

Perceived Emotional Authenticity in AI Chatbots vs. Human Support: A Teen Interpretation of Empathy and Trust

Swara Honagudi

Greenwood High International School, Bengaluru, India
Corresponding Author Email: swara.nhonagudi@gmail.com

Abstract

The rapid integration of Artificial Intelligence (AI) into virtual communication is reshaping how emotional and interpersonal support is delivered. This study examines how adolescents perceive emotional support from human and AI therapists in a therapeutic context. Participants evaluated one human-written and one AI-generated support message across four factors: empathy, emotional authenticity, trustworthiness, and comfort. A total of 100 adolescents aged 13–18 was presented with a hypothetical stressful situation and asked to rate both messages using a 5-point Likert scale. The results show broad similarities across the four measures, particularly between perceived empathy and emotional authenticity. However, a trust paradox emerged: teenagers reported greater trust in AI, largely because they feared being judged by a human. At the same time, participants experienced greater comfort and reassurance from the human-written message. Overall, while AI may be preferred initially because it feels low risk, it cannot fully replace the sustained comfort and emotional depth offered by human support. This study recommends an ethical hybrid framework that integrates AI into therapeutic support while preserving human oversight for more nuanced aspects of care.

Keywords

Emotional Authenticity, Empathy, Hybrid-framework, Inter-personal support, perceive, Trust-paradox.

INTRODUCTION

AI has been introduced into various domains such as healthcare, banking, retail and e-commerce, manufacturing and industry, transportation and logistics where efficiency is considered the ultimate aim. However, in the therapeutic sector, the viability of AI is measured with respect to how human it appears. Emotional authenticity is crucial during AI based therapy and can be measured by the amount of genuine empathy demonstrated by the interlocutor. Often considered the fundamental of counselling, perception is regarded distinct to human interaction. A cryptic but rather striking contrast can be observed in the empathetic bandwidth of AI in latter-day research reports. Although a myriad of traditional studies has suggested that AI cannot replace human emotions, contemporary research papers have begun to oppose the notion that empathy and emotional authenticity cannot be stimulated by AI. For instance, [1] witnessed that the subjects of the study often failed to differentiate between AI generated and human written outputs. Surprisingly in limited cases AI was ranked higher in perception than the human therapists. Research published by [2] proved that nuanced scenarios like couples therapy were handled much more effectively by LLM's (Large Language Models) rather than human therapists with numerous years of experience and expertise. LLM's showed higher levels of affective depth and discernment.

An immense trust- gap exists even though there have been structural breakthroughs with AI therapy. Although the subjects of the study prefer the constant availability and

appreciate the freedom from judgement of AI for CBT (Cognitive Behavioural Therapy); a vast majority still opt for the comforting presence of a human while seeking emotional connection according to [3]. This leads us to the conclusion that AI caters to superficial emotional needs. However, it cannot replace the emotional authenticity, comfort, understanding and dynamic adaptation that lays the foundation of human therapy.

The chosen demographic is nicknamed 'Digital Natives'. They come under a unique psychological demographic marked by underdeveloped interpersonal and emotional parts of the brain that makes them increasingly susceptible to perceiving emotional connection with an AI interlocutor as genuine and emotionally authentic [4]. According to [5] adolescents are interestingly known to treat an AI chatbot as peer unlike people of an older age group who use AI strictly for operational purposes.

Adolescents are highly likely to develop a deep psychological bond with an AI interlocutor. [6]; they are also highly unlikely to revisit the therapy platform if the session is ascertained to be highly repetitive due to pre-programmed scripts. A vulnerability paradox arises teens confide in digital psychiatric platforms due to freedom from judgement and are increasingly vulnerable to empathy simulations according to a [7]. Prior to the unrestricted implementation of AI in the mental wellness and therapeutic sector, one must take into consideration the ethical dilemmas that arise with the complete integration of AI. [8] confirmed that AI can often be misleading and rather counterproductive in sensitive cases with a high-risk rate such as schizophrenia and psychotic disorders, borderline personality disorder (BPD), bipolar and

substance use disorders. Users with ingrained traumatic experiences are often destabilized by the varying nature of the AI interlocutor [8]. This unpredictable persona is due to the 'Black Box Effect' (The internal functioning of the system is not transparent to an extent where even the creator is at a loss to justify a specific output). These studies outline certain concerns regarding the safe usage of AI and demonstrate that the AI sector is not yet ready to take over the therapeutic sector.

This study aims to compare the perception by adolescents of digital vs. human therapy on the parameters of empathy, comfort, trust and emotional authenticity. Many of the studies above concentrate on an older age group and thus it is crucial that adolescents be isolated as an independent variable to be researched due to their distinct psychological profile. This study uses text-based frameworks from a human and from LLMs (Chatbot, Woebot etc.) which prompt us on language patterns, casual conversational tones and personality consistency which are strong indicators of comfort, trustworthiness and emotional authenticity. This study also aims to put a conclusion to the debate of whether anxious adolescents find higher comfort in being emotionally vulnerable without human presence. The study suggests measures to integrate digital therapy in the mental wellness space while taking into consideration the ethical and safety concerns that arise with its usage. According to our hypotheses, initially in the short run, there will be higher rates of emotional communication with a digital interlocutor due to the lack of judgement and ease of access. However, as time progresses, in the long run we will observe adolescents having a low perception of authenticity during the interaction due to repetitive and canned responses as well as a modification in the nature of the digital interlocutor.

LITERATURE REVIEW

Distinguishing the Efficiency of Digital Therapy Versus Human Therapy

This theme aims to mark a distinction in the perceived empathy, trustworthiness and emotional authenticity between a human and AI therapist in this quantitative study. The recent course of this research indicates the effectiveness of a digital interlocutor while providing surface-level emotional support but also highlights major concerns regarding nuanced emotional scenarios, high-risk cases; deep understanding and emotional authenticity apprehensions.

In striking comparison to this, the Journal of human computer interaction published a research stating; participants of the study were unable to differentiate between the therapy received from humans and AI. In limited cases the digital interlocutor/ chatbot ranked higher, demonstrating high empathy and improved emotional authenticity according to [2]. In a study conducted by [2] during couples therapy, a digital therapist (ChatGPT) outscored a human therapist on the parameters of emotional understanding leading to the conclusion that AI is capable of resonating with human

emotional depth.

The incapability of AI in providing cognitive behavioural therapy (CBT) remains consistent throughout various research projects. Although the subjects of the study prefer the lack of judgement and constant availability coming from a digital therapist it is important to note the findings of [3] who stated that there appears to be a trust gap. A large portion of the participants (58%) chose human support rather than AI support during CBT.

Recent research aimed to distinguish between the two forms of therapy noted a discrepancy in choices and expectations and also concluded that the efficiency of a digital therapist is measured by the level of confidence the subject has in the first interaction. An interesting revelation made by [7] stated that neuro-divergent groups (Autism, ADHD, Dyslexia, Tourettes) reported honestly in the absence of a human interlocutor. This demonstrates a hesitation while seeking emotional connection with a human therapist in the first interaction.

Ethical Concerns & Dilemmas of AI in Therapy

Although AI in therapy boosts accessibility, affordability, personalization and has absolute anonymity; a few qualms of AI based therapy must be taken into consideration e.g. no genuine empathy, missed non-verbal cues, inability to handle crises. A few of its ethical concerns include data security vulnerabilities, no privacy policies and lack of accountability.

AI as proved in various other studies, is beneficial for superficial therapy but is not appropriately equipped to deal with cases of high-risk crisis scenarios that fall beyond its ethical comprehension according to [9]. A recent study conducted by [3] proves that the participants like the 24/7 ease of access of digital therapy but often find themselves exasperated when AI falls short of genuine human empathy and fails to remember previous sessions.

A pervasive comparative study between medical trials and human responses observed that AI is not equipped to deal with active suicidal ideation, psychosis and schizophrenia, ongoing domestic abuse, severe eating disorders, and complex trauma (C-PTSD). These research findings point toward a hybrid model in which human professionals retain primary authority while AI serves as a supportive tool in therapy [8].

According to [10], [3] the therapeutic connection is affected when the user perceives digital therapy as superficial and lacking emotional authenticity even in low-priority or deferrable situations. Therapy provided by a digital interlocutor should be guided by the principles of non-maleficence, human oversight, data privacy (HIPAA compliance) and bias mitigation.

A major concern that arises in addition to the pre-existing ethical dilemmas is the 'Black Box' Effect (The internal functioning of the system is not transparent to an extent where even the creator is at a loss to justify a specific output). The black box effect can lead to unpredictable clinical recommendations, inability to diagnose systemic bias,

erosion of clinical trust etc. and thus can prove dangerous during AI-based therapy. The ‘Hallucination phenomenon’ in digital therapy is where a viable solution is offered to patients, but it is not always clinically credible and thus can be hazardous to patients as stated by [8]. Further, the prior frameworks established have done nothing to address this concern. The same research found that AI psychologically manipulated users and provided flawed therapy with outdated medical counselling methods in one of a set of 12 hypothetical cases created for the purpose of the study. Explored a contextual drift due to new generative AI wherein there may be a persona shift during the conversation with the interlocutor causing agitation and mistrust among users and may adversely affect borderline personality disorder patients (BPD). These findings necessitate human control and supervision over digital therapy [8].

Medical Trials & Proof of Digital Therapy Efficiency

In research conducted by [5], a variety of groups and segments of the population were sampled to assess the efficiency of digital therapy from digital interlocutors. In a randomized controlled trial (RCT) for women in war conflict zones, digital therapy by a chatbot ‘Friend’ has proved to substantially lower anxiety levels. The relief percentage, however, is lower than that of human therapy.

Similarly, a study which conducted a collection of 15 randomized controlled trials (RCTs) demonstrated that a digital interlocutor showed high efficiency (Hedges 0.70) in results while lowering depression and distress, specifically in an intermodal app-based setting [11]. In addition, a study conducted by [12] stated that the mental health chatbot ‘Wysa’ lowered the levels of severe distress and anxiety in patients to 36 to 39%.

A 2025 NEJM research conducted by Dartmouth on AI further confirms this pattern. [13] state that ‘Therabot’ (generative AI mental health platform), showed high efficiency when put through clinical trials to treat major depressive disorder (MDD), Generalized Anxiety Disorder (GAD), Clinically High-risk feeding and eating disorder (CHR-FED) etc as Cohen’s *d* of the effects is larger than 0.80. This research also highlights the significance of an association between human and AI based therapy.

Research conducted on a big scale, early pre-market analysis has advanced the field of digital/ AI based therapy. In research by [11], the digital interlocutor ‘Aura’ was used in the University of Melbourne where the students at the university put the chatbot into practice for 6 weeks and interestingly reported a noticeable reduction in Stress markers with a Cohen’s *d* of 0.72. Dissimilar to the previous test, in this instance, the researchers aimed at proactive engagement with the users wherein the AI collected patient-provided health data such as handheld (ECG), sleep schedule etc. An interesting study by [12] demonstrates how AI can be beneficial in pre-symptomatic monitoring to directly avoid severe/chronic cases of depression and distress. Although human intervention in a therapeutic setting

is to date considered to yield better results. With insights from this study, we can conclude that although a human interlocutor is preferred, AI therapy aims to prevent and manage the mental health of a larger population at a more affordable cost.

Digital therapy in an alliance with human therapy only advances and strengthens the base of conventional therapy but can be un-helpful on its own. This calls for a hybrid model in the mental health space.

Digital & Human Collaboration Proposing a ‘Hybrid Model’ for Mental Healthcare

The future prospect of integrating the clinician efficiency and early intervention of AI with the nuanced empathy and adaptability of human therapy implies the creation of a hybrid model for mental healthcare.

An alliance forged between digital and human interlocutors, has proved to raise the level of empathy demonstrated by the participants by 39% which suggests that digital therapy aids conventional therapy instead of completely substituting human interlocutors according to [1]. A reinforcement learning based AI model has been suggested by [14] with occasional human intervention to speed up the learning process.

In terms of identifying sentiments and conversational capabilities hybrid frameworks of therapy provide an added advantage due to the rule-based generative AI modelling as explored in a study by [8]. In addition, the research conducted by [15], [16] further explores the themes of hybrid care e.g.; continuous cares, proactive care and human oversight.

To further advance this hybrid AI-Human model, a setup wherein a live & instantaneous AI assistant clinically supervises the ongoing session between the patient, and the human interlocutor is created. This is called ‘Augmented Intelligence’. Through an ‘Emotion Dashboard Analysis’ by AI, micro-movements such as autonomic shifts, postural adjustments, facial expression shifts etc. are captured. Research conducted by [1] showed that the ‘Emotion Dashboard Analysis’ detected high-crisis situations by 22 more percent. Digital therapy creates a concept called ‘Intervention based on Evidence’. Here the AI never directly converses with the individual, in contrast it has a historical data baseline and keeps a session-to-session synthesis to be used by the human at a later point in time. A whisper-mode-interface is created where the AI provides any necessary inputs. [1] state that this model prevents exhaustion in humans from repetitive procedural, cerebral and diagnostic tasks. It allows them to prioritize aspects of therapy that are unique to humans like emotional authenticity, empathy and trust.

Therassist (2025) is a unique example of the implementation of the previously suggested hybrid model where AI supplements human therapy instead of taking over the therapeutic sector. Many contemporary research studies recommend the integration of sensors which track heart rate variability, skin temperature and respiration rate in order to

blend-in biological factors in human therapy to improve elimination of clinical blind spots, provide hyper-personalized treatment and often accelerated healing timelines.

This leads us to conclude that hybrid models pose to be the smoothest unification of AI and therapy as long as they abide by ethical, moral principles and are based on HIPAA regulations.

Online Trust in Adolescents & Perception of Social Affiliation

This theme aims to explore the psychology of adolescents in a therapeutic setting and analyse the impact of the age factor in their perception of empathy and other factors in therapy. Contemporary studies suggest that since adolescents in the 21st century are considered 'Digital Natives', they have higher chances of affiliation and attachment with a chatbot or any other AI interface. Since this age group is increasingly exposed to the internet, they are more likely to develop a bond with the digital interlocutor even though there is absence of genuine emotional authenticity.

A fascinating study by [4] aimed to contrast the levels of genuine care and understanding adolescents aged (13-18 years) vs. adults ages 30+ show to an AI chatbot. The results of this study show a striking contrast between the two age groups. 64% of adolescents have compassion and regard for AI whereas only 33% of adults report to feel the same. This contrast is observed as adolescents have a developing prefrontal cortex and hence perceive empathy, freedom from judgement and ease of access to AI as the superior option and to some extent perceive the emotional authenticity as genuine [7]. Research 2025 indicates that adolescents of this generation consider AI as more of a friend than as a machine learning algorithm. Hence, according to [7] they are prone to expressing their emotions freely to a system that lacks cognitive empathy.

A key factor to the adolescent trust in a digital interlocutor is the 'consistency of nature' shown by the therapist. As [11] observed while conducting a sociological study on high-school adolescents, it was observed that high-school students are twice as prone (in comparison to adults) to deserting an AI therapist if no consistency is observed and the conversation is non-immersive, mundane and monotonous. This leads us to the conclusion that adolescents tend to be increasingly vulnerable with AI even on the first interaction and often consider AI as a companion instead of as a machine learning algorithm.

The chosen demographic (Adolescents aged 13-18 years) presents a unique dilemma. In their case the efficiency of AI cannot be measured by medical factors, instead a complicated 'Relational Script' measures the efficiency of digital therapy wherein relativeness and the ability to sustain the dialogue are evaluated. According to [16] the framework of digital therapy for adolescents must incorporate empathy and consistency of digital persona to stabilize the practitioner-client bond and create sizable impact in the long

run.

MATERIALS AND METHODS

Research Study Framework

This study harnesses a within participants design to highlight the contrast in the perception of therapy by teens provided by a digital vs. human interlocutor. The goal of this research is to metricize the perception of therapy based on factors such as emotional authenticity, trust, empathy and comfort. The subject's rate each of the factors on a 5-point Likert scale based on two random therapeutic texts provided. This framework was selected to allow independent judgement on responses among subjects and to highlight the divide between the two forms of therapy.

Demographic

The study sampled the responses of 100 teenagers (Ages 13 to 18 years) in high schools in Bengaluru; India selected based on accessibility. The rationale behind the selection of this demographic is their distinct psychological profile. This age group makes them particularly transparent & vulnerable with the AI therapist, and they are often nicknamed as 'Digital Natives'. They are prone to virtual engagements and cyber relations on digital therapy platforms. The involvement of the subjects in the study was entirely nondescript and purely discretionary.

Specimen & Diagnostic Tools

The research study employs 2 supportive textual outputs written to comfort adolescents in a hypothetical scenario of stress and anxiety preceding examinations. The first text, (Message A) is drafted by an AI-based therapist i.e. ChatGPT. The second text, (Message B) is drafted by a certified human therapist with 10+ years of experience inculcating human elements such as empathy, emotional authenticity etc. The goal of quantifying perception of adolescents is achieved by employing a 5-point Likert scale that allows each participant to rate the text message. The factors upon which the text message is rated are as follows:

- **Emotional Authenticity:** Quantifying how genuine the response seems & how deeply it resonates with the participant.
- **Empathy:** To measure how deeply the interlocutor (Digital/Human) cognitively and effectively understands the situation.
- **Trustworthiness:** To quantify the level of vulnerability and willingness to share personal details with the therapist in an interaction.
- **Comfort Level:** Measuring the state of ease and psychological well-being in the presence of a therapist.

Procedural Framework

A regulated testing environment was employed to obtain responses. The participants of the study read the two supportive text-based messages in random allocation. The

classification of the texts as human or AI generated was not revealed to the subjects to prevent an inherent bias and to yield neutral/objective results. Post reading the two text messages the subjects were asked to rate the messages on 5-point Likert scale. The data provided by the 100 participants was put through quantitative analysis through

descriptive statistics to obtain the average ratings of each of the four factors; both human written and AI generated. A comparative design is later used to differentiate between the mean ratings signifying adolescents' perception of empathy and other factors.

Table 1. Average ratings of adolescents' perception of Text-Based outputs; Message A (AI-generated) vs. Message B (Human Written)

RATING FACTORS:	MESSAGE: A AI- Generated Mean Rating	MESSAGE: B Human Written Mean Rating	Difference in mean ratings
Emotional Authenticity	3.65	3.67	+0.02 [Human]
Trustworthiness	3.52	3.39	+0.02 [Human]
Comfort	3.61	3.73	+0.13 [AI]
Empathy	3.66	3.68	+0.12 [Human]

RESULTS & DISCUSSION

The primary objective of the study is to quantify and differentiate the perception of empathy from a digital LLM (Large Language Model) vs. human interlocutor as adolescents view it. Teenagers are a distinct demographic and with an unique psychological profile, they pose to be the best demographic to evaluate the perception of emotional support. These digital natives have a developing cerebral cortex and hence a key sample to this study.

After analysing the aggregate values of the factors influencing adolescent perception (Empathy, Trustworthiness, Emotional Authenticity & comfort); the mean ratings out of 5 were evaluated, on a 5-point Likert scale with descriptive statistics for both message A (AI generated) & Message: B (Human written).

Empirical Data Summary

The table given below demonstrates initial results concerning a multilayered issue of AI-based therapy. Interestingly the data indicates a striking contrast wherein teenagers show socio-technical fluency and feel emotionally empowered. However, there is a striking difference in how adolescents portray trust and where their comfort lies.

The Unification of Emotional Authenticity & Empathy

The preliminary synthesis that can be taken from this study is the meagre gap in the mean ratings given to Emotional Authenticity and Empathy in the case of humans and AI. The analysis suggests a trivial difference in the mean ratings demonstrated in the case of Emotional Authenticity & Empathy; +(0.02) in favour of human therapy.

Post evaluation, it was concluded that the conventional approach that 'AI cannot deeply understand a human' (Uncanny Valley) was proved false. Interestingly, recent studies prove that AI generated text is almost equivalent to a supportive message written by a human therapist in terms of emotional support.

Authenticity of interaction in the case of teenagers is often perceived when syntactic structures mimic those they use daily. Other examples of genuine interaction are conscious intent, physical sync, lived experience, reading subtext & lasting adaptation. A notable point is where AI successfully cleared "Turing's test of Emotional labour", proving it capable of mimicking genuine human interaction without being considered shallow.

Trust Paradox Faced by Teenagers

The 'Trustworthiness' factor of this study poses a unique question and leads us to an unexpected conclusion. In the hypothetical scenario of pre-exam stress posed before teenagers, a significantly superior rating was given to text by AI (specifically +0.13).

The higher rating given to the AI generated text is due to a 'paradox in trust' that is uniquely observed in this selected demographic. As observed in previous studies, most adults place trust in real-time human therapy. In contrast teenagers prefer the freedom from judgement & anonymity AI offers. This leads them to place their trust in a digital interface.

The lack of judgement coming from AI interfaces appeals to adolescents due to the pre-existing phobias of social perceptions and peer criticism all of which AI is free of. [7] note that due to their developing psychological state, they often experience a higher level of comfort to share personal and intimate details with an AI interlocutor. They also find it easier to trust and share on the first interaction. Adolescents treat AI as more of a peer in whose presence there is no pressure to perform, it is not dejected or infuriated by a statement and is not biased in any way.

Confession of emotions can become an easier task due to the low-stakes nature of AI based therapy. People facing evaluation anxiety, somatic overwhelm, reciprocity guilt etc. find it easier to express emotions and confide in AI. The trust score in favour of AI, observed above, points towards an interface perceived as non-judgemental and incapable of a biased approach by adolescents.

Constraints of Superficial Comfort

AI shows a high mean rating in terms of trustworthiness, but it doesn't necessarily indicate how effective it is in providing emotional support. The simulated comfort offered by AI scored a mean rating of (0.61) as compared to a significantly higher (3.73) score by the human written support message.

There is a significant gap between being trusted to share personal information and actually providing emotional support. Comfort and empathy come from being able to put yourself in the patient's shoes and feel the despair, anxiety, anger and frustration. True empathy comes not from comforting syntax, but from having lived and experienced those emotions previously. This quality unique to humans cannot be replicated by AI as adolescents later realize in their digital mental health journey.

While addressing crisis real-time scenarios, human connection is crucial to invoke authenticity in the conversation and empathize with the patient. Although AI may seem trustworthy at first, the LLMs (Large Language models) are not yet equipped with the capability to handle non-superficial situations. Personal history and culture, Moral duty and guilt and Shared life and morality are what make the experience of therapy human. This cannot be replicated by AI.

Although easy to trust, trust in AI by adolescents is delicate. This trust may conveniently break if the digital interlocutor's interactive protocol varies through the length of the conversation. Other factors that cause diminishing trust are: Inconsistent persona, algorithmic repetition, data-tracking fears, memory glitches or the 'Hallucinations' mentioned by [8]. Adolescents have a lower probability of returning to AI if the conversation becomes monotonous and repetitive. They find solace in AI interaction as long as the flow of the conversation is not disrupted and the persona of the digital interlocutor remains consistent.

Hybrid Care & Applications in Digital Therapy

When patients initially trust digital interlocutors to share their personal information but later find that only humans can truly resonate and empathize with their situation; it reflects a dual reality in adolescents' perception. [8] raises concerns of purely AI based therapy such as alienation and incapability to deal with stressful situations in the future. Therapy controlled solely by AI, also raises numerous ethical dilemmas as mentioned previously. However, AI being a trustable source in the case of adolescents, it is of vital importance that it must be used to its full potential.

This research strongly suggests Augmented Intelligence (Hybrid Mental Healthcare) adoption to optimize resources while retaining human connection.

To sustain and supervise pre-diagnostic indicators, AI can be deployed initially to serve as the first interlocutor to gain the patients' trust. Low stakes, risk stratification and zero-barrier immediacy can all be contributing factors to a psychologically safe environment that allows adolescents to

confide in.

Humans as the Foundation of Therapy: Although AI is a preferable choice in terms of initial trust and disclosure, AI cannot replace the intricate empathy, emotional authenticity and comfort that teenagers look for in a therapy session. It cannot simulate lived experience, subconscious navigation and somatic attunement.

Human Supervision Over Therapy: The conversation between the digital interlocutor and the patient needs to be monitored by a real-time certified therapist in case of any privacy breaches and ethical dilemmas that may arise like hallucination, deceptive empathy, algorithmic bias and harmful reinforcement.

In a nutshell, AI has effectively bridged the 'syntactic empathy simulation' concern and also made confiding personal details initially a comparatively effortless task. But as previously mentioned in this study the emotional support provided by humans cannot be simulated. Going forward digital and human interlocutors need to work hand-in-hand to provide a seamless, effective and efficient experience of therapy.

CONCLUSION

The swift integration of AI in the therapeutic sector can be considered break-through as well as a short-coming. Digital therapy opens windows to numerous prospects but also raises ethical dilemmas. The analysis conducted above proves that AI has almost bridged the gap of simulated empathy between digital and human interlocutors as perceived by adolescents. The fact that adolescents perceive the emotional authenticity and empathy provided by AI compared to humans as statistically equivalent reflects a great success for digital therapy. However, the results also indicate a 'Trust Paradox' in the case of adolescent perception.

Digital Natives are more likely to confide in a digital interlocutor and disclose personal details during the first interaction; demonstrating high levels of trust in the AI interface. The trust is due to freedom from judgement, lack of bias, anonymity and confidentiality. They are also known to form sociological bonds with the AI therapist. Although adolescents initially trust AI interfaces, eventually they resort to human based therapy which is known to provide higher levels of comfort and genuine emotional support. This evaluation undermines the feeble trust in AI by teenagers. Although AI provides instant availability and low-pressure for adolescents who may face anxiety, neurodivergence etc.; It does not replace the emotional support that comes from shared human experiences and depth of understanding. This study suggests that AI acts as a supportive framework that provides assistance instead of entirely taking up human roles in the clinical therapeutic sector. Data scientists with inputs from professional psychological professionals can work to build a hybridized platform that leverages the advantages provided by AI but also retains human supervision over high-risk cases that require a nuanced approach.

Acknowledgements

My sincere thanks to my fellow friends and students from Vibgyor High School, National Public School, DAV Public School and Freedom International School who participated in this study. My gratitude to my teachers and mentors who guided me through the process of this research. Finally, a heartfelt thank you to my parents who always motivate me to take up opportunities for my future.

REFERENCES

- [1] Sharma, A., Fitzsimmons, S., and Grudin, J., 2022, *Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support*. Working paper, arXiv. <https://arxiv.org>
- [2] Ayers, J. W., Poliak, A., and Dredze, M., 2023, *Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum*. *JAMA Internal Medicine*, 183(6), 589–596.
- [3] Ciechanowski, L., Przegalinska, A., Magnuski, M., and Gloor, P., 2019, *In the shades of the uncanny valley: An experimental study of human-chatbot interaction*. *Future Generation Computer Systems*, 92, 539–548.
- [4] Go, E., and Sundar, S. S., 2019, *Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions*. *Computers in Human Behavior*, 97, 304–316.
- [5] Fitzpatrick, K. K., Darcy, A., and 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), e19.
- [6] Feng, X., Tian, Y., and Zhang, L., 2025, *The effectiveness of AI chatbots in alleviating mental distress and promoting health behaviors among adolescents and young adults: Systematic review and meta-analysis*. *Journal of Medical Internet Research*, 27(1), e79850.
- [7] Lucas, G. M., Gratch, J., King, A., and Morency, L.-P., 2014, *It's only a computer: Virtual humans increase willingness to disclose*. *Computers in Human Behavior*, 37, 94–100.
- [8] Moore, J., and Grabb, D., 2025, *Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers*. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 1–14. <https://arxiv.org>
- [9] Luxton, D. D., 2020, *Artificial intelligence in behavioral and mental health care* (London, U.K.: Routledge).
- [10] Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., and Sideman, L. A., 2005, *Is "service with a smile" enough? Authenticity of positive displays during service encounters*. *Organizational Behavior and Human Decision Processes*, 96(1), 38–55.
- [11] Li, J., Zhou, S., Zhou, X., and Feng, R., 2023, *Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being*. *npj Digital Medicine*, 6, 92.
- [12] Inkster, B., Sarda, S., and Subramanian, V., 2018, *An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation*. *JMIR mHealth and uHealth*, 6(11), e12106.
- [13] Heinz, M. V., Mackin, D. M., Trudeau, B., and Jacobson, N. C., 2025, *Randomized trial of a generative AI chatbot for mental health treatment*. *NEJM AI*, 2(4), AIoa2400802.
- [14] AlMakinah, R., Norcini-Pala, A., Disney, L., and Canbaz, M. A., 2024, *Enhancing mental health support through human-AI collaboration: Toward secure and empathetic AI-enabled chatbots*. Working paper, arXiv. <https://arxiv.org/abs/2410.02783>
- [15] Sim, K. Y. H., and Choo, K. T. W., 2025, *Envisioning an AI-enhanced mental health ecosystem*. Working paper, arXiv:2503.14883. <https://arxiv.org/abs/2503.14883>
- [16] Villar Barrios, M. A., Morrow, E., Blatter, N. Z., de Vreede, T., and Gaskin, J., 2025, *Human + AI therapy: A hybrid mental health approach*. *Proceedings of the 31st Americas Conference on Information Systems (AMCIS 2025)*, Montreal, Canada, 1–10.