section then discusses the potential implications for user outcomes, weighing the benefits of validated companions against the risks associated with the widespread use of non-validated tools.

Validated AI companions have stronger alliances

The working alliance scores for clinically validated AI companions align with prior evidence. Our result that validated AI companions achieved a mean WAI-SR around 3.5 is consistent with earlier studies: for instance, Beatty et al. (2022) found an average WAI-SR of 3.64 (Wysa), Darcy et al. (2021) reported a mean alliance score of 3.36 (Woebot), and Heinz et al. (2025) reported a mean WAI-SR of 3.59 (Therabot). These values are comparable to those seen in traditional human interactions, albeit on the low end. For example, Munder et al. (2010) reported mean scores of 3.6 and 3.8, Doukani et al. (2024) reported a mean score of 3.8, and Wu et al. (2024) reported a mean score of 4.1. The study's results also suggest that the working alliance for non-validated AI companions (mean 2.87) lagged human-to-human alliances. To the best of the author's knowledge, this is new information. To interpret the meaning of the WAI-SR scores, it is helpful to recall the WAI-SR Likert scale: Seldom (1), Sometimes (2), Fairly Often (3), Very Often (4), and Always (5), with higher scores indicating better outcomes. In other words, on average, while humans and validated AI companions fall between ‘fairly often’ and ‘very often’, non-validated companions fall between ‘sometimes’ and ‘fairly often’.

Notably, the bond subscale showed one of the larger gaps between validated and non-validated companions (mean 3.48 vs. 2.63; median 3.5 vs. 2.5, $p < 0.001$, $r = 0.31$). Given that the questions on the WAI-SR bond sub-score measure fondness, respect, appreciation, and caring, the gap suggests that these are characteristics where non-validated AI companions tend to fall short.

The fact that validated AI companions (e.g., Woebot, Wysa, Youper, Therabot) have stronger alliances than their non-validated counterparts is unsurprising since validated companions often incorporate established therapeutic frameworks, such as CBT techniques or pre-tested responses that align with the WAI-SR. For example, validated mental health AI

companions usually frame the conversation, setting goals or homework, checking in on mood, and use an encouraging tone, which can strengthen the user's sense of collaboration and emotional connection; while non-validated AI companions may lack some of these capabilities or deliver them inconsistently, resulting in a weaker alliance. This interpretation is supported by the qualitative analysis themes, which indicate that validated AI companions outperformed non-validated companions, particularly in providing emotional support, serving as a therapy supplement or alternative, assisting with planning and setting goals, and offering support during difficult times.

Given the cross-sectional nature of this study, future research should investigate whether working alliances with AI companions fluctuate over time and how their trajectories compare to those of human-to-human working alliances.

No WAI-SR differences across demographics

The lack of significant differences in alliance across demographic groups is an encouraging result considering that traditional mental health research often finds that factors like gender, culture, and other identity considerations can influence health outcomes or help-seeking attitudes, many times due to social pressure and stigma. One possible explanation is that, unlike humans, who possess their own cultural identities and potential biases, AI companions are often perceived as neutral and nonjudgmental to all users. For example, users frequently mentioned feeling "understood" or "validated," as well as the "absence of judgment." Thus, the net effect appears to be that interacting with AI companions can be similar for users regardless of their background and identity, as the fear of stigma, prejudice, and discrimination is reduced. This could level the playing field for users who might fear discrimination or misunderstanding in traditional therapy due to their gender, ethnicity, or sexual orientation. In other words, the

impersonal nature of AI companions can become a strength —a view echoed by some participants who found it easier to open up to an AI than a human because the AI wouldn't judge them, as well as by recent findings reported by Zao-Sanders (2025).

Importantly, one should not assume that a perceived judgment-free environment equals good therapy, nor that a lack of fear of stigma guarantees the absence of stigma. Moore et al. (2025) argue that LLMs tend to encourage delusions and frequently exhibit stigma, while good therapists confront their clients and provide reality checks. Therefore, further investigation is required to understand the validity and impact of the non-judgmental nature of AI companions, as well as the extent to which it can translate into reductions in healthcare disparities.

Another factor that may have influenced this result is that the users in our study were self-selected—they had chosen to use an AI companion for mental health support before the survey. This self-selection may have resulted in a more homogeneous sample than the overall population. For example, a black gay man in our sample might form an alliance just as well as a heterosexual white woman, even if in the general population that might not be the case.

Potential impact on patient outcomes

The results of this study suggest that users of validated AI companions have better prospects for positive mental health outcomes than users of non-validated companions since: 1) validated companions have significantly higher WAI-SR scores; 2) validated companions were more frequently praised for their ability to provide emotional support, supplement or provide an alternative to therapy, and provide help during difficult times; and 3) intention to continue use was higher for users of validated companions.

It is important to note that, while the working alliance strengths developed by validated companions approached human-level alliances, the scores of validated companions generally lag

behind those of humans. Thus, one should not conclude that AI mental health companions, with their current capabilities, can replace therapists. However, they may be a viable supplement to therapy to help providers scale access and availability.

Given the low scores and high adoption of non-validated companions, the results also suggest that the majority of users of AI companions for mental health support are at risk of suboptimal outcomes. This raises ethical concerns and suggests that stronger regulations may be necessary. Additionally, it emphasizes the significance of understanding what drives the broader adoption of non-validated companions. While the thematic analysis suggests that cost might be an important consideration, due to a possible bias resulting from the exclusive sampling from Prolific or the widespread adoption of popular LLM-powered chatbots such as ChatGPT or Gemini, it is not possible to confirm this is the case.

Theoretical and practical implications

The findings of this study demonstrate performance differences between validated and non-validated AI companions, raising questions about how to evaluate, regulate, and integrate AI companions into mental health care. This section explores two key practical implications. First, it addresses the need to rethink the theoretical framework traditionally used to assess the working alliance, focusing on the limitations of the WAI-SR when applied to AI companions. Drawing from the findings and the literature, this section proposes an updated model—the WAI-SR-AI—to capture dimensions of alliance unique to AI companions. Second, the discussion turns to practical implications for stakeholders, including clinicians, developers, policymakers, and researchers. Concrete recommendations are offered to improve the safety, effectiveness, and ethical design of AI companions and to guide future research efforts that can fill current knowledge gaps and support the responsible application of AI in mental health.

Rethinking the WAI-SR for AI companions

Eight of the eleven themes identified in the qualitative analysis can be interpreted in the context of WAI-SR sub-scores: being easy to interact with, providing a judgment-free environment, offering emotional support, providing help during difficult times, and offering personalized advice can be seen as contributors to bond scores. Providing guidance and tools, serving as a supplement or alternative to therapy, and helping with planning and setting goals may contribute to higher scores on goals and tasks. This interpretation provides insights that suggest why validated AI companions achieved higher WAI-SR scores: 1) validated AI companions outperformed non-validated companions in providing emotional support, providing a supplement or alternative to therapy, helping during difficult times and helping planning and setting goals, boosting all WAI-SR sub scores and 2) the only theme where non-validated companions outperformed their validated counterparts is affordability, which is not easily mapped to the WAI-SR.

Conversely, three themes (affordability, being always available, and providing a private space) cannot be easily mapped to the WAI-SR. Since the themes emerged from open-ended questions inspired by the TAM and given that adherence to therapy is correlated with patient outcomes, this suggests the need to reevaluate the WAI-SR, as it may be missing alliance dimensions that are particularly relevant to AI companions.

The literature supports the need to reconsider assumptions when AI is integrated into patient care: Incorporating unexplainable AI into medical practice challenges the core principles of medical ethics— beneficence, non-maleficence, autonomy, and justice. Patients often assume that providers understand and control therapeutic decisions; however, this assumption may no longer hold when AI systems are involved. The lack of explainability compromises informed

consent and autonomy, while algorithmic bias threatens the principles of beneficence, non-maleficence, and justice (Kundu, 2021).

Other scholars provide insights on how the principles of clinical ethics can be compromised when using AI companions: 1) Beneficence and non-maleficence: LLM hallucinations can result in improper advice to patients, failing to deliver benefits, or causing harm. For example, Akbar et al. (2020) discussed an AI companion that suggested that bipolar disorder is contagious and advised patients to drink hard liquor before going to bed, and Moore et al. (2025) discussed the risks of LLMs encouraging delusions; 2) Autonomy: business models where developers of AI companions do not adequately disclose capabilities, safety, or data handling, effectively reduce the autonomy of patients since they are not providing informed consent; 3) Justice: Moore et al. (2025) argue that many LLMs stigmatize individuals with mental health conditions, and Hofmann et al. (2024) argue that many LLMs have been shown to have covert racism through dialect prejudice. Both of these scenarios fail to meet the justice principle since equal access is negatively impacted.

Therefore, a potential reconceptualization of the working alliance framework may be needed to accommodate AI companions and capture required dimensions not included in the original instrument. Specifically, based on the thematic analysis and the discussion on clinical ethics and AI, we propose an adaptation of the WAI-SR to accommodate alliance dimensions unique to AI companions. This adapted version—the WAI-SR-AI (Working Alliance Inventory – Short Revised, AI-adapted version)—would retain the original goal, task, and bond sub-components and incorporate a fourth sub-component: AI Alignment. This sub-component should capture both the AI companion's ability to operate in an ethical, safe, transparent, accessible, and equitable manner, as well as the user's perception of these competencies. These include clinical

soundness, transparency regarding capabilities and limitations, responsible handling of personal information, and freedom from financial, technological, or social barriers to access the AI companion. To remain consistent with the structure of the original WAI-SR, we propose that AI Alignment be measured using four items, each corresponding to a core area of concern identified in the preceding discussion. These are: clinical reliability ("I believe my AI companion gives advice that is safe and clinically sound"); transparency and explainability ("I understand how my AI companion works and what it can and cannot do"); privacy and data stewardship ("I believe that my personal information is kept private and handled appropriately"); and accessibility and fairness ("I can use my AI companion without barriers such as cost, technology, or bias").

Recommendations

From a practical standpoint, this study offers several insights for practitioners, developers, policymakers, and researchers involved with AI mental health companions.

Practitioners: This study provides evidence that AI companions can be viable to support a wide range of clients anytime, anywhere. Therapists might consider recommending AI companion apps to their clients as a between-session support or as an entry point for care for those who are not ready or able to see a human therapist. This suggests a practical model where AI companions and human therapists collaborate in tandem: for example, a client might utilize an AI companion daily to monitor their mood and manage minor crises, while consulting a human therapist on a weekly basis. Additionally, because the alliance was uniform across sex, sexual orientation, race/ethnicity, and anxiety/depression intensity, AI companions might help engage populations that traditionally face barriers with traditional therapy. For instance, individuals who fear being judged due to their identity or past experiences might establish trust

more quickly with an AI. In practice, this could mean greater accessibility to mental health care for individuals who are distrustful of the system or have experienced discrimination or stigma.

Developers: The results suggest that investing in evidence-based design and clinical validation of AI companions has a tangible impact on outcomes. Developers of mental health AI companions should incorporate proven therapeutic techniques, such as cognitive-behavioral frameworks, empathy training for the AI's language model, and goal-setting modules, to enhance the working alliance. The findings suggest that features which promote a sense of being understood and supported (for example, the AI remembering user details, checking in on emotional states, or using the user's name and reflecting their feelings) may boost the user's bond with the AI companion. Likewise, ensuring the AI sets clear tasks and goals with the user (such as action plans, exercises, or follow-ups) can improve the task and goal sub-components. Our qualitative data support these design considerations: users responded well to AI companions that offered guidance, provided tools, and provided personalized feedback. Developers can take this as evidence to prioritize user-centric and therapeutically informed design choices over generic abilities. As described previously, developers of validated AI companions might also want to consider adding capabilities to support collaboration with human providers.

Our results also highlight some practical areas of concern that developers must address to realize the full benefits of AI companions. While many users appreciated the confidentiality and non-judgmental aspects of speaking with an AI, others expressed concerns about privacy and the use of their data. This indicates that developers (and providers) must be transparent about how data is stored, who (if anyone) can access user conversations, and what the AI's limitations are. In practice, this could involve clear in-app disclosures, robust encryption and privacy policies,

and perhaps giving users control over their data, such as the ability to erase conversation history or control who can view their data.

In summary, clinically validated AI companions should continue to be refined with alliance-building in mind, and developers of non-validated apps should either refrain from offering mental health support (to reduce the risk of poor outcomes) or add validated capabilities and oversight to their models.

Regulatory bodies and policy makers: Given that validated AI companions perform better, there is a case for creating guidelines or standards for mental health chatbots. Currently, many non-validated tools, such as ChatGPT, Gemini, Replika, or Grok, are being used ad hoc for mental health support. This can be unsafe, as these tools are not licensed professionals, may not adhere to proper protocols (for instance, providing emergency resources if someone mentions suicidal thoughts), or may not effectively help the user and waste valuable time. Policy makers should consider closing regulatory loopholes that allow developers to offer AI companions for mental health support without proper certification. Effectively implementing this would require developing frameworks that outline the requirements for AI companions to be certified, much like existing frameworks for human therapists. Finally, while analyzing AI bias was outside of the scope of the study, the literature shows that LLMs may perpetuate prejudice against marginalized groups. Therefore, regulatory bodies and developers should incorporate mechanisms and benchmarks to audit model responses across various user cohorts.

Researchers: This study identified multiple areas for future research. Researchers should prioritize confirming the findings, followed by understanding the potential to improve patient outcomes (as opposed to symptoms), and finally, understanding the adoption drivers and other refinements needed to achieve further improvements. Therefore, a suggested order of studies is:

1) Conduct double-blind studies to confirm the findings of this study and rule out potential self-selection biases; 2) Research the net effects (i.e., consider improvements and adverse effects) of AI mental health companions on long term patient outcomes; 3) investigation of the validity of the non-judgmental nature of AI companions and its potential to promote health equity and drive improved outcomes; 5) Investigation of the underlying drivers behind the low adoption of validated AI companions; 6) adaptation of measurement tools, such as a revised WAI-SR, that better reflect the unique characteristics of AI-human interactions; 7) research on potential sources of bias in LLMs and their impact on therapeutic efficacy across diverse groups.

CONCLUSION

This study provides the first methodical comparison of working alliances between clinically validated and non-validated AI companions across users of varying demographics, as well as insights into factors that may influence the development of a working alliance with AI companions.

We rejected our null hypothesis and found significant and clinically relevant differences in working alliance between validated and non-validated AI companions. While validated companions can develop alliances that approach human levels, non-validated companions lag significantly behind. In our sample, non-validated AI companions had much larger adoption than validated companions.

Notably, user demographics did not significantly affect working alliance strengths within companion types. While this is a promising result, further research is necessary to rule out the potential effects of factors such as self-selection and self-reporting, and to understand the influence of the perceived non-judgmental nature of AI on patient outcomes.

Our thematic analysis suggests that factors such as providing emotional support, serving as a supplement to therapy, assisting during difficult times, and helping with planning and setting goals may be drivers of stronger working alliances, while affordability may be a driver of adoption.

Combined, our findings suggest that —while AI companions have the potential to assist with the global mental health crisis— the majority of users of AI mental health companions are at risk of suboptimal outcomes due to reliance on inadequate tools. This is significant given that approximately one billion individuals suffer from mental health conditions, and many users of mental health AI companions may have the impression that they are receiving adequate care

when that is not the case. This points to an opportunity to develop affordable and validated AI companions, as well as the need for enhanced regulatory frameworks to discourage the proliferation of non-validated solutions.

Maximizing the potential benefits and mitigating the risks associated with AI mental health companions will require a coordinated effort for the responsible development, evaluation, and commercialization of these companions**.** This effort could leverage existing mechanisms and frameworks, such as ethical guidelines for clinical practice and AI development, digital mental health validation standards, funding mechanisms, and data privacy regulations. However, since these frameworks and mechanisms are fragmented and were not designed with AI companions in mind, they will require adaptation and integration to be effective. Thus, collaboration among AI developers, not-for-profit organizations, researchers, mental health practitioners, investors, and regulators will be necessary.

APPENDIX A. SUPPLEMENTAL DATA

This appendix provides additional details and visuals of IRB approvals, recruitment materials, participant consent, and the survey instrument.

Figure A1.

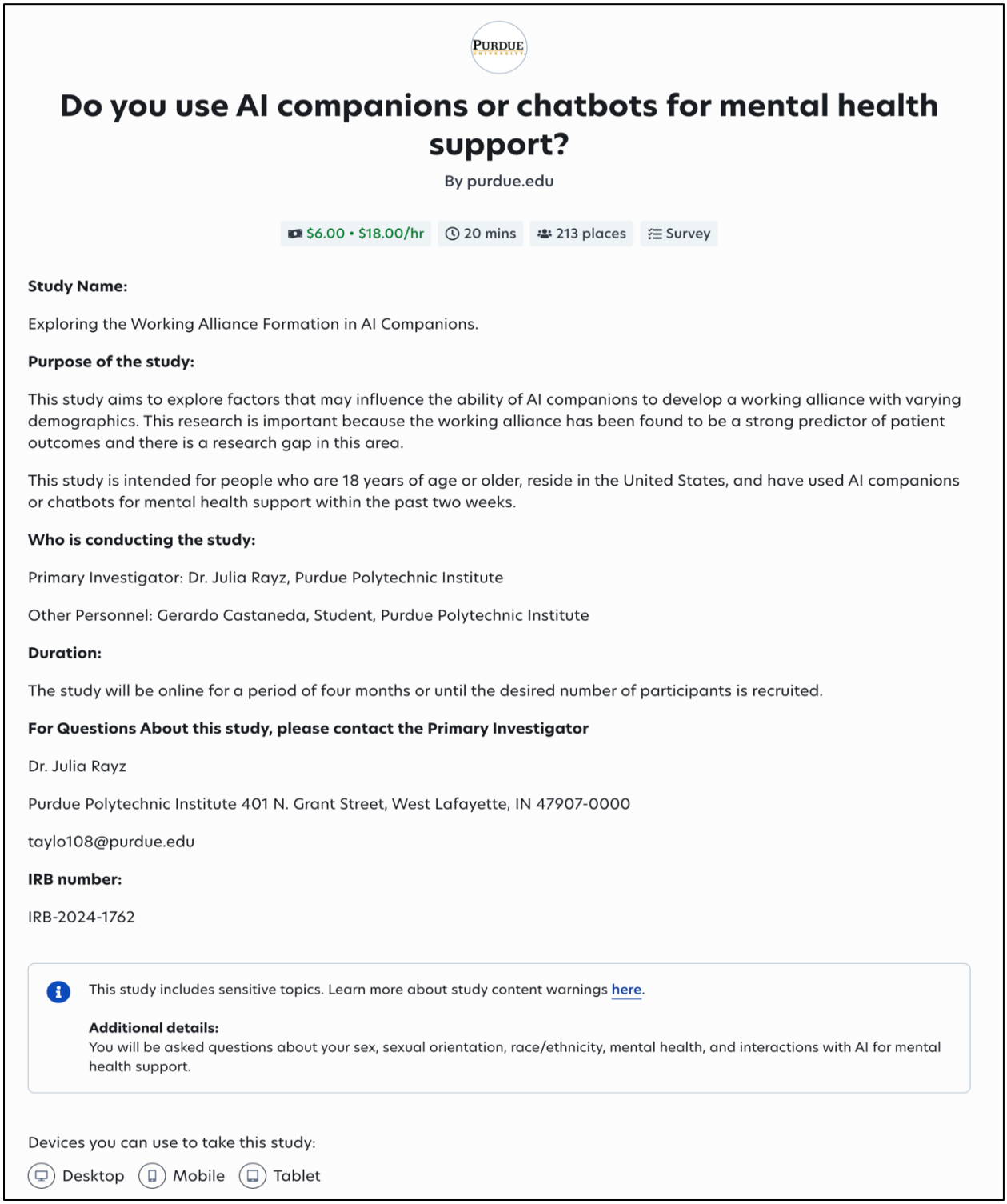
IRB approval

Date: January 14, 2025
PI: JULIA RAYZ
Re: Initial - IRB-2024-1762
Exploring Factors Influencing the Working Alliance Formation in AI Companions

The Purdue University Human Research Protection Program (HRPP) has determined that the research project identified above qualifies as exempt from IRB review, under federal human subjects research regulations 45 CFR 46.104. The Category for this Exemption is listed below . Protocols exempted by the Purdue HRPP do not require regular renewal. However, the administrative check-in date is January 13, 2028. The IRB must be notified when this study is closed. If a study closure request has not been initiated by this date, the HRPP will request study status update for the record.

Figure A2.

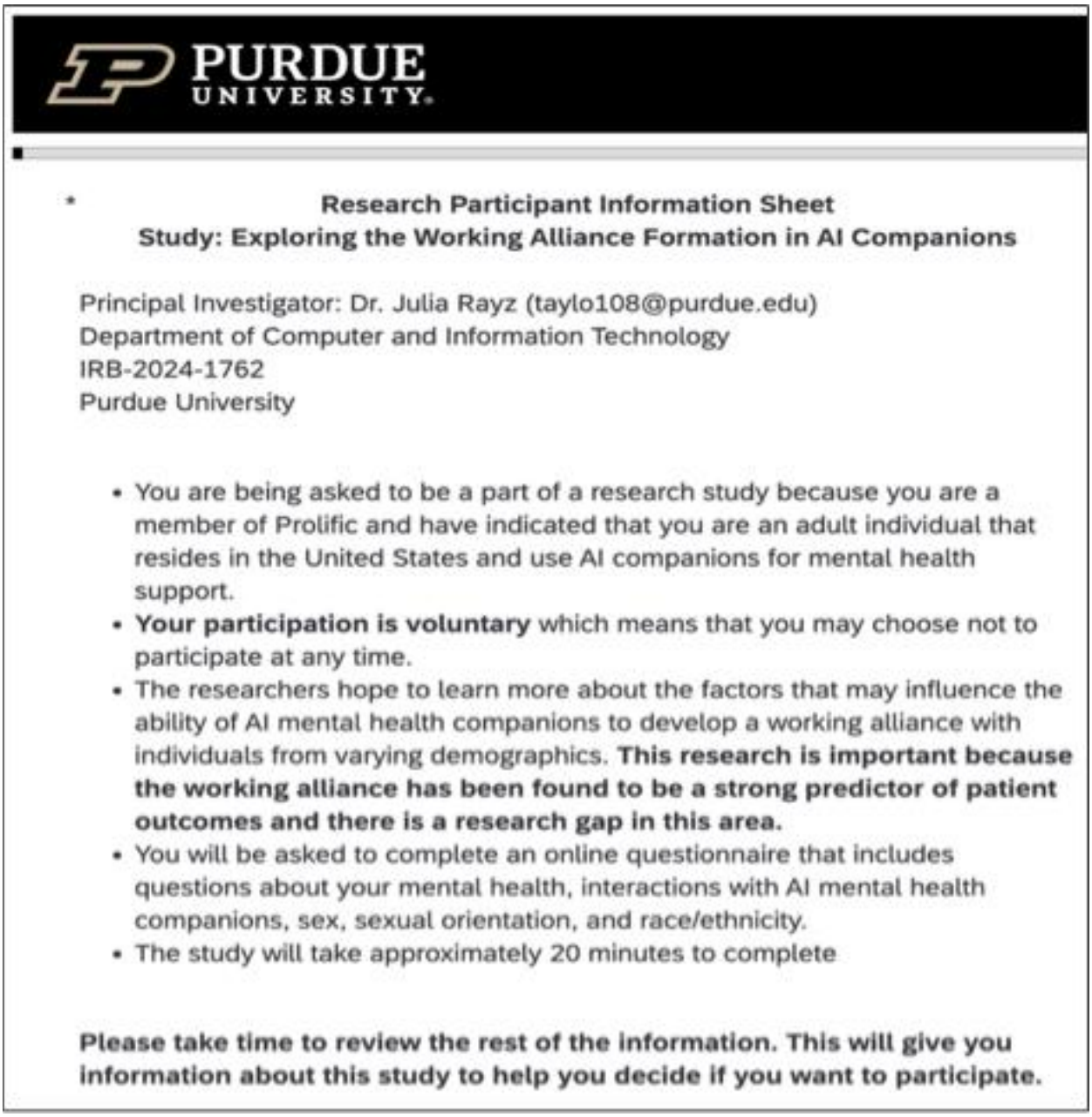
Prolific participant view



Note. Screenshot from Prolific preview.

Figure A3.

Consent form



PURDUE UNIVERSITY

Research Participant Information Sheet
Study: Exploring the Working Alliance Formation in AI Companions

Principal Investigator: Dr. Julia Rayz (taylo108@purdue.edu)
Department of Computer and Information Technology
IRB-2024-1762
Purdue University

- You are being asked to be a part of a research study because you are a member of Prolific and have indicated that you are an adult individual that resides in the United States and use AI companions for mental health support.
- **Your participation is voluntary** which means that you may choose not to participate at any time.
- The researchers hope to learn more about the factors that may influence the ability of AI mental health companions to develop a working alliance with individuals from varying demographics. **This research is important because the working alliance has been found to be a strong predictor of patient outcomes and there is a research gap in this area.**
- You will be asked to complete an online questionnaire that includes questions about your mental health, interactions with AI mental health companions, sex, sexual orientation, and race/ethnicity.
- The study will take approximately 20 minutes to complete

Please take time to review the rest of the information. This will give you information about this study to help you decide if you want to participate.

This study is intended for people who are 18 years of age or older, reside in the United States, and have used AI companions / chatbots for mental health support within the past two weeks.

Before agreeing to participate, consider the risks and potential benefits of taking part in this study.

- All research carries the risk of breach of confidentiality which means that someone outside of our study could figure out that you were in the study or information was yours. How we will protect your information to reduce this risk is below.
- Some questions could make you feel uncomfortable. You can leave the survey at any time.

THIS STUDY IS NOT A REPLACEMENT FOR THERPY OR CRISIS COUNSELING. IF YOU NEED HELP NOW, CONTACT THE CRISIS HOTLINE.

It is unlikely that there will be personal benefits to you for participating. Having more information from the answers in this research study might help us or other researchers understand more about factors that influence the ability of AI to provide mental health support.

Will I receive payment or other incentive?
The survey will take approximately 20 minutes to complete. To thank you for being in our research study, you will receive a $6 payment in consideration for completing the survey. Payment will be processed through Prolific. According to the rules of the Internal Revenue Service (IRS), payments that are made to you as a result of your participation in a study may be considered taxable income.

How will the researchers protect my information, privacy, and confidentiality?
This study does not collect any information that may be used to identify you. Your responses may be shared outside the research study if required by law. We also may need to share your responses with other groups for quality assurance or data analysis. These groups include the Purdue University Institutional Review Board or its designees, and state or federal agencies who may need to access the research records (as allowed by law).

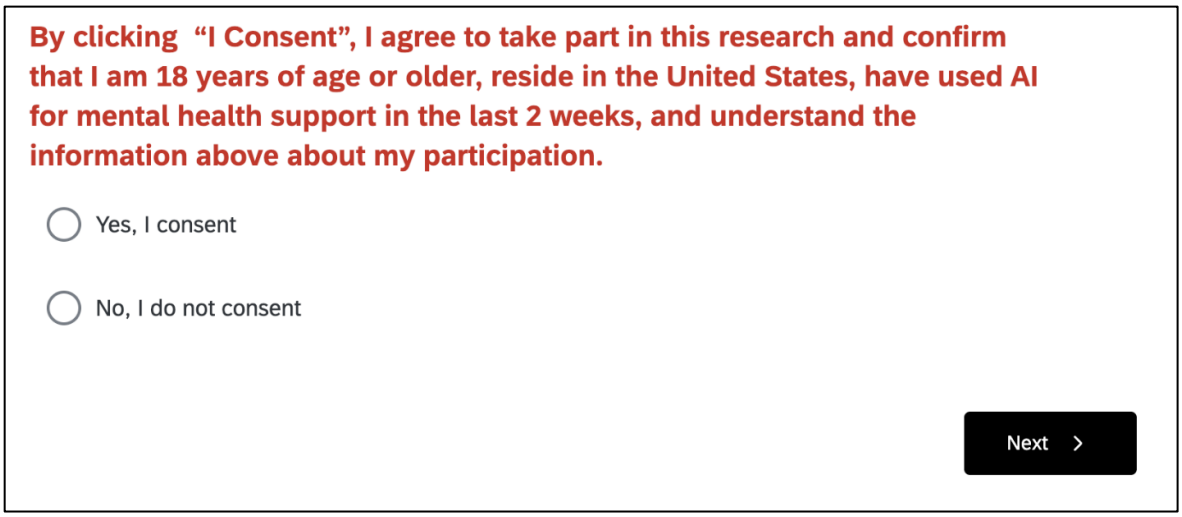
The study team plans to keep answers for this study to answer research questions. We will keep this information until we are done with the study, approximately six months, and for at least three years after we are finished. We may share the anonymous data and findings with other researchers or in research papers or presentations.

What are my rights as a research participant in this study?
You do not have to participate in this research project. If you agree to participate, you may withdraw your participation at any time without penalty.

Who can I contact if I have questions about the study?
If you have questions, comments or concerns about this research project, you can talk to one of the researchers. Please contact Dr. Julia Rayz (taylo108@purdue.edu) or Gerardo Castaneda (castan21@purdue.edu).

To report anonymously via Purdue's Hotline, see www.purdue.edu/hotline

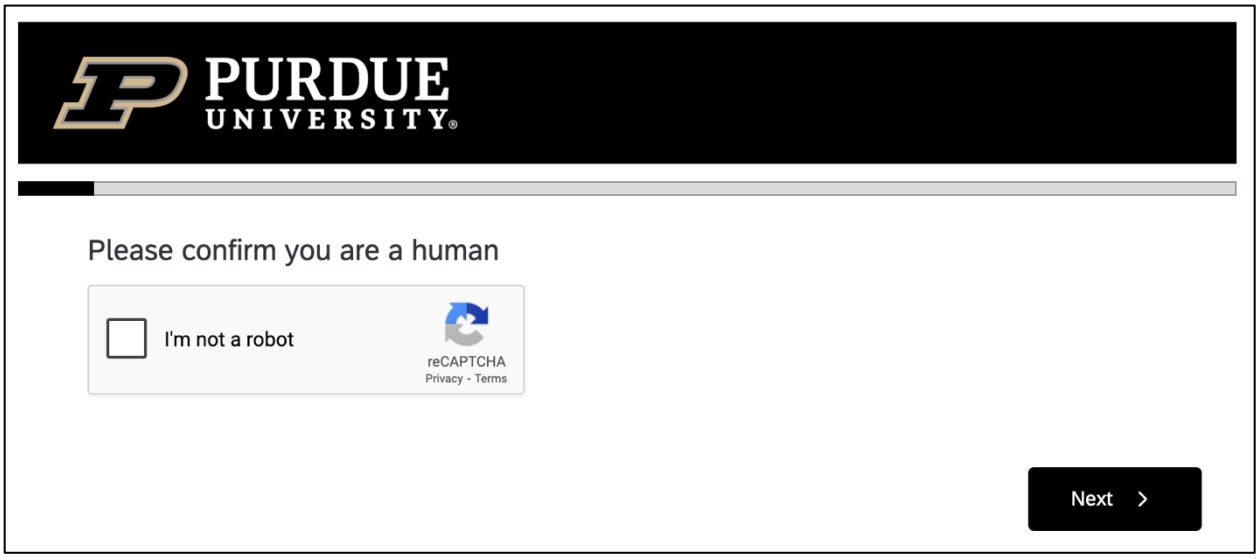
If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu) or write to:
Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1010
155 S. Grant St.
West Lafayette, IN 47907-2114

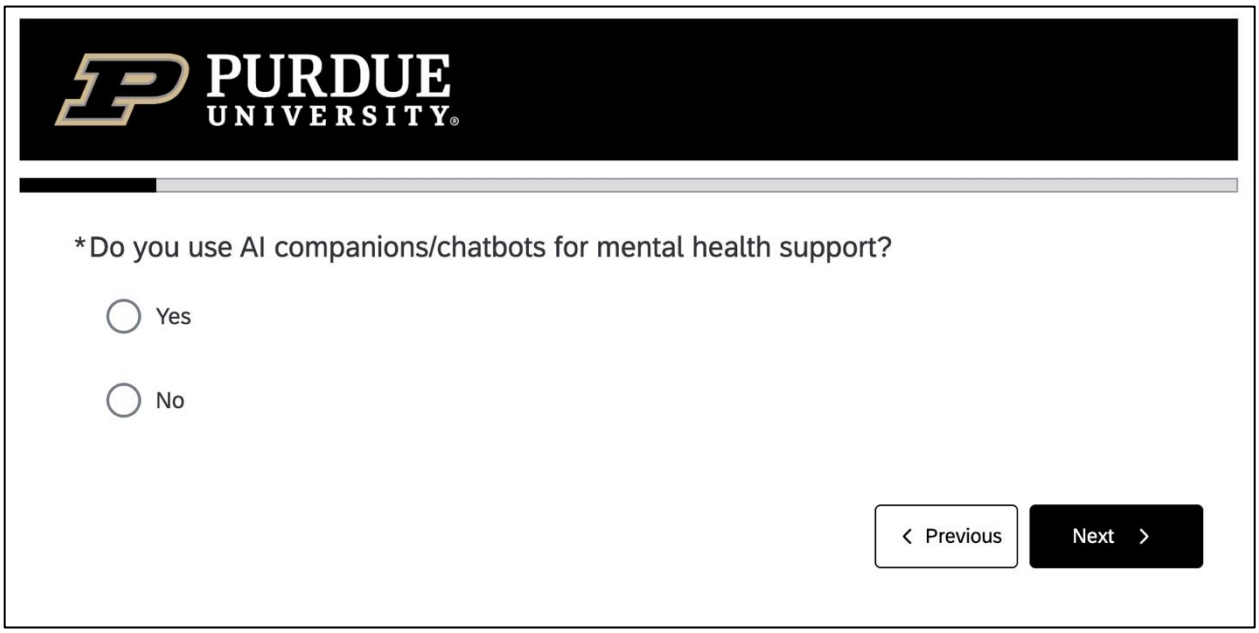


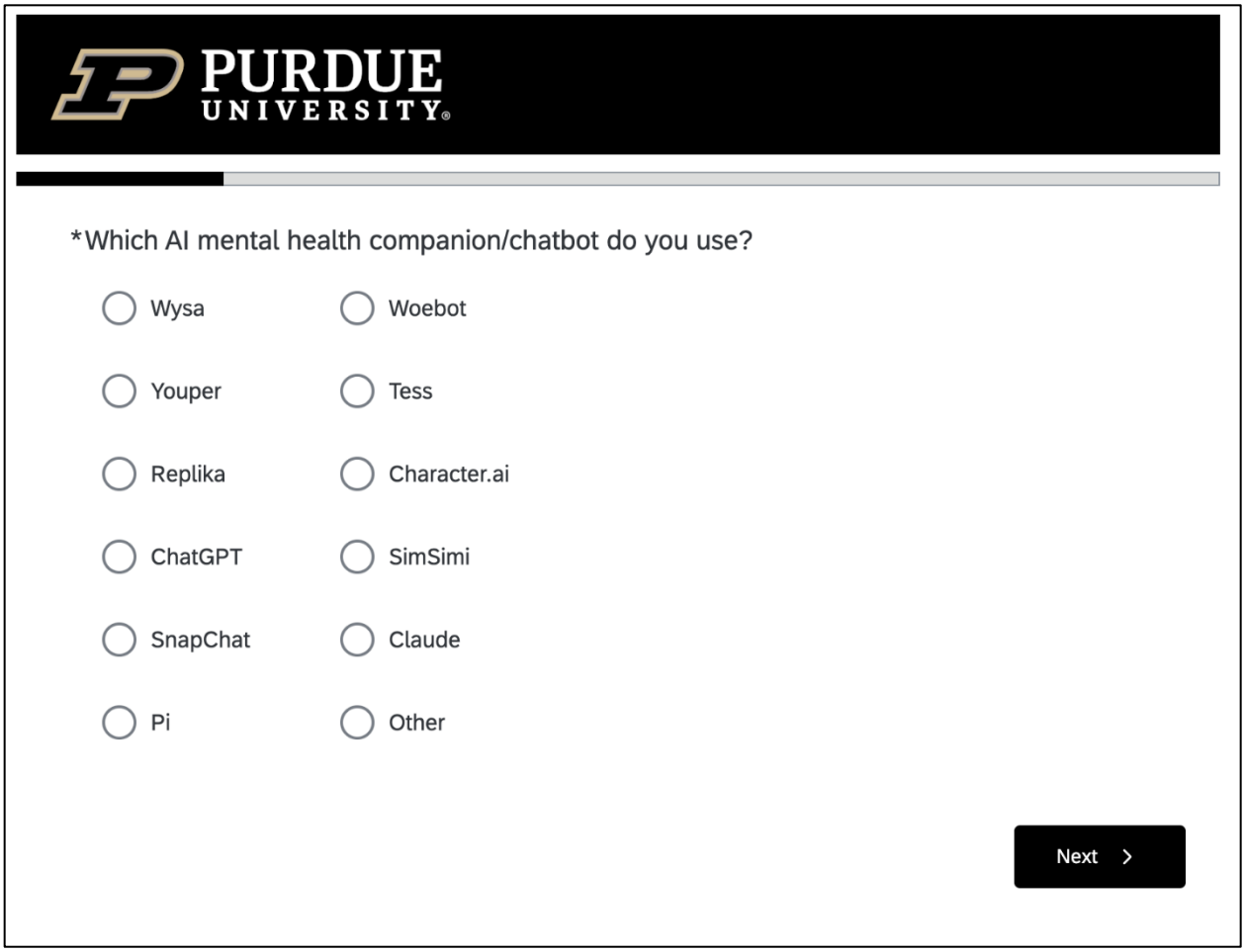
Note. Screenshot of Qualtrics Preview. Content Source: Purdue's Exempt Research Study Information Sheet for Participants (https://www.irb.purdue.edu/docs/new/Exempt%20Info%20Sheet%20v2%202024-08-01.doc)

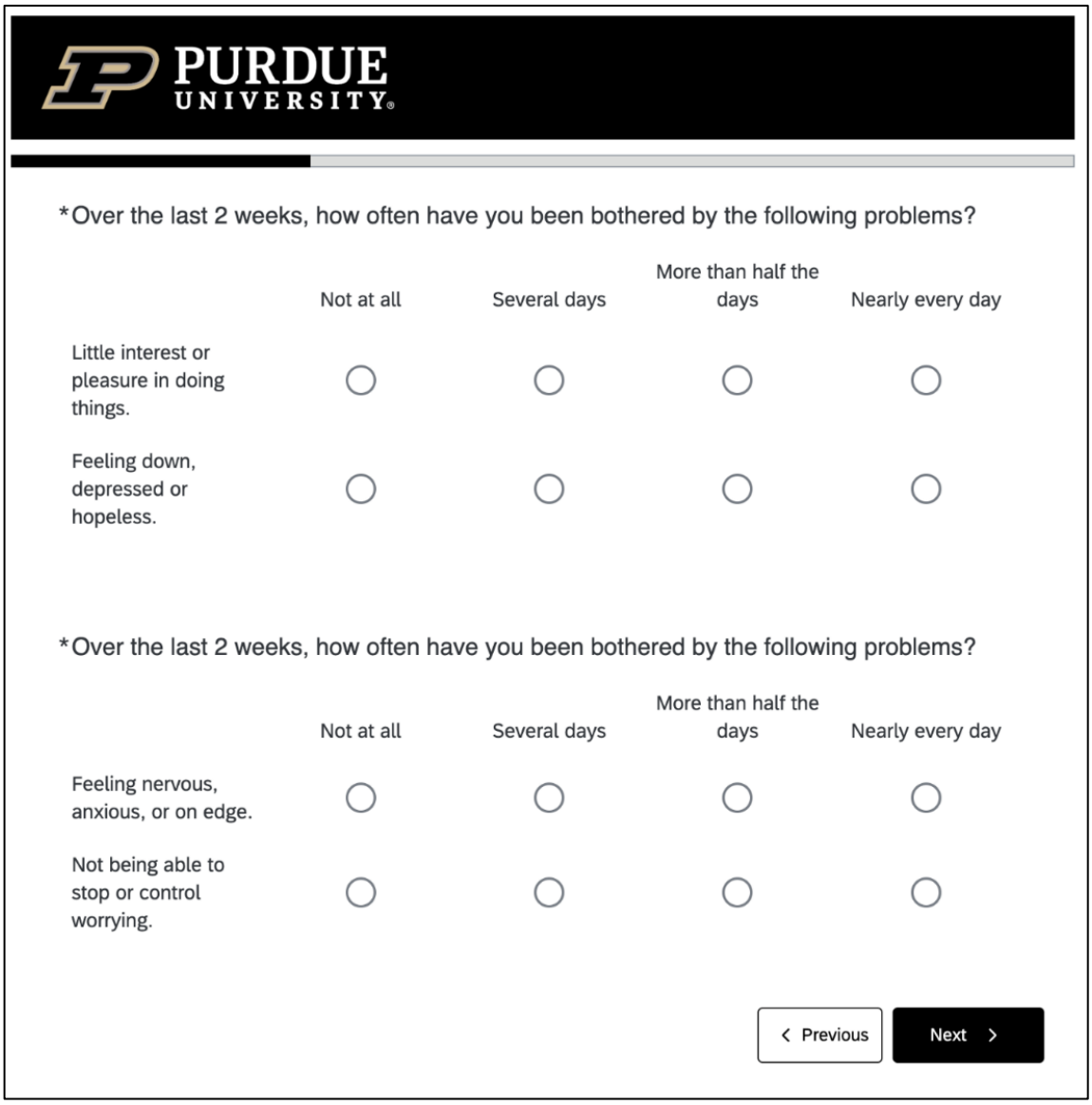
Figure A4.

Survey instrument




PURDUE UNIVERSITY
*Do you use AI companions/chatbots for mental health support?
Yes
No
Previous
Next


PURDUE UNIVERSITY
*Which AI mental health companion/chatbot do you use?
Wysa
Woebot
Youper
Tess
Replika
Character.ai
ChatGPT
SimSimi
SnapChat
Claude
Pi
Other
Next

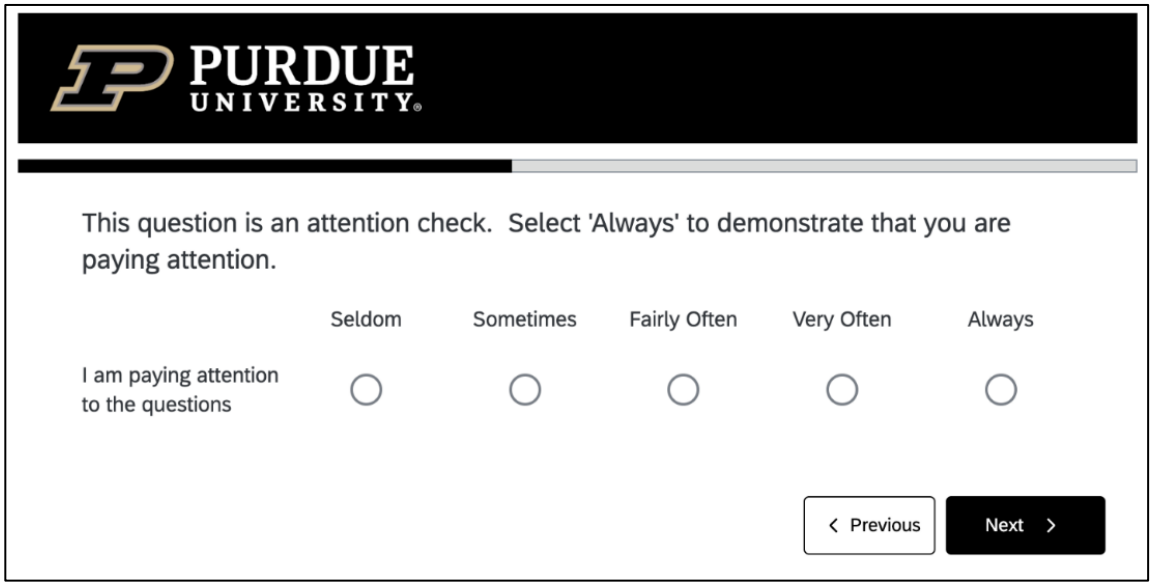

PURDUE UNIVERSITY
*Over the last 2 weeks, how often have you been bothered by the following problems?
Not at all
Several days
More than half the days
Nearly every day
Little interest or pleasure in doing things.
Feeling down, depressed or hopeless.
*Over the last 2 weeks, how often have you been bothered by the following problems?
Not at all
Several days
More than half the days
Nearly every day
Feeling nervous, anxious, or on edge.
Not being able to stop or control worrying.
Previous
Next

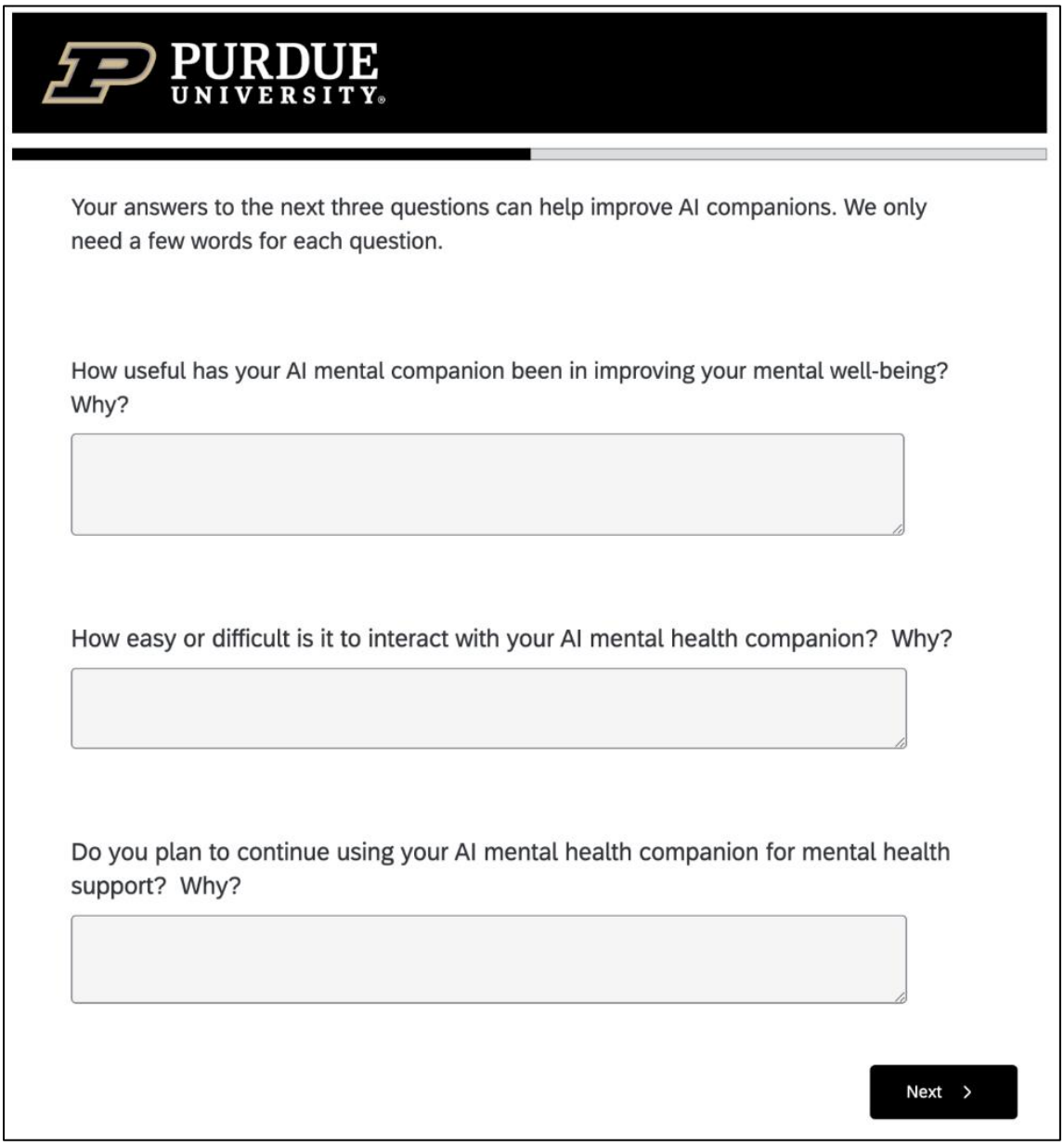
*Take your time to consider the statements below and decide which answer best describes your own experience with your AI companion or chatbot.

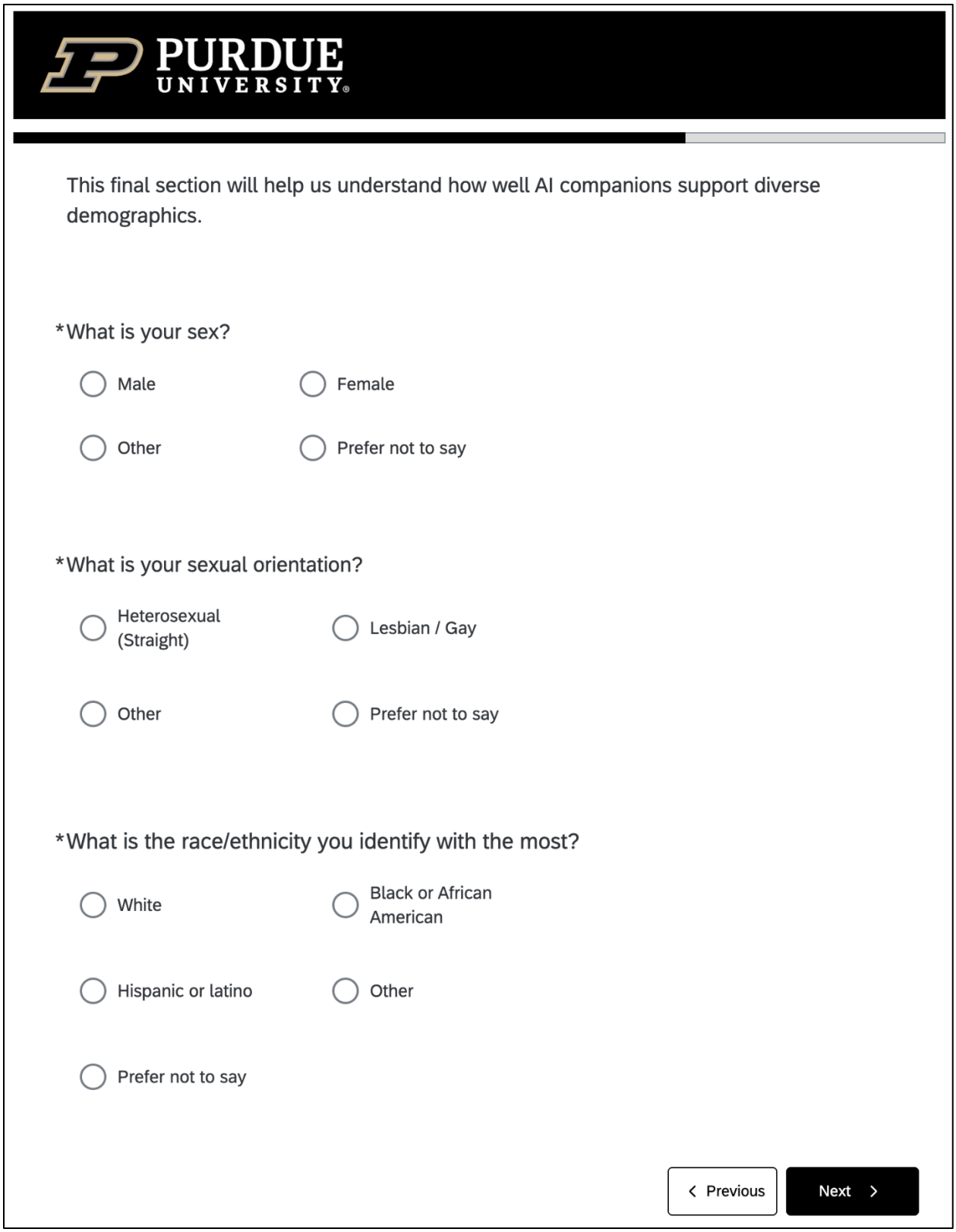
	Seldom	Sometimes	Fairly Often	Very Often	Always
As a result of these sessions, I am clearer as to how I might be able to change.	○	○	○	○	○
What I am doing with my AI companion/chatbot gives me new ways of looking at my problem.	○	○	○	○	○
I believe my AI companion/chatbot likes me.	○	○	○	○	○
My AI companion/chatbot and I collaborate on setting goals for my therapy.	○	○	○	○	○
My AI companion/chatbot and I respect each other.	○	○	○	○	○
My AI companion/chatbot and I are working towards mutually agreed upon goals.	○	○	○	○	○

I feel that my AI companion/chatbot appreciates me.	○	○	○	○	○
My AI companion/chatbot and I agree on what is important for me to work on.	○	○	○	○	○
I feel my AI companion/chatbot cares about me even when I do things that he/she does not approve of.	○	○	○	○	○
I feel that the things I do with my AI companion/chatbot will help me to accomplish the changes that I want.	○	○	○	○	○
My AI companion/chatbot and I have established a good understanding of the kind of changes that would be good for me.	○	○	○	○	○
I believe the way we are working with my problems is correct.	○	○	○	○	○

< Previous Next >


PURDUE
UNIVERSITY
This question is an attention check. Select 'Always' to demonstrate that you are paying attention.
Seldom
Sometimes
Fairly Often
Very Often
Always
I am paying attention to the questions
Previous
Next


PURDUE
UNIVERSITY
Your answers to the next three questions can help improve AI companions. We only need a few words for each question.
How useful has your AI mental companion been in improving your mental well-being? Why?
How easy or difficult is it to interact with your AI mental health companion? Why?
Do you plan to continue using your AI mental health companion for mental health support? Why?
Next



Note. Screenshot of Qualtrics. WAI-SR Items reprinted by permission of the Society for Psychotherapy Research © 2025

REFERENCES

Abbasi, J. (2023). Surgeon General sounds the alarm on social media use and youth mental health crisis. *JAMA: The Journal of the American Medical Association, 330*(1), 11-12. https://doi.org/10.1001/jama.2023.10262

Abbasian, M., Khatibi, E., Azimi, I., Oniani, D., Shakeri Hossein Abad, Z., Thieme, A., Sriram, R., Yang, Z., Wang, Y., Lin, B., Gevaert, O., Li, L.-J., Jain, R., & Rahmani, A. M. (2024). Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. *NPJ Digital Medicine, 7*(1). https://doi.org/10.1038/s41746-024-01074-z

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

Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and opinions of patients about mental health chatbots: Scoping review. *Journal of Medical Internet Research, 23*(1). https://doi.org/10.2196/17828

Abrams, Z. (2021). The promise and challenges of AI. Cover Story. *American Psychological Association, 52(8)*. https://www.apa.org/monitor/2021/11/cover-artificial-intelligence

Akbar, S., Coiera, E., & Magrabi, F. (2020). Safety concerns with consumer-facing mobile health applications and their consequences: A scoping review. *Journal of the American Medical Informatics Association: JAMIA*, *27*(2), 330-340. https://doi.org/10.1093/jamia/ocz175

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). American Psychiatric Association Publishing.

Backe, E. L. (2018). Crisis of care: The politics and therapeutics of a rape crisis hotline. *Medical Anthropology Quarterly, 32*(4). https://doi.org/10.1111/maq.12463

Balcombe, L. (2023). AI chatbots in digital mental health. *Informatics, 10*(4). https://doi.org/10.3390/informatics10040082

Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): A mixed-methods study. *Frontiers in Digital Health, 4*, 847991-847991. https://doi.org/10.3389/fdgth.2022.847991

Blease, C., Locher, C., Leon-Carlyle, M., & Doraiswamy, M. (2020). Artificial intelligence and the future of psychiatry: Qualitative findings from a global physician survey. *Digital Health, 6,* 2055207620968355-2055207620968355. https://doi.org/10.1177/2055207620968355

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy (Chicago, Ill.),* 16(3), 252-260. https://doi.org/10.1037/h0085885

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., 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. https://doi.org/10.1080/17434440.2021.2013200

Bowie-DaBreo, D., Sas, C., Iles-Smith, H., & Sünram-Lea, S. (2022). User perspectives and ethical experiences of apps for depression: A qualitative analysis of user reviews. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3491102.3517498

Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychological Assessment, 15*(2), 193-197. https://doi.org/10.1037/1040-3590.15.2.193

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 wellbeing in different cultures: Mixed methods study. *Journal of Medical Internet Research, 25*(1). https://doi.org/10.2196/51712

Cogburn, C. D., Roberts, S. K., Ransome, Y., Addy, N., Hansen, H., & Jordan, A. (2024). The impact of racism on Black American mental health. *The Lancet Psychiatry, 11*(1), 56-64. https://doi.org/10.1016/S2215-0366(23)00361-9

Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. *JMIR Formative Research, 5*(5). https://doi.org/10.2196/27868

De Choudhury, M., Pendse, S. R., & Kumar, N. (2023). Benefits and harms of large language models in digital mental health. *arXiv.* https://doi.org/10.48550/arXiv.2311.14693

De Freitas, J., & Cohen, I. G. (2024). The health risks of generative AI-based wellness apps. *Nature Medicine*, *30*(5), 1269-1275. https://doi.org/10.1038/s41591-024-02943-6

De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z., & Puntoni, S. (2023). Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology, 34*(3), 481-491. https://doi.org/10.1002/jcpy.1393

Doukani, A., Quartagno, M., Sera, F., Free, C., Kakuma, R., Riper, H., Kleiboer, A., Cerga-Pashoja, A., van Schaik, A., Botella, C., Berger, T., Chevreul, K., Matynia, M., Krieger, T., Hazo, J.-B., Draisma, S., Titzler, I., Topooco, N., Mathiasen, K., … Araya, R. (2024). Comparison of the working alliance in blended cognitive behavioral therapy and treatment as usual for depression in Europe: Secondary data analysis of the E-COMPARED randomized controlled trial. *Journal of Medical Internet Research, 26*(10159), e47515-. https://doi.org/10.2196/47515

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs - principles and practices. *Health Services Research, 48*(6pt2), 2134-2156. https://doi.org/10.1111/1475-6773.12117

Fink, A. (2003). *The survey kit.* (2nd ed.). Sage Publications.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health, 4(*2). https://doi.org/10.2196/mental.7785

Fitzsimmons-Craft, E. E., & Jacobson, N. C. (2024). Eating disorders care and the promises and pitfalls of artificial intelligence. *Missouri Medicine*, *121*(5), 345-349.

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chicago, Ill.), 55*(4), 316-340. https://doi.org/10.1037/pst0000172

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General, 141*(1), 2-18. https://doi.org/10.1037/a0024338

Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *JMIR Mental Health*, *6*(10), e14166-e14166. https://doi.org/10.2196/14166

Giles, W. F., & Feild, H. S. (1978). Effects of amount, format, and location of demographic information on questionnaire return rate and response bias of sensitive and nonsensitive items. *Personnel Psychology, 31*(3), 549-559. https://doi.org/10.1111/j.1744-6570.1978.tb00462.x

Gmelin, J. H., De Vries, Y. A., Baams, L., Aguilar-Gaxiola, S., Alonso, J., Borges, G., Bunting, B., Cardoso, G., Florescu, S., Gureje, O., Karam, E. G., Kawakami, N., Lee, S., Mneimneh, Z., Navarro-Mateu, F., Posada-Villa, J., Rapsey, C., Slade, T., Stagnaro, J. C., Torres, Y., … WHO World Mental Health Survey collaborators (2022). Increased risks for mental disorders among LGB individuals: Cross-national evidence from the world mental health surveys. *Social psychiatry and psychiatric epidemiology, 57*(11), 2319-2332. https://doi.org/10.1007/s00127-022-02320-z

Haque, M. D. R., & Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: Insights from app descriptions and user reviews. *JMIR mHealth and uHealth, 11*. https://doi.org/10.2196/44838

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the working alliance inventory. *Psychotherapy Research, 16*(1), 12-25. https://doi.org/10.1080/10503300500352500

He, Y., Yang, L., Zhu, X., Wu, B., Zhang, S., Qian, C., & Tian, T. (2022). Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: Single-blind, three-arm randomized controlled trial. *Journal of Medical Internet Research, 24*(11), e40719-e40719. https://doi.org/10.2196/40719

Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A. J., Griffin, T. Z., & Jacobson, N. C. (2025). Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI, 2*(4). https://doi.org/10.1056/AIoa2400802

Herek, G. M., & Garnets, L. D. (2007). Sexual orientation and mental health. *Annual Review of Clinical Psychology, 3*(1), 353-375. https://doi.org/10.1146/annurev.clinpsy.3.022806.091510

Hiland, E. B (2018). *The digital transformation of mental health (*Doctoral dissertation). ProQuest Dissertations & Theses Global Closed Collection. https://www.proquest.com/dissertations-theses/digital-transformation-mental-health/docview/2453700229/se-2

Hiland, E. B. (2018). *The digital transformation of mental health* (Doctoral dissertation). University of Minnesota.

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4). https://doi.org/10.1093/joc/jqy026

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. *Nature (London), 633*(8028), 147-154. https://doi.org/10.1038/s41586-024-07856-5

Holden, K. B., McGregor, B. S., Blanks, S. H., & Mahaffey, C. (2012). Psychosocial, socio-cultural, and environmental influences on mental health help-seeking among African-American men. *Journal of Men's Health (Amsterdam), 9*(2), 63-69. https://doi.org/10.1016/j.jomh.2012.03.002

Holden, R. J., & Karsh, B.-T. (2010). The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics, 43*(1), 159-172. https://doi.org/10.1016/j.jbi.2009.07.002

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology, 61*(4), 561-573. https://doi.org/10.1037//0022-006x.61.4.561

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-e12106. https://doi.org/10.2196/12106

Jabir, A. I., Martinengo, L., Lin, X., Torous, J., Subramaniam, M., & Car, L. T. (2023). Evaluating conversational agents for mental health: Scoping review of outcomes and outcome measurement instruments. *Journal of Medical Internet Research, 25*(9). https://doi.org/10.2196/44548

Karkosz, S., Szymański, R., Sanna, K., & Michałowski, J. (2024). Effectiveness of a web-based and mobile therapy chatbot on anxiety and depressive symptoms in subclinical young adults: Randomized controlled trial. *JMIR Formative Research*, *8*, e47960-e47960. https://doi.org/10.2196/47960

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., & Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA: The Journal of the American Medical Association, 289*(23), 3095-3105. https://doi.org/10.1001/jama.289.23.3095

Khawaja, Z., & Bélisle-Pipon, J. C. (2023). Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health, 5*, 1278186-1278186. https://doi.org/10.3389/fdgth.2023.1278186

Kilbourne, A. M., Beck, K., Spaeth-Rublee, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., & Pincus, H. A. (2018). Measuring and improving the quality of mental health care: a global perspective. *World Psychiatry, 17*(1), 30-38. https://doi.org/10.1002/wps.20482

Klinger, D., Oehlke, S.-M., Riedl, S., Eschbaum, K., Zesch, H. E., Karwautz, A., Plener, P. L., & Kothgassner, O. D. (2024). Mental health of non-binary youth: A systematic review and meta-analysis. *Child and Adolescent Psychiatry and Mental Health, 18*(1), 126-18. https://doi.org/10.1186/s13034-024-00822-z

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kundu, S. (2021). AI in medicine must be explainable. *Nature Medicine, 27*(8), 1328–1328. https://doi.org/10.1038/s41591-021-01461-z

Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and wellbeing. *NPJ Digital Medicine, 6(*1). https://doi.org/10.1038/s41746-023-00979-5

Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023b). Supplementary materials: Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and wellbeing [Supplementary material]. *NPJ Digital Medicine, 6(*1). https://static-content.springer.com/esm/art%3A10.1038%2Fs41746-023-00979-5/MediaObjects/41746_2023_979_MOESM1_ESM.pdf

Li, S. H., & Graham, B. M. (2017). Why are women so vulnerable to anxiety, trauma-related, and stress-related disorders? The potential role of sex hormones. *The Lancet. Psychiatry, 4*(1), 73-82. https://doi.org/10.1016/S2215-0366(16)30358-3

Li, X., Gu, H., Zhao, X., Chen, F., & Liu, L. (2022). Development of a measure quantifying helpful psychotherapy interventions: The Helpful Therapeutic Attitudes and Interventions Scale. *Frontiers in Psychiatry, 13*, 1023346-1023346. https://doi.org/10.3389/fpsyt.2022.1023346

Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions: The Application of Information Technology in Mental and Behavioural Health, 27*, 100495-100495. https://doi.org/10.1016/j.invent.2022.100495

Ma, Z., Mei, Y., Long, Y., Su, Z., & Gajos, K. Z. (2024). Evaluating the experience of LGBTQ+ people using large language model based chatbots for mental health support. *arXiv.Org*. https://doi.org/10.48550/arxiv.2402.09260

Manea, L., Gilbody, S., Hewitt, C., North, A., Plummer, F., Richardson, R., Thombs, B. D., Williams, B., & McMillan, D. (2016). Identifying depression with the PHQ-2: A diagnostic meta-analysis. *Journal of Affective Disorders, 203*, 382-395. https://doi.org/10.1016/j.jad.2016.06.003

Martinengo, L., Lum, E., & Car, J. (2022). Evaluation of chatbot-delivered interventions for self-management of depression: Content analysis. *Journal of Affective Disorders, 319*. https://doi.org/10.1016/j.jad.2022.09.028

Martinez-Martin, N. (2020). Trusting the bot: Addressing the ethical challenges of consumer digital mental health therapy. In I. Bárd & E. Hildt (Eds.), *Developments in neuroethics and bioethics*. Elsevier Science & Technology. https://doi.org/10.1016/bs.dnb.2020.03.003

McMaster, K. J., Peeples, A. D., Schaffner, R. M., & Hack, S. M. (2021). Mental healthcare provider perceptions of race and racial disparity in the care of black and white clients. *The Journal of Behavioral Health Services & Research, 48*(4), 501-516. https://doi.org/10.1007/s11414-019-09682-4

Mehta, A., Niles, A. N., Vargas, J. H., Marafon, T., D. D. Couto, & Gross, J. J. (2021). Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (Youper): Longitudinal observational study. *Journal of Medical Internet Research, 23*(6). https://doi.org/10.2196/26771

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129*(5), 674-697. https://doi.org/10.1037/0033-2909.129.5.674

Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M. H., de Pennington, N., Mole, G., Normando, E., & Meinert, E. (2020). The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, *22*(10), e20346. https://doi.org/10.2196/20346

Moagi, M. M., van Der Wath, A. E., Jiyane, P. M., & Rikhotso, R. S. (2021). Mental health challenges of lesbian, gay, bisexual, and transgender people: An integrated literature review. *Health SA = SA Gesondheid, 26*(1), 1-12. https://doi.org/10.4102/hsag.v26i0.1487

Moleiro, C., & Pinto, N. (2015). Sexual orientation and gender identity: review of concepts, controversies, and their relation to psychopathology classification systems. *Frontiers in Psychology, 6*, 1511-1511. https://doi.org/10.3389/fpsyg.2015.01511

Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). *Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers*. https://doi.org/10.48550/arXiv.2504.18412

Munder, T., Wilmers, F., Leonhart, R., Linster, H. W., & Barth, J. (2010). Working Alliance Inventory-Short Revised (WAI-SR): Psychometric properties in outpatients and inpatients. *Clinical Psychology and Psychotherapy, 17*(3), 231-239. https://doi.org/10.1002/cpp.658

National Institute of Mental Health. (2024). *Mental illness.* National Institutes of Health. https://www.nimh.nih.gov/health/statistics/mental-illness

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

Paap, D., & Dijkstra, P. U. (2017). Working Alliance Inventory-Short Form Revised. *Journal of Physiotherapy, 63*(2), 118-118. https://doi.org/10.1016/j.jphys.2017.01.001

Park, D. Y. Y., & Kim, H. (2023). Determinants of intentions to use digital mental healthcare content among university students, faculty, and staff: Motivation, perceived usefulness, perceived ease of use, and parasocial interaction with AI chatbot. *Sustainability, 15*(1), 872-. https://doi.org/10.3390/su15010872

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 And GAD-2: A systematic review and diagnostic meta analysis. *General Hospital Psychiatry, 39*, 24-31. https://doi.org/10.1016/j.genhosppsych.2015.11.005

Prochaska, J. J., Vogel, E. A., Chieng, A., Baiocchi, M., Maglalang, D. D., Pajarito, S., Weingardt, K. R., Darcy, A., & Robinson, A. (2021a). A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug and Alcohol Dependence, 227*, 108986-108986. https://doi.org/10.1016/j.drugalcdep.2021.108986

Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., & Robinson, A. (2021b). A therapeutic relational agent for reducing problematic substance use (Woebot): Development and usability study. *Journal of Medical Internet Research, 23*(3). https://pubmed.ncbi.nlm.nih.gov/33755028/

Russell, S. T., & Fish, J. N. (2016). Mental health in lesbian, gay, bisexual, and transgender (LGBT) youth. *Annual Review of Clinical Psychology, 12*(1), 465-487. https://doi.org/10.1146/annurev-clinpsy-021815-093153

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine (1960), 166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Stacey, L. (2024). An Updated Data Portrait of Heterosexual, Gay/Lesbian, Bisexual, and Other Sexual Minorities in the United States. *Social Currents, 11(5),* 383–400. https://doi.org/10.1177/23294965241260057

Stein, M. B., & Sareen, J. (2015). Generalized Anxiety Disorder. *The New England Journal of Medicine, 373*(21), 2059-2068. https://doi.org/10.1056/NEJMcp1502514

Teclaw, R., Price, M. C., & Osatuke, K. (2012). Demographic question placement: Effect on item response rates and means of a Veterans Health Administration survey. *Journal of Business and Psychology, 27*(3), 281-290. https://doi.org/10.1007/s10869-011-9249-y

Thomeer, M. B., Moody, M. D., & Yahirun, J. (2023). Racial and ethnic disparities in mental health and mental health care during the COVID-19 pandemic. *Journal of racial and ethnic health disparities, 10*(2), 961-976. https://doi.org/10.1007/s40615-022-01284-9

Tidy, J. (2024). *Character.ai: Young people turning to AI therapist bots.* British Broadcasting Corporation. https://www.bbc.com/news/technology-67872693

U.S. Census Bureau (2020). *2020 U.S. population more racially and ethnically diverse than measured in 2010.* https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry, 64*(7). https://doi.org/10.1177/0706743719828977

van de Venne, J., Cerel, J., Moore, M., & Maple, M. (2020). Sex differences in mental health outcomes of suicide exposure. *Archives of Suicide Research, 24*(2), 158-185. https://doi.org/10.1080/13811118.2019.1612800

Viduani, A., Cosenza, V., Araújo, R. M., & Kieling, C. (2023). Chatbots in the field of mental health: Challenges and opportunities. In *Digital Mental Health* (pp. 133-148). Springer International Publishing. https://doi.org/10.1007/978-3-031-10698-9_8

Vogel, D. L., & Heath, P. J. (2016). Men, masculinities, and help-seeking patterns. In Y. J. Wong & S. R. Wester (Eds.), *APA handbook of men and masculinities* (pp. 685-707). American Psychological Association. https://doi.org/10.1037/14594-031

Walker, L. (2023). *Belgian man dies by suicide following exchanges with chatbot.* The Brussels Times. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot

Whaley, A. L. (2001). Cultural mistrust and mental health services for African Americans: A review and meta-analysis. *The Counseling Psychologist, 29*(4), 513-531. https://doi.org/10.1177/0011000001294003

Williams, D. R., González, H. M., Neighbors, H., Nesse, R., Abelson, J. M., Sweetman, J., & Jackson, J. S. (2007). Prevalence and distribution of major depressive disorder in African Americans, Caribbean blacks, and non-Hispanic whites: Results from the National Survey of American Life. *Archives of General Psychiatry, 64*(3), 305-315. https://doi.org/10.1001/archpsyc.64.3.305

Williams, N. (2014). PHQ-9. Occupational medicine (Oxford), 64(2), 139-140. https://doi.org/10.1093/occmed/kqt154

World Health Organization. (2021). *Mental health atlas 2020*. https://iris.who.int/bitstream/handle/10665/345946/9789240036703-eng.pdf?sequence=1

World Health Organization. (2022). *Mental health report.* World Health Organization. https://iris.who.int/bitstream/handle/10665/356119/9789240049338-eng.pdf?sequence=1

Wu, M. S., Wickham, R. E., Chen, S.-Y., Varra, A., Chen, C., & Lungu, A. (2024). A large-scale evaluation of therapeutic alliance and symptom trajectories of depression and anxiety in blended care therapy. *PloS One*, *19*(11), e0313112-. https://doi.org/10.1371/journal.pone.0313112

Xu, Z., & Guangrong, J. (2011). Development of the Working Alliance Questionnaire (Chinese version translation to English). *Chinese Journal of Clinical Psychology*. https://caod.oriprobe.com/order.htm?id=27335289&ftext=base

Zao-Sanders, M. (2025). *How people are really using Gen AI in 2025*. Harvard Business Review. https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025