Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Pilot Study of Human Empathy Expansion Through Facilitated AI Persona Engagement in Counselor Training

Rodney J. Luster

rodney.luster@phoenix.edu

University of Phoenix

Aaron Smith

Yorkville University

Research Article

Keywords: empathy, artificial intelligence, AI-human interaction, empathic attunement, empathic responsivity, performance based assessment, interpersonal reactivity index, relational engagement, pilot study

Posted Date: July 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10094062/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Pilot Study of Human Empathy Expansion Through Facilitated AI Persona Engagement in Counselor Training

Rodney J. Luster[1,2], Aaron Smith[2]

[1] College of Doctoral Studies, University of Phoenix, Phoenix, Arizona, USA

[2] Master of Arts in Counselling Psychology (MACP), Yorkville University, Fredericton, New Brunswick, Canada

Corresponding author: Rodney J. Luster (rodney.luster@phoenix.edu)

Abstract

Emerging autonomous applications have opened the door to more simulated relational forms of engagement. As AI–human interactions (AI–HI) increase, a gap remains in understanding how these interactions may shape the development of empathic responsiveness, particularly within counseling training settings. This pilot study explored the relationship between counselor-facilitated AI–HI encounters and the development of empathic responsiveness in counseling students. A mixed-methods pilot design was used, with participants (N = 10) engaged in structured, multi-session, counselor-led dialogic encounters with AI-generated personas responsive to empathic cues under a consistent interaction protocol. Quantitative data were collected using an 8-item empathy scale informed by the Interpersonal Reactivity Index, and qualitative data were deductively coded from session transcripts using an embedded Unlock Tier System to track changes in empathic responsivity. Participants demonstrated consistent movement toward greater empathic responsivity across sessions. Mean empathy scores increased from baseline (M = 23.10, SD = 5.02) to post-intervention (M = 25.80, SD = 2.90), indicating a large effect (Cohen's d = 0.89). Qualitative findings revealed growth within Perspective Taking and Empathic Concern, and tier coding indicated progression from basic emotional recognition toward more integrative, reflective empathic responding. These findings suggest that counselor-led AI–HI encounters may strengthen empathic precision and capacity in trainees, supporting an underdeveloped area in graduate training, with potential value for broader healthcare and leadership contexts. Despite its preliminary nature, the study supports the view that counselor-led AI–HI simulations may serve as scalable, auxiliary tools for developing empathic skills within counselor education and other helping professions.

Keywords

empathy; artificial intelligence; AI-human interaction; empathic attunement; empathic responsivity; performance-based assessment; interpersonal reactivity index; relational engagement; pilot study

1. INTRODUCTION

1.1 Foundations of Empathy

Empathy is universally recognized as a foundational part of effective helping relationships because it establishes the essential trust, safety, and understanding. Empathy is important for relational framing across counselor education, in addition to some contemporary leadership programs. Rogers [1] initially positioned empathy as the counselor's capacity to *sense a client's internal world* and communicate that pivotal understanding with reflected authenticity and unconditional positive regard, thus framing empathy as a "core" relational condition of therapeutic change. Clinical Psychologist Martin Hoffman [2] further expanded the conceptualization of empathy into facets of developmental and moral psychology, stating emphatically that "empathy is the 'spark' of human concern for others, and the emotional contingent that makes social life possible." Hoffman again emphasized the important role of empathy in motivating care, reducing aggression, and enhancing fairness within social interactions. Recent studies have reported the practice of engagement and self-regulated learning in educational contexts mediated by technology. For example, Lin and Wang reported how enjoyment of a foreign language and willingness to communicate in digital speaking contexts affect each other [3]. Wang and Guo showed that having a growth mindset and being able to adapt are important factors in learner engagement in AI-mediated informal learning settings [4]. Also, structural equation modeling to demonstrate the influence of L2 self-guides and achievement emotions on engagement in AI-supported digital learning [5]. These results show the general significance of structured, emotionally attuned engagement in AI-mediated contexts, a principle that applies to the cultivation of empathic responsiveness explored in the current study.

Over the years, in subsequent contributive scholarship that followed, attempts were made to refine definitions regarding our understanding of empathy beyond the purely "clinical" realm. Hoffman (2) further described empathy as a *developmental* and *moral capacity* that heavily supports prosocial behavior and mature perspective-taking, while contemporary Batson [6] had additionally emphasized empathic concern as a driver of *altruistic motivation* that can operate independently of one's own self-interest. These scholars together described empathy as layered and multidimensional, and shaped by one's relational experience, the resulting effectuation of caring enacted in the experience of interaction, and a moral invocation towards prosocial action.

Empathy has long been a focus of study and debate, and the field still contains noticeable gaps, especially in how different helping professions attempt to teach, assess, and support the phenomenon. These gaps become even more apparent in contemporary counseling programs, where student empathic development remains a difficult component to assess and measure. Programs preparing future psychologists and helping professionals must face the reality that, although empathy is central to their work, because of its nature, it can also be difficult to measure consistently, and perhaps even harder to cultivate in any deliberate or scalable way.

In professional counselor training programs and related helping fields, empathy is typically cultivated and evaluated through role-plays, supervisory feedback dialogues, formative academic feedback, and reflective practice assignments. As well, outcomes can vary across domains of empathic attunement (one's inherent capacity to accurately perceive, emotionally resonate, and responsively align to another's emotions). This is further impacted through unbalanced variables ranging from basic instructor modeling practices to subjective assessments, feedback and more. Yet compliance bodies like the American Counseling Association (ACA) make it transparent through standards such as their Code of Ethics, where ethical practice is noted as "inseparable" from empathy's central role in supporting respect, dignity, and the primary therapeutic alliance [7]. Beyond such academic and helper training-oriented environments, the domains of business and leadership training equally position empathic content as imperative to leadership and central to cultivating trust, fostering inclusion, and enhancing the relational effectiveness necessary for organizational growth by Weng et al. [11].

However, despite its centrality to training contexts, empathy in scholarly circles maintains its purists who contend that over-indulgence in *empathic focus* could lead to discordant forms of responding that are not uniformly beneficial and can be exposed to distortions. Phenomena contained within the domain of empathy, such as *empathic bias*, may conversely lead people to align more readily with those they perceive as *similar*, while something like *empathic distress*, if left unaddressed, may alternately contribute to issues like burnout or compassion fatigue, especially amongst helping

professionals [9]. Empathy is fundamentally a complex human component, influenced by a variety of factors such as one's capacity to tolerate emotional complexity, as well as one's cognitive understanding of how empathy develops and is ultimately expressed. These elements can encumber attempts to measure empathy, where some argue that it cannot be captured so adequately by any single tool or format. Even if such a measure were attainable, it would need to differentiate *other-oriented* empathic concern from *self-oriented* distress. Compounding this further, research also suggests that empathy may actually "decline" during intensive professional training, thereby highlighting a need for more deliberate and systematic approaches to support such aspects for progressive development [10]. This depiction of empathy and how to address it is further obscured as newer technologies and algorithmic systems begin encroaching on the human element of empathy through simulated empathic responding capabilities, where tools such as Character AI and other autonomous persona-based applications are now entering more mainstream conventional use. The shift suggests that something significant may be unfolding in how the psychological construction of empathy is understood and engaged within contemporary society.

It is important to further understand how empathy has made its way into the world of technology. Although there is no single agreed-upon definition of what artificial intelligence (AI) means, most descriptions tend to frame it as the use of computer systems to approximate or "mirror" aspects of human intelligence and behavior [11]. As AI continues to evolve, developers behind the scenes have continuously looked for ways to truly demonstrate its potential to interact with human beings and are now beginning to imbue autonomous AI with early forms of what might be termed *empathic expression,* in other words, the ability to emulate empathic concern for another. This shift signifies a consequential transition from what we once considered "automated" tools towards increasingly *autonomous systems*, introducing a larger and more looming sociological presence of considerable consequence when we consider the growing numbers of individuals who have turned to AI applications to address their effective needs. Moreover, innovations that model human-like emotional responses have been at the forefront of affective computing [11], including improvements to things like synchronized affect, differentiated emotional expression, and conversational coherence, where chatbots, character-based AI, and other autonomous agents can now act as artificial interlocutors engaging in exchanges that progressively blur the distinction between human and machine-mediated dialogue. These newly proliferating technologies have also introduced critical ethical concerns, particularly regarding the possibility of unregulated adoption within counseling and leadership sectors, and as a potential for advancing their liberal application within other structured and professionally guided networks. In the absence of a regulatory framework, the misuse of such public-facing AI tools remains a considerable and foreseeable risk, and researchers have re-emphasized the need for employing safeguards against rogue misappropriations and negligent deployment. Notably, as these concerns perseverate in public discourse, with documented instances of individuals acting on interactions within AI encounters, a parallel body of research suggests that AI-generated responses to human emotional disclosures may also yield beneficial outcomes [11], and are frequently perceived by users as highly compassionate, thereby prompting deeper inquiry into how individuals experience and interpret empathic cues emanating from these kinds of artificial agents [12]. Within counseling and training environments, however, the most prevalent question is not whether AI will ever supplant the gravitas of any kind of human-delivered care, particularly given that a considerable portion of the population still remains cautious and disinclined towards forming such connections with AI [12], but rather, whether structured, professionally facilitated integrations of AI personas integrated into human engagement encounters could be leveraged as an adjunctive resource for deliberate, and focused empathy development.

Accordingly, the present pilot does not attempt to explore the dimensions of *psychological anthropomorphism* (the assignation of human-like qualities to an artificial agent) or the boundaries inherent in AI-autonomous artificial expression, but rather, methodically, endeavors to explore a more practitioner-useful inquiry into the instrumental values that may be a bridge to partnering with such technological applications as it concerns the pragmatic aspects within the approach to empathy measurement and development. This approach brings together classical and modern ideas about empathy. These ideas are used in a new resource for therapy sessions with AI-driven, programmed empathic responses. The goal is to help participants become more empathetic. The inverse inquiry of the few studies proliferating in the research world typically seeks to inquire how AI robots and autonomous applications might enrich social experience or improve on AI's own responsivity patterns. The present study seeks to distinguish itself within this discourse by examining how AI personas, when coupled with scaffolded artificial empathy constructs deployed as training and assessment instruments, might serve to guide counseling students and other analogist programs toward focused dimensions of empathy development and expansion that have largely remained inaccessible within conventional training approaches. Central to the study's design was the creation of an experiential environment capable of

meaningfully engaging participants in authentic empathic encounters that were based on the aforementioned groundings of theoretical underpinnings. Within the present study, participants were required to engage in strategically structured, counselor-facilitated dialogic encounters with AI-generated personas. These personas were specifically developed around the principles of emotional expression and behavioral synchrony and were anchored by a constructed personality spine (P.S.), an artificially constructed system of enduring personality traits, relational patterns, and narrative history embedded within each AI persona. This architectural foundation would serve to facilitate a more consistent and environmentally balanced response style that is congruent with the naturalistic cadence of what we might understand as typical person-to-person discourse.

The PS was further designed to inform the AI persona's interactional style and response patterns, enabling more precise calibration to the naturalistic demands of human engagement and discourse as they would occur in these planned encounters. The adoption of a dyadic interactional framework was also configured for the AI-human interaction encounter, through which the AI persona could be systematically coded and positioned as a "legitimate" relational partner within such an exchange. The embedded characteristics would then allow for many possibilities in response to encounters with human participants. AI–HI encounter transactions would then be explicitly designed to elicit and measure *empathic precision* (referring to the accuracy and contextual sensitivity where one identity reflects, and responds in synchrony to another's emotional experience), attunement, rupture-repair, emotional signaling, and an understanding of empathic expansion (emotional growth). This experiment was further positioned to explore whether simulated interaction could meaningfully support and cultivate the competencies integral to established counseling programs, while concurrently influencing trainee therapists' own empathic resonance during the course of their professional development tenure. Such an inquiry additionally opens the consideration for the expansion of experiential learning paradigms, including the deployment of interventional strategies like exposure-based approaches administered perhaps within such AI-simulated environments. Grounded in the view that empathy can be strengthened through practice, insightful feedback, and reflective attenuation to discourse and language, this research afforded a way to possibly meet the aforementioned components within a facilitated AI engagement environment where measurable change, alongside observable markers of empathic development, could yield a strong basis for assessment, measurement and future inquiry.

1.2 Measurement, Psychometrics, Empathy and AI

One of the major and long-standing challenges involved in empathy research regards operationalizing a construct that can track the landscape of cognitive, affective and behavioral domains, with no single measure expressly capable of adequately capturing the full dimensionality of empathic expression. Before going further, it is necessary that we delineate the triadic dimensions of empathy that served as the primary targets of examination that informed and anchored the present study. *Cognitive empathy* is the first of this collective and is broadly conceptualized as the capacity to identify and interpret another's individual deeper perspectives and affective experiences [13]. *Affective empathy,* according to empathy literature, is best understood as the refinement of one's emotional attunement towards another, or more intently, the capacity to demonstrate genuine sensitivity and compassionate concern for another's emotional experience [13]. Lastly, *Behavioral empathy* refers to the verbal and nonverbal expressions of understanding, positive regard, and support that are consciously or otherwise unconsciously enacted and conveyed by one individual towards another within the interplay of interpersonal discourse [13]. Some scholars also describe a fourth layer, conceived as *moral empathy*, which resembles compassion, but is aligned more directly with altruistic actions and the desire to reduce another's suffering [13]. Given that empathy is also widely regarded within the literature as a *layered* and multi-dimensional psychological construct, the present study sought to investigate whether AI-HI encounters might function as a viable mechanism for calibrating and strengthening empathic precision across these constituent dimensions. Recognizing, however, that the inherent complexity and phenomenon of empathy render its measurement a nuanced methodological undertaking, a multi-modal assessment approach was employed utilizing several complementary instruments designed to capture discrete facets and hierarchical levels of empathic functioning.

A central measure in this work was the Interpersonal Reactivity Index (IRI), a 28-item self-report scale assessing Perspective Taking (PT), Empathic Concern (EC), Fantasy (FS), and Personal Distress (PD) [14-15]. Although widely used, self-report tools still remain vulnerable to response distortion, social desirability, limited introspective access, recall bias, and culturally shaped response styles. The PD subscale has also been critiqued for conflating empathy with personal vulnerability or maladaptive distress [16]. In addition, since the IRI measures dispositional tendencies rather than

moment-to-moment behavior, it cannot truly capture anything like real-time empathic responding as it might occur within a conversation.

Other instruments, however, might offer different perspectives. The Balanced Emotional Empathy Scale (BEES) is one that emphasizes affective resonance and emotional responsivity [17], though it can also blur distinctions between cognitive and affective empathy. Another is the Jefferson Scale of Physician Empathy (JSPE), which is useful in clinical environments but may not generalize well to other non-clinical settings, and where there is overlap with something like sympathy, which can complicate interpretations [18].

Together, these tools both highlight the inherent promise and also the limitations of current empathy assessment methods. Self-perceptions in these measures do not always match observable behaviors (ego-dystonia), additionally raising construct-validity concerns for this kind of self-report approaches [19]. Instruments such as the Accurate Empathy Scale (AES) offer structured observational ratings [20], but they also require trained raters, interrater reliability procedures, and significant institutional resources, making them difficult to scale in counseling and leadership training programs. As a result, such limitations have prompted calls for more behavioral and performance-based assessment development that may better correlate with empathy as it unfolds in real-time.

In the world of counseling as part of programmatic instruction, integration of AI into this landscape could likely aid in burden reduction in certain areas, such as reducing issues related to cost, infrastructure, and human capital, while easing time and cognitive demands on educators and supervisors. AI-supported practice sessions might additionally enhance training in counseling, healthcare, and leadership programs by offering “ongoing calibration” assessment and more precise ways to capture and train in the phenomenological dimensions of empathic experience. And as contemporary trends in neuroscience and physiological research continue to evolve, findings implicating structures such as the anterior insula, anterior cingulate cortex, and autonomic markers like heart-rate variability may help further complement the development of AI systems designed to navigate the full scope of human response, particularly as it pertains to mental and affective experience and their interplay with physiology which is closely tied to mental well-being [21].

As AI–HI encounters expand, and counseling programs remain reliant on measures with well-known limitations, such as self-reporting, observational ratings, performance tasks, and reflective assignments, an increasingly pressing need thus proliferates for approaches that could equip counseling trainees to more adequately understand and cultivate their own empathic capacities, while simultaneously addressing the needs of a population that is turning with growing frequency towards autonomous AI for emotional support. This burgeoning issue elevates the need for more consistent and scalable approaches to empathic capacity building, with the broader objective of strengthening empathic precision and ultimately enhancing the quality of client and patient encounters post-graduation. The strategic integration of AI as an assistive instrument within such training programs may prove instrumental in narrowing this widening gap

1.3 Could AI Fill the Gap in Empathy Development? The Inertia for This Study

The persistent challenges surrounding the study and application of empathy across research, training, and applied settings, where the measurement and assessment of empathy appear to further confound these efforts, necessitated a more focused and systematic review of existing assessment instruments, with particular attention given to their respective strengths and limitations. This review ultimately informed the design of an innovative Unlocked Tier System (UTS) application, a performance-based coding framework that was operationalized directly within the AI personas and utilized in the present study. Specifically constructed to monitor and assess empathic responsivity as it manifested across transactional encounters in communication, the UTS functioned as an integrated, performance-based evaluative complement to the experiential protocol researchers engaged, systematically tracking empathic responsivity during AI persona-driven interactions.

The impetus for this work was initially informed by a first-hand account known to one member of the research team. To contextualize the conceptual development of the study, the following account is presented as an illustrative example rather than empirical evidence. In recounting the experience, the individual described a recent argument they were involved in with their partner, which had become increasingly tense in the moment, prompting them to step away and return home. They had used an AI-based character platform that evening to continue ventilating their emotional frustration related to the earlier argument by interacting with an AI persona loosely modeled on their partner. As the

artificial exchange unfolded, the AI persona's reflective responses were visibly displayed within a chat-based interface as the persona's internal narrative (i.e. "What he's saying really hurts my feelings"). He explained further that this response by the AI surfaced deeper layers of reflection in him and the emotional significance of the words he used, in ways he had not anticipated. This unexpected mirroring and introspection prompted an internal shift as he continued to reflect on the event overnight. He explained that the next day, he and his partner decided to meet again, even though there was still fear and trepidation in both parties' minds from vestiges of the former day. What he also explained seemed to change the dynamics of their next encounter; he attributed it to the engagement of the AI persona the night before and the chance to reflect upon the narrative of the AI-character. As a result, he felt that he was better able to regulate his own nervous system and emotions, meeting up again and approaching their evening in a deliberate and compassionate way. This approach, he described, with his newfound clarity, allowed them both to open up to each other, allowing room to calmly discuss the argument with greater clarity, less emotional polarization, and a willingness to repair the rupture. His partner, he explained, was pleasantly surprised by the outcome, and the couple had a "wonderful evening" in his recounting of their follow-up. His ability to foster such reflection and control for any dysregulation allowed him to co-regulate with his partner, which is a part of what therapists do with their clients. This experience would be the catalytic flashpoint for launching the present study's exploration into such AI–HI encounters and their viability for developing empathic attunement and enhancing emotional precision.

1.4 Technology, AI, and the Advent of a Unique Unlock Tier System

Early systems such as Tess, Sara, and Woebot were among the first autonomous chat platforms that were first studied for their promise in supporting users experiencing forms of mild anxiety and related issues [22]. These innovations have progressed further from their former ancestors, where earlier generations of AI relied more heavily on scripted, rules-based responses. Updated versions, however, presented users with increasingly sophisticated conversational cues and anthropomorphic features that effaced more human-like, offering a "sense" of authentic social interaction for users. In some cases, these developments enabled AI to help people reveal issues of user distress, identify emotions, recognize thought patterns, and even provide some mild temporary relief. As simulated interactions continue to incentivize substantial investment from the technology sector in sophisticated AI engineering, it becomes worthwhile to potentially consider the broader, sweeping implications of deploying such technology within sound, professionally facilitated environments. If counseling programs and supervisory frameworks were to strategically integrate these tools, the possibilities for extending empathic development through training could be considerable, not only advancing the curriculum in a progressively meaningful way but also forging a deliberate and purposeful partnering between human-centered practice and growing technology applications. This trajectory could be of particular significance for institutions where healthcare and human-to-human practice already constitute the foundational pillars of professional preparation.

Some of AI's more promising functionality can already be seen in deployed AI-based support agents designed for business institutions and industry that create efficiencies, such as answering business-related questions, appointment setting, handling customer complaints, and more. The continued integration of artificial intelligence in healthcare, primarily within "facilitated" parameters, could offer additional emotional and social value to the industry. These systems may serve as a "temporary bridge" for many by providing immediate access to support services, such as filtering higher-escalation concerns to providers, while, in tandem, offering mild stabilization and regulation support until individuals can connect with a human provider. Some of this is already being researched and engaged for its potential by insurance providers and other industries. AI agents may also hold considerable promise as a preventative resource, capable of fostering smaller elements of psychological resilience while offering some temporary relief from mildly anxious arousal and stress that permeate work-life conditions.

All of this, though, must be approached with a line of caution as AI systems continue to develop and perhaps move into this sphere of conversational complexities, and must always be humanly managed, modified as needed, and have explicit boundary-set conditions when a person's health or impending issues might be concerned. To this end, however sophisticated their designs may be, AI agents remain fundamentally unable to approximate the lived psychological depth that surrounds an authentic empathic exchange, which imbues things like shared vulnerability, conscious experience, and the deeply human capacity for mutual recognition and co-regulation; no algorithmic construct can match such an effective and intersubjective experience. In the absence of the deeper psychological substrates of human consciousness, introspective capacities, and the affective and cognitive presence that are the singular trademarks afforded through authentic human-originated encounters, these tools should be regarded, unequivocally, as assistive and supportive in

nature, adjunctive by design and intention, and not as viable substitutes for genuine human connection [10, 23-24].

The more recent emergence of Large Language Models such as ChatGPT has substantially expanded the conversational capabilities of AI agents, ushering in a generation of situationally sensitive and emotionally coherent responsiveness that continuously tracks and adapts to user input in sophisticated ways. Ongoing scholarship in this domain further suggests that transparency in AI design may meaningfully shape the degree to which *users* extend genuine empathic responses towards AI agents [25], a finding that carries particular psychological significance when considered along the relational dynamics inherent to the therapeutic context. The conditions under which individuals feel genuinely seen, heard, and psychologically safe are not merely preferable but foundational to the depth and authenticity of what they're willing to disclose. This matters in counseling settings because clients often withhold intimate material initially due to fear of judgment, whereas AI-facilitated dialogue may neutralize or, at a minimum, reduce this inhibition altogether [26]. Preliminary research into AI use in facilitated dialogue shows evidence in its "assistive nature" in supporting mildly guided reflection and rehearsal cues that evince simulated empathic responding [10, 23-24]. What AI cannot manifest is bodily human attunement, supported by complex neural systems such as evidenced by the mirror neuron network and right-brain-to-right- brain resonance between humans, which are critically mandated for relational depth to take place in therapy and mental health practice. Amidst the meaningful progress in the AI landscape, this research seeks to position AI personas as more of a "supportive" application tool that, when incorporated alongside professionally guided training, may meaningfully expand current training paradigms for developing counselors, where the cultivation of metacognitive awareness and the intentional practice of empathy represent not merely supplementary competencies but foundational imperatives of effective clinical practice This study sought to explore these possibilities in the following ways.

1.5 The Present Study

To feasibly advance the methodological aims of the current investigation and operationalize empathic engagement as a viable, measurable construct within AI-HI dialogic encounters, it was necessary to establish a systematic and replicable means by which empathy could be reliably *recognized*, *recorded*, *categorized*, and *assessed* across the proposed experimental encounters. To achieve this, an analytic operational structure termed the Unlock Tier System (UTS) was created and would be integrated within an AI–GPT persona. The UTS would then function as a performance-based tool for identifying and measuring the depth of empathic responsivity as it unfolded in real-time during AI–HI interactions. It also allowed for direct comparison with IRI-coded domains utilized across these same participant sessions. By combining the ecological strengths of observational coding with the scalability of AI-mediated practice, the system would collect a rich source of transactional data. And when tier activation aligned with IRI domains such as Perspective Taking and Empathic Concern, it could offer evidence of convergent validity within the study. In addition, the UTS would make it possible to capture engagement elements and relational nuances that are often missed through things like self-reporting measures, observational ratings, or other traditional assessment methods. The real-time synchrony herein would be necessary and central to the experiment's design.

The Unlock Tier System application used for this experimental endeavor would record expressed empathic behaviors while assessing them simultaneously. Using AI in this way allowed for concurrent upstream measurement of social and affective processes if they were effectuated in sessions, allowing for continuous data on empathic recognition, validation, and synthesis as it evolved naturally during structured sessions while also being recorded. The results would either demonstrate the system's relevance or reveal limitations that could be scientifically falsified through post-study results.

To increase situational realism and support in-session empathic activations, the AI personas were additionally built with an embedded personality spine (PS) that included distinct backstories, unique characteristic traits, dispositions, and relational patterns. This design exposed participants to a range of interpersonal exchanges requiring flexibility in their empathic responding. For example, one persona might present with guarded reluctance, requiring patience and steady validation, while another might share more complex distress narratives that require careful engagement without escalating emotional intensity. This adaptability allowed for repeated, but varied practice that could evidence interesting and dynamic data. AI personas would display their internal narratives, obvious to participants throughout the sessions as part of the chat stream, and a means to elevate engagement while perhaps effectuating empathic propensities in participants.

In seeking a viable means of capturing and evaluating empathic engagement within AI-mediated dialogic encounters, the

present study developed the Unlocked Tier System (UTS) as a structured performance-based coding framework designed to operationalize empathic depth through the real-time analysis of participant responses. Informed by the conceptual logic of progressive pathways to healing described in prior work, the UTS organizes empathic interactions into three hierarchically structured levels of increasing emotionally relevant depth and communicative complexity. Those three categorical realms comprised the following tier levelling. Tier 1 actuation was built to explore *Rapport*, which would reflect surface-level acknowledgments marked by attentiveness and basic validation from the participant. Tier 2 actuation would be centered on *Emotional Recognition,* reflecting the participant's ability to identify and respond to emotions expressed by the persona, demonstrating movement beyond content and into the realm of affective engagement. Tier 3 actuation would focus then on *Core Wound Disclosure*, representing the deepest layers of engagement, where the participant's effectuated empathic responsivity would activate the AI persona's internal signal recognitions regarding response and the detection of psychological safety in their interaction with the participant. The UTS was further developed around the conceptual foundations of the *Theory of Relational Engagement* (TRE) [27]. TRE conceptualizes relational depth as evidenced through the convergence of *perceptions*, *actions*, *presentation*, and *context* [27]. Within this model, Tier 3 moments would also reflect much of TRE's heightened and referenced relational engagement propensities, as participants would align to interpretive stances, communication behaviors displayed, effectuated self-presentation, and derived situational attunement in ways that would drive deeper disclosure, meaning-making, and relational depth. TRE would thus support our investigation and serve as the driving theoretical reference for the study. In all of this, a person's interactions and analyzed dialogue in this simulated environment would then be the key needed to understand their resonating depth and track it along this specifically designed system of detection.

With this framework established, the present study was positioned to systematically examine the range of empathic engagement elicited among participants through their encounters with the constructed AI personas and the novel Unlock Tier System. If the study revealed meaningful shifts in empathy, the findings could offer a practical pathway for integrating a similar AI application into training within institutional counseling programs. Compared with things like "live" role-plays and labor-intensive observational methods, such a system would allow for deeper insights and analysis through data-rich transcripts and coding to support scalable analysis with greater methodological consistency.

Ultimately, by comparing tier attainment with established self-report measures such as the IRI, this study could also test for convergent validity and understand whether performance indicators arising in dialogue would align well with traditional empathy constructs. The tier model would also align well with process-based accounts of empathy. Zaki [28], for example, argues that empathy is better understood as a set of processes rather than a single trait. The tier model would be capable of reflecting this view by treating empathy as something that "unfolds" rather than a static, unchangeable quality, providing a clearer window into how empathic engagement develops and under what dialogic circumstances. With these theoretical and technological foundations in place, the study moved forward into further methodological design and eventual implementation. As an exploratory pilot study, the main goals were to find out if the AI–HI facilitated encounter model could work, to see if the Unlock Tier System could be used as a performance-based measure, and to come up with preliminary effect size estimates that could be used in future, larger-scale confirmatory research.

2. MATERIALS AND METHODS

2.1 The AI Persona and the Imbued Personality "Spine"

The term *AI persona* would be used to denote the researcher-constructed artificial agents that were developed for use in this research, where each would be encoded with a defined *personality spine* that contained a relational backstory and programmed empathic response patterning. For understanding its usage, the *AI persona* is herein differentiated from the broader term *AI robot*, which is still retained when referencing older literature and externally developed systems. To further augment the psychological depth and relational authenticity of the constructed AI personas, each was anchored by an *Enneagram*-based personality archetype, a theoretically grounded typological taxonomy whose nine distinct personality structures would offer a rich and clinically resonant foundation for shaping the motivational orientations, relational patterns, and effective dispositions that would govern each AI persona's interactional style. This component would be combined with a backstory from a mosaic of de-identified thematic clinical themes and modified vignettes to create more realistic layers to the AI personas engaging participants. The personas were also programmed to exhibit

distinctive disclosure patterns, respond with contextual adaptability, social resonance, behavioral and affective synchronization, visible narrative interiority, effectuated psychological defensive styles, simulated introspection, and reflection potentials. Participants would engage the AI personas through a secure, non-clinical digital communication platform, within which no protected health information was collected and full de-identification of all participants was maintained throughout. Each session was then structured as a 15-minute protocol comprised of an opening period of facilitated delivered instructions and reaffirmation of informed consent, two 15-minute dialogic encounter sessions with the designated AI personas of choice, each initiated by a participant-originated salutation, a structured closing interaction with the facilitator, and an immediate post-encounter assessment administered upon conclusion of the session.

The four constructed personas, imbued with the aforementioned components, are designated by human naming conventions alongside brief descriptive character marquee entries as presented below (Figure 1).

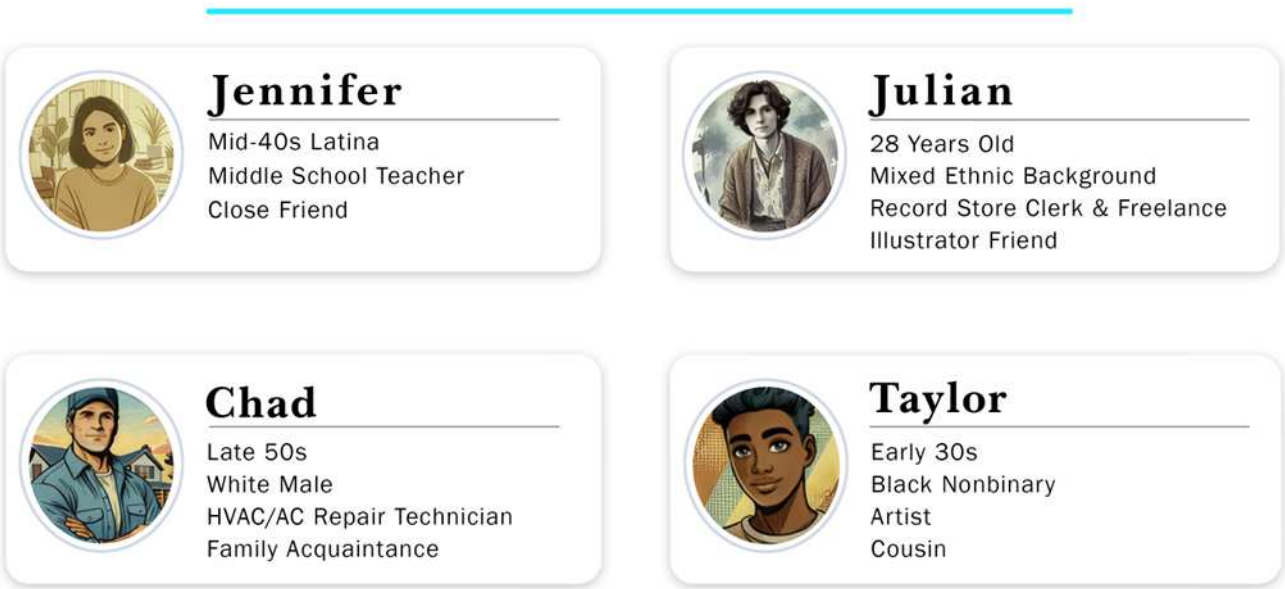


Fig. 1

Constructed GPT personas with embedded backstories. Four distinct personas (Jennifer, Julian, Chad, Taylor) were designed to simulate varied personality styles, disclosure tendencies, and embedded with relational backstories.

2.2 Participants

For this study, 11 adults were recruited through professional and social networks. One participant was excluded from the final analysis due to incomplete session participation and incomplete post-test data, resulting in a final analytic sample of N = 10. In line with the exploratory objectives of a pilot study, this sample size was deemed sufficient for evaluating the feasibility of the protocol. It also allowed refining the UTS coding framework and yielding preliminary estimates of intervention effects to guide power analyses for subsequent research. Participants resided in both Canada (n = 4) and the United States (n = 6), providing a modest cross-national sample. Eight participants had identified as female and two as male, with ages ranging from 25 to over 60 years. Prior research on AI–HI engagement has disproportionately focused on primarily younger populations, likely reflecting their higher rates of engagement with digital technologies. However, with limited representation of adults over the age of 60 as seen in the research [8], this study did include several participants aged 60 and older, thus offering a unique additional frame less frequently depicted in any currently existing AI–HI literature. As a guiding principle of the research, and in order to preserve confidentiality, participants were assigned unique identifier codes combining participant number with the first letter of a name (for example, P1-Z). This study was also limited to participants of *consenting age*. No participants under the age of 18 were included in this study.

The research received approval from the Research Ethics Board (REB).

2.3 Ethics approval and informed consent

This research was approved by Yorkville University's Research Ethics Board (REB) (Protocol/Certificate No. 202500624RL) on June 24, 2025. The study launched in July 2025, and all participants were provided with their electronic informed consent prior to any participation or engagement in sessions. Participants also received complete study information in advance, and all consent was documented via electronic signature and an "I agree" confirmation at the start of the online workflow. Electronic consent records were then stored in accordance with institutional data security requirements in a secure, encrypted environment in accordance with REB research requisites. The research was supported by a research grant from Yorkville University. The study was conducted in accordance with REB ethical standards and the principles outlined in the Declaration of Helsinki for research involving human participants.

2.4 Procedure

The study was conducted between July 2025 and October 2025. Participants completed an 8-item baseline empathy self-assessment prior to engaging in any AI persona sessions. Sessions were conducted on a secure, privacy-protected digital platform, with strict de-identification measures in place. To reduce any expectancy effects, participants were informed that the intent of the study was to explore AI usage in online interactions; the main focus of this study on empathy was not disclosed. Once all consent and communications had been satisfied, along with securing reserved dates of engagement, participants then engaged in 50-minute sessions conducted independently of one another.

Following a brief orientation to the session procedures, participants were instructed to select two of the four available AI personas prior to engagement. Each encounter was initiated by a participant-originated greeting directed toward the selected AI persona, effectively launching the dialogic exchange. Sessions were structured to last up to 50 minutes, with a stipulated minimum engagement threshold of 15 minutes to ensure sufficient depth of interaction. Although voice-to-voice capability was available within the platform, all persona encounters were conducted via text-based chat to facilitate more accurate and reliable transcription of session content. Participants were additionally encouraged to engage in purposeful moments of silence as needed, affording opportunities for reflective pausing and deliberate consideration of conversational elements prior to responding to AI-generated commentary. The AI personas themselves were designed to respond dynamically and adaptively to participant input, sustaining conversational engagement and maintaining relational continuity throughout the duration of each session. The artificial empathy architecture encoded within each persona was deliberately constructed to embed sufficient empathic material to support deeper resonance and more penetrative, effective engagement between participant and persona. The underlying premise of this design was that the back-end programming constituting each persona's psychological and relational infrastructure would enable increasingly dynamic and responsive communication across the arc of each session, while sustaining important participant engagement throughout.

Following the conclusion of the encounter sessions, participants then completed a post-test empathy self-assessment instrument. All sessions were recorded, and participant identities were systematically de-identified through the assignment of participant codes prior to analysis. Session transcripts and embedded scoring data derived from the unlocked tier system (UTS) were subsequently extracted, de-identified, and prepared for formal outcome analysis.

2.5 Instrumentation

Three instruments were used in this study which included (a) an adapted 8-item empathy self-assessment scored on a 0–4 Likert scale (total range: 0–32), items were adapted to capture core aspects of affective and cognitive empathy in brief self-report format (b) transcript coding guided by the four domains of the Interpersonal Reactivity Index (IRI): Perspective Taking, Empathic Concern, Fantasy, Personal Distress; and (c) the Unlock Tier System (UTS), a study-specific structure designed to classify levels of empathic engagement. As the IRI is a copyrighted instrument, only the domain structure was used; no original items were reproduced.

The UTS classified empathic engagement into three distinct levels of affective arousal: Tier 1 (Rapport), Tier 2 (Emotional

Recognition), and Tier 3 (Core Wound Disclosure). In the Supplementary Materials (see "The Unlock System: Conceptual Overview, Scoring Model, and Validation Framework"), you can find a full explanation of the UTS scoring logic, signal weighting, and theoretical basis. Tier classifications were applied based on predefined criteria to ensure consistency across transcript evaluations. Higher-tier classifications typically indicated greater levels of empathic engagement occurring during participant-AI interactions.

2.6 Data Analysis

Data analysis was conducted along two distinctive, complementary phases. The first phase involved an analysis of pre- and post-test scores from the 8-item empathy self-assessment (0–32), which were summarized using descriptive statistics due to the smaller sample size (N = 10). Quantitative data were further analyzed using a paired samples comparison to examine pre-post differences, with effect sizes that were calculated to support interpretation in accord with a smaller sample. The second phase of this study involved deductive coding of session transcripts guided by the Interpersonal Reactivity Index (IRI) domains of Perspective Taking, Empathic Concern, Fantasy, and Personal Distress as targeted areas. In this, the UTS levels (Tier 1-Rapport; Tier 2-Emotional Recognition; Tier 3-Core Wound Disclosure) would also be a part of the larger data analysis. Coding was conducted by the primary researcher, with a second coder independently coding a subset of transcripts for reliability assessment. Transcript coding reliability was assessed using Cohen's kappa ($\kappa = .70$), indicating substantial agreement. Discrepancies that existed between coders were primarily observed at adjacent UTS thresholds or in borderline distinctions between empathic concern and neutral responses, revealing some slight variation in judgments of empathic depth rather than any inconsistency in the coding framework. Consensus procedures were applied prior to final qualitative analysis. Additionally, to support transparency and replicability, coding and convergence checks across self-report scores, IRI-coded domains, and UTS attainment were managed using a purpose-built spreadsheet-based application developed for this study rather than specialized qualitative software (for example, NVivo, MAXQDA), which was not warranted given the pilot nature and limited sample size.

RESULTS

After taking all assessments and measures into account, and after thorough analysis by the researchers, the results presented below depict data analysis outcomes, including sample transcript excerpts (illustrative samples which reveal emotional expressiveness, attunement to the conversation, and empathic resonance observed) as extracted from AI–HI guided encounters. We begin with self-reported empathy scoring.

3.1 Self-Report Empathy Scores

Relevant to the study's inquiry regarding empathic enlargement (effectuation, attunement, resonance), 10 participants completed both pre- and post-test assessments, and mean empathy scores demonstrated an increase from 23.10 (SD = 5.02) at pre-test to 25.80 (SD = 2.90) at post-test, recording an average improvement of 2.70 points on the 0–32 scale. Most participants *did* demonstrate noticeable increases, as variability was observed and accounted for by one participant's nominal decrease. To this end, the pre-to-post change reflected a large effect size (Cohen's d = 0.89). Overall, participants demonstrated upward trends in empathy scores. Below, group/participant level changes are shown in Figures 2, 3, and 4 and Table 1, illustrating the deeper data dive revealing means and standard deviation data, along with scores and outcomes of participants in the experiment.

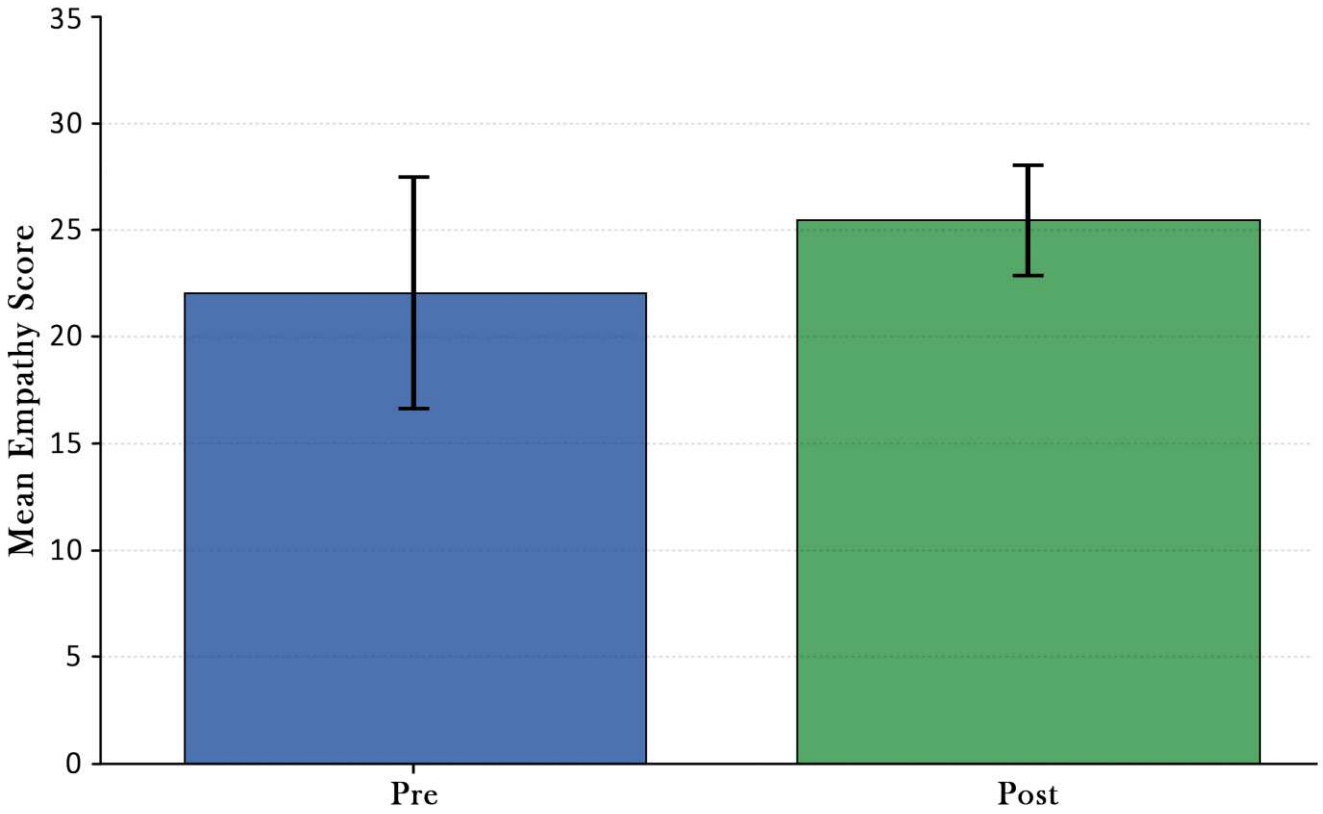


Fig. 2

Mean pre- and post-test empathy scores with standard deviations (N = 10) are shown here. Participants showed a moderate mean increase in empathy self-ratings following two AI-based interactions, with a positive trend in direction.

Table 1.

Means and Standard Deviations of Empathy Self-Assessment Scores (n = 10)

Measure	Pre-Test M	Pre-Test SD	Post-Test M	Post-Test SD
Empathy	23.10	5.02	25.80	2.90

Note* Scores are based on the 8-item empathy self-assessment scale (0–32).

Figure 3 below presents participant-levels and the trends in both pre- and post-empathy results, while Figure 4 provides an additional visualization of these changes.

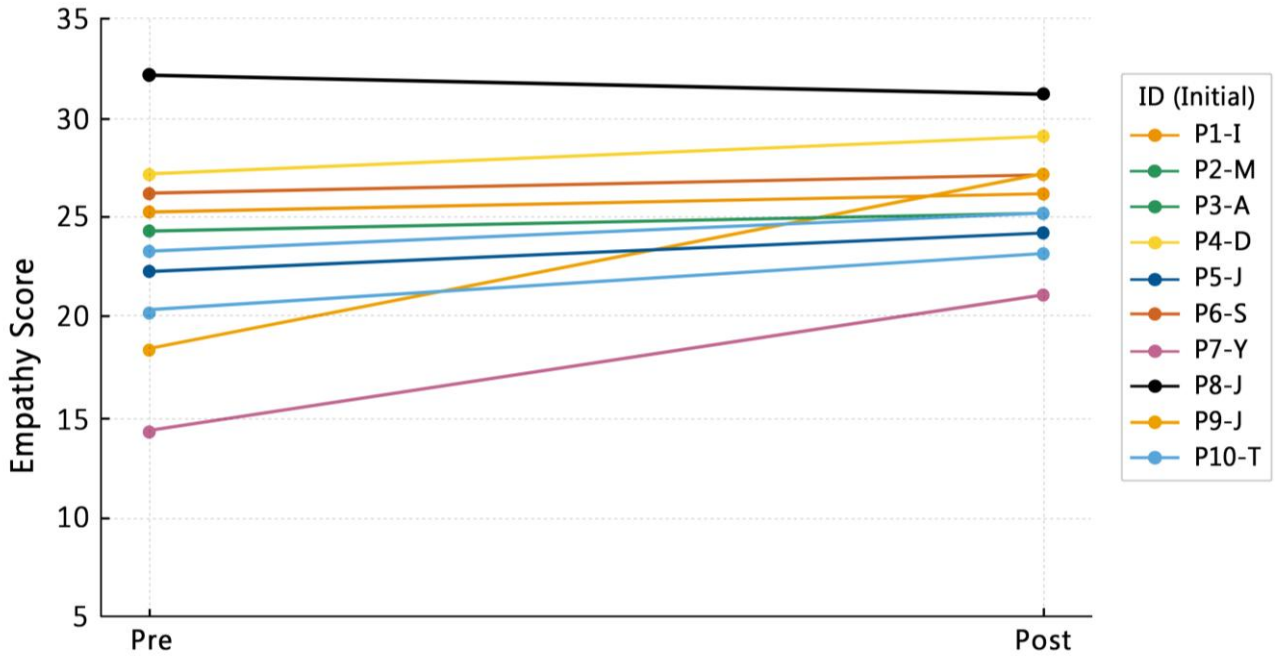


Fig. 3.

Individual trajectories of participant empathy scores depicted before and after AI-based interactions (N = 10). Most participants demonstrated meaningful increases in empathic enlargement (effectuation, attunement, resonance), where some degree of variability in individual change was recorded.

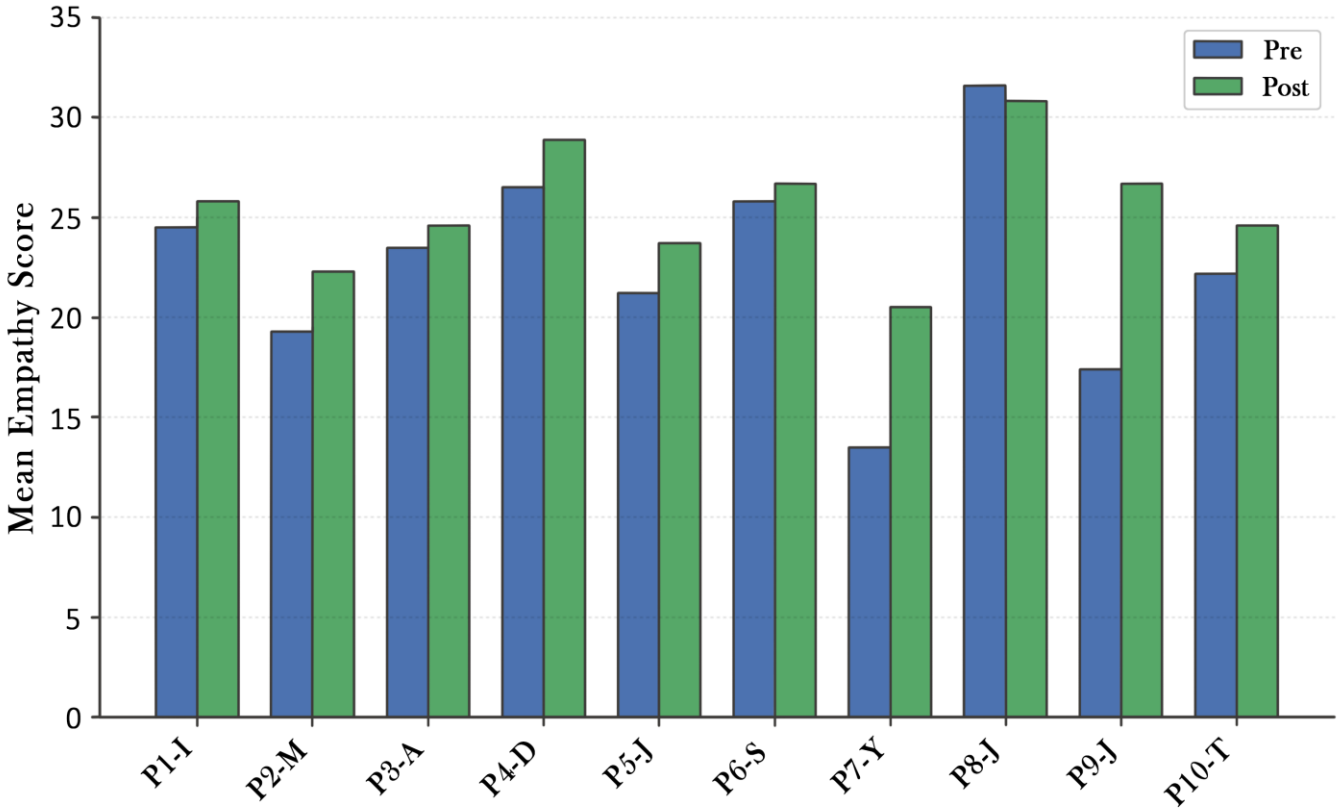


Fig. 4.

Individual participant pre- and post-test empathy scores (N = 10). Side-by-side bars illustrate relative changes across participants.

3.2 Unlock Tier System (UTS) Outcomes

UTS Tiers were the progressive scaffolds of the embedded AI persona embodied construct. As participants engaged with the AI personas in the 2 round 15-minute encounter sessions, the AI persona scanned participants' responsivity for behavioral cues signaling deeper relational synchronicity around *trust*, *safety* and *care*, which would trigger greater back story disclosures elicited from the constructed personas. Tiers 1-3 were accessible only through the effectuation of the following progressive paths:

Tier 1 interactions in the UTS emphasized rapport and acknowledgment, aligning strongly with cognitive empathy and perspective-taking. Tier 2 involved explicit recognition and naming of emotions, corresponding with empathic concern and affective attunement. Tier 3 reflected deeper "relational" components around safety, trust and vulnerability, characterized by core wound disclosure and the motivational dimension of empathy described in Batson's [6] empathy-altruism hypothesis.

Conceptually, the tiered structure approach offered a practical way to describe how empathic engagement can deepen during immersive interactions. It also aligns with relational views of empathy. The *Theory of Relational Engagement* (TRE), for example, proposes that empathic depth emerges through the convergence of perception, action, presentation, and context [27]. In this study, reaching Tier 3 reflected that convergence: participants not only adopted the persona's perspective, but also demonstrated empathic behaviors through sensitive communication and the creation of conditions that supported conversational synchronicity, leading to deeper disclosures.

The outcomes showed that 9 of the 10 participants reached Tier 3, the highest level of empathic resonance on the UTS scale. Nine of ten participants (90%) reached tier three, the highest level of empathic engagement on the UTS scale. Tier 3 attainment indicated that participants were able to effectuate deeper vulnerability and disclosure within the AI personas. One participant reached Tier 2, which still reflects positive upward movement along the empathic scale. Together, the data revealed a latent construct within the UTS results that the researchers view as empirically coherent and worthy of further study. These findings, however, do not imply sustainable or durable empathic skills as a result of a one-time encounter, but rather, a form of empathic conditioning that manifested in the sessions. This points to the experiential conditions present within AI–HI encounters where participants did in fact temporarily activate components of deeper empathic resonance. Identifying that latent capacity was a central aim of the study. Such depth of engagement, though, appeared contingent upon sequential mediation of intimacy and reciprocal emotional attunement within those AI–HI interactions. Participant-level outcomes are seen and summarized in Table 2 below.

Table 2.

Participant Demographics, Empathy Scores, and Unlock Tier Outcomes (N = 10)

ID	Country	Gender	Pre Totals	Post Totals	Unlock Tier
P1-I	Canada	Female	25	26.0	Tier 3
P2-M	Canada	Female	20	23.0	Tier3
P3-A	Canada	Female	24	25.0	Tier 3
P4-D	US	Female	27	29.0	Tier 3
P5-J	US	Male	22	24.0	Tier 2
P6-S	US	Female	26	27.0	Tier 3
P7-Y	Canada	Female	14	21.0	Tier 3
P8-J	US	Female	32	31.0	Tier 3
P9-J	US	Male	18	27.0	Tier 3
P10-T	US	Female	23	25.0	Tier 3

Note* One participant (originally coded P1) was excluded from analysis due to incomplete post-test data, resulting in a final sample of N = 10.

Beyond pedagogical application, the results demonstrate possible implications for understanding how empathy as a "multidimensional" construct along the lines of cognitive, affective, and motivational components of the mind, can be tapped via AI–HI encounters like the ones in this experiment [2, 15]. The results of this pilot research also provide some significant insight into AI–HI encounters where the UTS could be a viable means for measuring and developing empathic aptitude.

With respect to the phenomenological dimensions of the present study, selected transcript excerpts are additionally employed to eliminate participant perspectives as they were experienced in *situ* during engagement with the AI personas, simultaneously functioning as qualitative evidence of empathic movement as it organically unfolded within each encounter. The following examples offer an illustrative wedge of insight into how participants began to evince change and empathic responding as was recorded through the domains of the Interpersonal Reactivity Index (IRI) and the UTS during actual AI–HI encounters.

3.3 Transcript Excerpts

The following excerpts provide a brief purview into how empathic responses were expressed and varied across interactions, as reflected in UTS tier levels and aligned with IRI-coded domains.

P7-S-Response to AI persona story/dialogue encounter-(Tier 3)-Use of curiosity or meaning-oriented inquiry. Participant response after AI persona shared something personal- "Why do you think you pulled away at that moment?" Here, the participant engaged in curiosity while their inquiry became an access point to deeper, relevant material and insights with the AI persona as the session progressed. (Perspective taking, Meaning-oriented inquiry, and Empathic concern).

P8-Y-Response to AI persona story/dialogue encounter-(Tier 3)-Use of metaphor elaboration for affective attunement. AI persona response - "Sometimes it feels like I'm just floating." Participant response: "Floating can mean freedom or being unanchored." In this response dialogue, the participant is engaging in metaphor expansion and emotional granularity, which are sophisticated empathic moves that help the AI persona further articulate their internal state. Here, the participant sought to help alleviate anguish, distilling AI persona conveyances with increased granularity of expression (Meaning-making, Affective attunement, Empathic concern).

P9-J-Response to AI persona story/dialogue encounter-(Tier 3)-Use of continuity tracking. This participant responded with continuity tracking, linking to prior disclosures from AI persona responses: "It reminds me of when you described losing touch with your sister, similar?" Here, the participant held earlier disclosures in mind, linking them to the present and attempting to establish episodic connectivity. This level of empathic behavior denotes narrative memory, associative linkage, and sustained attunement activation. (Perspective taking, Attenuation, Narrative attunement).

P2-I-Response to AI persona story/dialogue encounter-(Tier 2)-Use of affect recognition, validation and empathic concern. The AI persona reveals some personal information. Participant responds- "That sounds painful. I hear how important it was." In this response, the participant used mirroring of emotional content and affirmed the significance, but did not move into meaning-making or deeper associative work to advance Tier-step progression. (Emotional recognition, Empathic concern).

These sample transcript extracts provide a few of the instances observed in the study. Across them, progressive shifts in participant responses were evidenced further as sessions continued with the AI personas. Observed changes took place when the AI persona detected markers for *trust*, *safety*, and general *concern*, allowing room for deeper deposits of disclosure by them. Across sessions, the facilitator noted recurring "exception" moments where participants demonstrated increased *metacognitive awareness* in how their communication began to shape and influence the AI persona's reactions; moments that often surfaced mid-session as participants became more immersed in the evolving dialogue. What was evident to the researchers across all transcripts was the strong presence of indicators around Perspective Taking and Empathic Concern, with moderate levels of imaginative perspective-taking (Fantasy), and fewer indicators of Personal Distress. This pattern parallels earlier findings showing that an AI persona's programmed cognitive and affective empathy cues can transactionally define how users approach, or even avoid, certain responses [29]. The qualitative themes in this study appear to align with the quantitative results ascertained, suggesting that AI–HI interactions can offer moments of meaningful empathic engagement for participants. Taken together, these findings indicate convergence across self-report scores, UTS tier classifications, and transcript-based IRI coded responses.

Collectively, these results provide a starting point for considering the broader theoretical and practical implications of AI- facilitated resourcing.

4. DISCUSSION

The present study contributes to what is a growing body of research on artificial empathy by shifting focus from the development of systems that simulate empathic behavior to the examination of how empathic engagement tends to *unfold* within such AI-human encounters. Rather than centering on algorithmic recognition alone, the study explored empathic enlargement, which was operationalized through *effectuation*, *attunement* and *resonance*, within AI-HI interactions.

The primary trend in the body of research on artificial empathy tends to emphasize the "development of systems" within the technological space that can recognize or simulate empathic behaviors, such as pain recognition across facial, vocal, and bodily cues [30]. Artificial empathy of this kind represents an advanced level of functioning in the areas of perspective taking, empathic concern, emotional contagion (mirroring), along with facial and bodily cues [11]. As well, noted again in the work of Zhang *et al.* [11], AI robots that adhere algorithmically to interpersonal norms *can* facilitate areas around self-disclosure [31] and promote psychological identification [32]. Building on this prior work, the present study extended this line of inquiry by centering the human response to AI-HI encounters through the theoretical lens of empathic enlargement (effectuation, attunement, resonance) as would effectuate given a structured interactional setting [11]. If artificial intelligence could serve as an assistive technology capable of approximating empathic cues with sufficient fidelity to be recognized and trusted in helping domains, then structured, intentional AI–HI facilitated encounters by trained counselors or healthcare-approved professionals may also serve as a catalyst for promoting empathic performance and capacity in students enrolled in developing counselor programs. This framing additionally supports the replicability of AI-HI encounter sessions as a systematically reproducible augmentative modality for the ongoing assessment and cultivation of empathic responsivity and capacity. It simultaneously offers a methodologically sound and scalable means of accumulating empirical data across successive iterations for the purposes of continued scholarly inquiry. What arises from the present study is a compelling indication that empathic engagement that is conducted through AI-HI encounters is not merely incidental, but can be deliberately stimulated, activated, and systematically observed across multiple layers of analysis Participants demonstrated moderate increases in self-reported empathy scores, including strong engagement within the UTS classification system, with the majority reaching Tier 3 levels indicative of deeper relational attunement and disclosure. These patterns were further reflected in transcript-based analysis, where participant responses aligned with IRI-coded domains such as perspective taking and empathic concern. The convergence observed across self-report, behavioral coding, and tiered engagement indicators also implies that empathic processes were not merely expressed but actively enacted within the interactional discourse of each session, lending preliminary support for the ecological validity of the UTS as a performance-based measure that is capable of capturing empathic engagement as it emerges within concurrent dialogue.

As the expanding body of evidence continues to come to light in support of this study's purposive area of inquiry, it is worth noting that AI-HI facilitated encounters have been perceived by users as *comparable* in quality and authenticity to human-authored responses [33]. This can be engendered through such dyadic fusion, as contained in this experiment, where an AI persona and a human participant begin to center on affective attunement around empathic concern. What these findings permit us to reasonably infer is that, with deliberate construction of AI-HI discourse environments for the purposes of empathic expansion, such environments may hold meaningful promise as a formative adjunctive aid within programs responsible for the professional development and training of counselors and allied healthcare providers. This plausibility is further substantiated by a growing body of research supporting the integration of formalized AI environments as constituent components of structured training applications, designed to afford repeated and purposeful exposure opportunities for practicing engagement while grounding development and cognitive accuracy.

4.1 Comparisons

How does this approach then compare to what has been historically relied upon to train counselors and others in traditional training contexts? Empathy development in counselor education is nuanced at best and has long been embedded in roleplays, supervised practice, reflective journaling, curriculum-based instruction, and group process activities. These traditional methods still remain foundational to relational skill development. Roleplays, for instance, may offer structured opportunities to practice responding to interpersonal disclosures, while supervision provides corrective feedback tied to in-session performance. Reflective journaling, as well, can deepen introspection and awareness of emotional reactions and personal biases. At the same time, these pedagogies present limitations. Roleplays,

for instance, depend heavily on peer performance and may not capture the full complexity of real client encounters, where multiple dynamics typically evolve simultaneously. Supervision is also resource-intensive and often constrained by time, scheduling, and supervisor availability. Lastly, journaling, though valuable, lacks the potency of immediate feedback and does not allow for real-time rehearsal of micro-skills. These constraints can reduce the potentialities in students and their development of empathic precision, a core component of psychological synchrony in human relationships.

In contrast, AI–HI encounters offer a different set of possibilities. AI systems have been studied for decades, beginning with early customer-service applications, and this body of research has contributed to the development of contemporary AI personas and relationally oriented robotic systems. This progression reflects a shift from surface-level automation towards systems capable of inviting more intimate self-disclosure and responding to emotional cues through increasingly sophisticated signaling and dialogic patterns. These artificially designed relational exchanges are now more commonly used for social engagement, fantasy roleplay, exploratory dialogue, and other role-based interactions, which are becoming more prevalent with increased accessibility to such technologies that now proliferate.

Drawing from *Computers as Social Actors* (CASA) theory, prior research also demonstrates that individuals often apply human *social norms* to AI interactions, even when they are completely aware of the system's algorithmic nature [11]. This was evident in the present study, where participants frequently concluded interactions using normative closings such as "Farewell," "See you later," or "Talk soon." These observable behaviors suggest that structured AI–HI exchanges can have a deeper resonance and may support multidimensional empathic processes, particularly for individuals developing competence within complex communicative situations [34]. Within counselor education, this highlights the value of supervised AI–HI encounters as a supplemental training tool but not a substitute for traditional methods.

At the same time, critics cautiously forewarn that AI-generated empathic expressions, while compelling, do not reflect genuine "feeling states" grounded in the confines of human consciousness. A critical counterpoint to this promise, however, lies in the possibility that the convenience and accessibility of AI-HI encounters could inadvertently displace rather than supplement the development of empathic competencies as they are naturally forged within lived interpersonal [35]. Additional concerns include anthropomorphism and the development of parasocial, one-sided bonding, especially among individuals seeking connection in the absence of supportive interpersonal relationships. *Speciesism* arguments further underscore the risk of blurring boundaries between human and non-human emotional exchanges [11]. Halpern [23] argues that while AI can simulate empathic responses, it cannot truly care, as genuine empathy requires *moral commitment* and *subjective responsivity*. Rubin et al. [36] similarly note that AI may approximate cognitive empathy and support psychoeducation, but the empathic labor essential to therapeutic alliance undoubtedly remains within human experience. In the end, these considerations reinforce the need for partnered approaches if we are to engage technology within counseling education, and other arenas, where there is potential, and its use to extend "practice" opportunities while the core of therapeutic connection and ethical responsibility can remain human-centric. Of course, this would take thoughtful consideration to consider such emerging technologies to serve as meaningful extensions within counselor and psychological training. Real-world evidence may offer us a glimpse into narratives gleaned and research that explored the COVID-19 pandemic, which further illustrates this potential. Jiang et al. [24] found in their research on that event that individuals who engaged with socially assistive AI systems during periods of heightened psychological stress reported benefits such as stress relief, emotional support, and short-term resilience gains, including strengthened identity anchors and improved emotional states. These findings point toward AI-assisted interactions' contributive possibilities within human-centric support when integrated as part of a broader, ethically grounded guiding structure.

4.2 Considerations

Up to this point, the findings underscore the pedagogical value of AI-guided persona work as a promising tool for future use in counselor education. Empathy, identified by accrediting bodies such as CACREP [37] as a core professional competency, may be approaching a more "partnered" application use in this assistive technology that could support its use in enlarging students' empathic potentials within counseling. But we may not be there yet. Yet, many healthcare and educational training programs continue to face challenges in providing consistent, measurable, and scalable opportunities for students to practice and develop empathic accuracy. And although empathy *in toto* can be difficult to track, AI–HI therapist-led interactions may help fill the void. Recent evidence posits that patients may at times rate AI-

based responses as equally or even "more" *empathic* than their physicians' responses, further affirming the value behind AI–HI interactions as a supplement, rather than a replacement for human encounters [38].

Another key consideration is its perspective benefit in the area of *preparatory skills training*, where AI–HI facilitated encounters may offer usefulness in intervention-based role playing for intervention tools like the *empty chair technique* and *role reversal* in an experiential format. Students may also benefit from using strategically constructed AI personas to rehearse empathic responses across a range of situational contexts and storylines prior to engaging in peer roleplays or utilizing them within skills labs. This form of practice may help reduce student-counselor anxiety, strengthen confidence, and support the development of an internal locus of control, while also creating space for iterative experimentation.

A second consideration involves the need for oversight and supervision. Supervision remains central to counselor development; however, practical constraints often limit the extent of repeated practice and in-depth reviewing that supervision can provide. Within counselor education, supervision reflects both an affective commitment to trainee development and the managerial responsibility of conducting clinical appraisal and ongoing professional assessment [39]. When developmental coaching, administrative responsibilities, and teaching demands are combined, supervisory capacity can well become overstretched. Gustavsson and Bivall [40] note that clinical supervisors must "balance" routine professional duties alongside the time and attention required for effective supervision. Chief among these challenges is ensuring that students meet their learning objectives, determining when students can perform w more independently in practicum settings [40].

AI-mediated dialogues may extend such supervision processes in a complementary way. For instance, AI-HI interactions could be incorporated into supervision assignments, with students submitting transcripts from AI–HI encounters for review. Supervisors could then evaluate interaction patterns, identify areas of growth, and provide targeted feedback on domains such as empathic resonance, attunement, and client synchrony while conserving direct supervision time. The transcript format further allows for detailed coding and analysis, offering supervisors concrete material to score, guide and discuss in greater depth with trainees.

AI–HI encounters can broaden the range of interpersonal experiences available to counseling students. AI personas can be constructed to reflect diverse backgrounds, communication styles, and relational challenges, providing exposure to a wider range of empathic encounters prior to working with actual clients. This approach aligns with multicultural and social justice lenses as applied within counselor education by allowing students to practice empathy in safe, simulated environments that mirror real-world cultural dynamics at work, and can allow deeper reflection on how their own *positionality* shapes their responses.

Scalability also represents an additional practical advantage. AI personas can be deployed at a lower cost than hiring additional human capital, making AI personas a case-use-advantage, a more feasible pathway for institutions seeking to expand training and development. In this sense, AI applications may foreseeably help programs meet such increasing demands without overburdening valuable resources.

4.3 Integrated Synthesis of AI–HI Empathic Training for Industry Leaders

The value of AI–HI–based empathy training could also possibly traverse beyond counselor education and into leadership and supervisory development, where human resource departments may adopt and standardize such tools. In organizational settings, empathy has additionally been linked to higher *employee engagement*, *stronger commitment*, and *healthier workplace climates* [8]. Leaders who clearly demonstrate empathy tend to build trust and loyalty, resolve conflicts more effectively, and strengthen team cohesion. With these outcomes in mind, AI-supported empathy tools may be adapted for leadership training and organizational development, helping future leaders-in-training strengthen their communication and relational skills.

Managers may engage with personas that can simulate common workplace scenarios such as employee conflict, performance concerns, or moments of co-worker distress. These simulations could provide immediate feedback on how responses may be perceived, along with alternative options for refining their communication skills. This reflects the tenets of aspirational and experiential leadership training models through accessible and useful practice simulation.

While the primary focus of the present study was counselor education, the findings do carry some broader implications that meaningfully extend into the domain of programs focused on leadership development. The inherent adaptability and scalability of AI personas, particularly when deployed within a structured facilitator-moderated training environment, can become a viable and significant instrument in organizational settings. When paired with the unlock tier system UTS, institutions may be afforded more deliberate and replicable pathways for then cultivating empathic competencies across a range of diverse professional domains.

LIMITATIONS AND FUTURE DIRECTIONS

The following limitations should be considered when interpreting the findings, and each of these points toward specific future research directions.

Sample Size and Generalizability- The small sample size (N = 10) limited statistical power and generalizability. However, this study was designed as a pilot with the primary aim of assessing feasibility and identifying preliminary patterns rather than establishing causal effects.

Replication with a larger, multi-site, and demographically diverse sample is needed to validate the UTS framework. Thereby, determine whether observed effects generalize across different training contexts and cultural settings.

Absence of a Comparison Group- The lack of a control or comparison condition limits the ability to attribute observed changes solely to the AI–HI interactions. Participant-level factors, such as prior experience with AI personas or dispositional empathy, may have influenced engagement and outcomes.

Future studies should use a waitlist control design or directly compare AI–HI training against standard role-play or reflective journaling methods to isolate the specific contribution of AI persona engagement.

Facilitator Presence- The consistent presence of a facilitator may have unknowingly shaped participant engagement and relational depth.

Future assessments should systematically vary facilitator involvement (for example, fully facilitated vs. partially automated conditions) to uncouple the effects of human facilitation from those of the AI persona interaction itself.

Durability of Effects- Empathy was assessed over a relatively short time frame without longitudinal follow-up. As a result, conclusions regarding the durability of observed changes remain limited.

Longitudinal designs with follow-up research at 3- and 6-month intervals are necessary to evaluate whether gains in empathic responsivity are sustained, decay, or require booster sessions.

Despite these limitations, the study does provide an initial foundational model for examining empathic engagement within structured and facilitated AI-human interactional contexts. Also, it offers clear paths to continue practical study.

CONCLUSION

The aforementioned study contributes to the growing body of evidence on artificial empathy by redirecting attention from the simulation of empathic behaviors within AI systems towards the ways in which empathic engagement can be manifested and meaningfully activated, cultivated within human participants through structured AI-HI encounters. Rather than positioning AI as a purely empathic conduit, the findings imply that AI personas, when thoughtfully constructed and deliberately embedded within guided, intentional, and professionally facilitated helping contexts, hold considerable promise for supporting the activation, expression, and development of empathic potentials in human participants.

The introduction of the Unlocked Tier System (UTS) as a novel convention created for this study provided a methodologically complementary means of observing empathic engagement well beyond the limitations of traditional

self-reporting measures, thus capturing the ways in which empathy could *unfold* across the arc of an encounter with interactional depth and relational exchange being recorded concurrently. In this regard, the present study underscores the value of performance-based indicators as meaningful and ecologically valid additions to existing empathy assessment approaches and the literature on empathy development and measurement.

On a closing note, McStay and Bakir [41] caution that "unintentional empathy" in autonomous systems may introduce regulatory gaps, underscoring the need for governance structures and ethical guardrails in affect-recognition and emotionally responsive technologies. Within this understanding, embracing a "deliberately partnered" approach may offer the profession its most constructive and ethically defensible path forward in the integration of assistive technologies, contingent, however, upon an unwavering preservation of human authority, clinical authenticity, and professional accountability. The imperative then carries particular weight within the confines of mental health, where the stakes of relational integrity, psychological safety, and ethical practices are not merely institutional obligations but fundamental to the well-being and dignity of those served. As effective AI continues to evolve and permeate helping professions, the maintenance of human oversight must not be treated as a peripheral consideration, but rather, as the central and non-negotiable governing principle upon which any responsible integration of such technologies must be predicated.

AUTHORS CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: R.L.: served as the lead and primary investigator, designed the study methodology, constructed the Unlock Tier System (UTS), facilitated all participant sessions as a licensed professional counselor, conducted data analysis, and led all writing, editing, and drafts of the final review. A.S.: contributed to the important theoretical framework and TRE, served as co-investigator, interrater, and participated in manuscript drafting. Chantal Hanna served as Research Assistant, compiled preliminary data, transcribed recorded sessions, and assisted with initial invitations to participants. Both of the authors approved the final manuscript.

LIST OF ABBREVIATIONS

AI = Artificial Intelligence

AI-HI= Artificial Intelligence-Human Interaction

CACREP = Council for Accreditation of Counseling and Related Educational Programs

IRI = Interpersonal Reactivity Index

LLM = Large Language Model

PD = Personal Distress

PT = Perspective Taking

EC= Empathic concern

FS= Fantasy

TRE = Theory of Relational Engagement

UTS = Unlock Tier System

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study received ethics approval from the Yorkville University Research Ethics Board (REB) (Protocol/Certificate No. 202500624RL; approval date June 24, 2025). All participants were adults over the age of 18 and provided electronic informed consent prior to participation. Sessions were conducted on a secure, non-clinical digital communication platform, with no collection of protected health information and full participant de-identification. All consent records and study data were stored in accordance with institutional data-security requirements in a secure, encrypted environment. Participation was voluntary, and participants were informed of their right to withdraw at any

time without penalty. All procedures were conducted in accordance with applicable institutional and ethical standards for research involving human participants. All methods were carried out in accordance with the relevant guidelines and regulations, and the study was conducted in accordance with the principles of the Declaration of Helsinki.

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of this study are stored securely at Yorkville University and are not publicly available due to ethical restrictions related to participant confidentiality and institutional data governance requirements. The Unlock Tier System (UTS), developed by the lead author, is a proprietary analytic framework used in this study. Detailed methodological descriptions are provided in the manuscript; however, the underlying system architecture and embedded scoring logic are not publicly available.

FUNDING

This research was funded by a grant from Yorkville University.

CONFLICT OF INTEREST

The authors declare no institutional, financial, or commercial conflicts of interest related to this study.

ACKNOWLEDGEMENTS

The authors would like to thank the funding bodies for our grant to support the research herein, as well, our YU Research Assistant, CH, for her support and valued assistance. This research is dedicated to Dr. Diane Gavin, a consummate researcher and educator whom I had the pleasure of knowing in this lifetime.

References

1. Rogers CR. The necessary and sufficient conditions of therapeutic personality change. J Consult Psychol. 1957;21(2):95-103. https://doi.org/10.1037/h0045357

2. Hoffman ML. Empathy and Moral Development: Implications for Caring and Justice. Cambridge: Cambridge University Press; 2000. https://doi.org/10.1017/CBO9780511805851

3. Lin J, Wang Y. Exploring Chinese university students' foreign language enjoyment, engagement and willingness to communicate in EFL speaking classes. Humanit Soc Sci Commun. 2025;12:650. https://doi.org/10.1057/s41599-025-04948-z

4. Wang Y, Guo Y. Investigating the Interplay of Chinese English as Foreign Language Learners' Growth Mindset, Enjoyment, Adaptability, and Engagement in AI-Mediated Informal Digital Learning of English. J Comput Assist Learn. 2026;42(1):e70189. https://doi.org/10.1002/jcal.70189

5. Wang X, Gao Y, Reynolds BL. L2 Self-Guides, Achievement Emotions, and Engagement in AI-Mediated Informal Digital Learning of English: Insights from Structural Equation Modeling and Psychological Network Analysis. Asia Pac Educ Res. 2026. https://doi.org/10.1007/s40299-025-01073-y

6. Batson CD. Altruism in Humans. Oxford: Oxford University Press; 2011. https://doi.org/10.1093/acprof:oso/9780195341065.001.0001

7. American Counseling Association. ACA code of ethics. [cited 2026 Apr 18]. https://www.counseling.org/resources/ethics

8. Goleman D, Boyatzis R, McKee A. Primal Leadership: Unleashing the Power of Emotional Intelligence. 10th anniversary ed. Boston: Harvard Business Review Press; 2013.

9. Figley CR. Compassion fatigue: Psychotherapists' chronic lack of self care. J Clin Psychol. 2002;58(11):1433-1441. https://doi.org/10.1002/jclp.10090

10. Neumann M, et al. Empathy decline and its reasons: A systematic review of studies with medical students and residents. Acad Med. 2011;86(8):996-1009. https://doi.org/10.1097/ACM.0b013e318221e615

11. Weng Z, Huang Y, Weng S. From Code to Care: How Artificial Empathy Enhances Customer Experience in Human-Robot Interaction. J Bus Res. 2026;206:115969. https://doi.org/10.1016/j.jbusres.2026.115969

12. Ovsyannikova D, de Mello VO, Inzlicht M. Third-party evaluators perceive AI as more compassionate than expert humans. Commun Psychol. 2025;3(1):4. https://doi.org/10.1038/s44271-024-00182-6

13. Hepner A, Cerutti B, Lüchinger R, Perron NJ. Measuring empathy in OSCEs: A comparison of student, standardized patient, and observer rating tools. PEC Innov. 2026;8:100452. https://doi.org/10.1016/j.pecinn.2025.100452

14. Davis MH. A multidimensional approach to individual differences in empathy. JSAS Cat Sel Doc Psychol. 1980;10:85.

15. Davis MH. Measuring individual differences in empathy: Evidence for a multidimensional approach. J Pers Soc Psychol. 1983;44(1):113-126. https://doi.org/10.1037/0022-3514.44.1.113

16. Chrysikou EG, Thompson WJ. Assessing cognitive and affective empathy through the Interpersonal Reactivity Index: An argument against a two-factor model. Assessment. 2016;23(6):769-777. https://doi.org/10.1177/1073191115599055

17. Mehrabian A, Epstein N. A measure of emotional empathy. J Pers. 1972;40(4):525-543. https://doi.org/10.1111/j.1467-6494.1972.tb00078.x

18. Hojat M, Gonnella JS, Mangione S, Nasca TJ, Magee M. Physician empathy in medical education and practice: Experience with the Jefferson Scale of Physician Empathy. Semin Integr Med. 2003;1(1):25-41. https://doi.org/10.1016/S1543-1150(03)00002-4

19. Zaki J, Ochsner K. The neuroscience of empathy: Progress, pitfalls and promise. Nat Neurosci. 2012;15(5):675-680. https://doi.org/10.1038/nn.3085

20. Truax CB, Carkhuff RR. Toward Effective Counseling and Psychotherapy: Training and Practice. Chicago: Aldine Publishing Company; 1967.

21. Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. Empathy for Pain Involves the Affective but Not Sensory Components of Pain. Science. 2004;303(5661):1157-1162. https://doi.org/10.1126/science.1093535

22. Luster R. The potentials of artificial intelligence and psychotherapy: Are we there yet? Phoenix Scholar. 2024;7(1):2-9.

23. Montemayor C, Halpern J, Fairweather A. In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. AI Soc. 2022;37(4):1353-1359. https://doi.org/10.1007/s00146-021-01230-z

24. Jiang Q, Zhang Y, Pian W. Chatbot as an emergency exit: Mediated empathy for resilience via human-AI interaction during the COVID-19 pandemic. Inf Process Manag. 2022;59(6):103074. https://doi.org/10.1016/j.ipm.2022.103074

25. Shen J, DiPaola D, Ali S, Sap M, Park HW, Breazeal C. Empathy toward artificial intelligence versus human experiences and the role of transparency in mental health chatbot design: Comparative study. JMIR Ment Health. 2024;11:e62679. https://doi.org/10.2196/62679

26. Hill CE, Lent RW. A narrative and meta-analytic review of helping skills training: Time to revive a dormant area of inquiry. Psychotherapy: Theory, Research, Practice, Training. 2006;43(2):154-172. https://doi.org/10.1037/0033-3204.43.2.154

27. Smith A. Artificial Relational Intelligence (ARI): Relational engagement and the future of human-AI interaction. ARI Applications. 2026. Manuscript in preparation.

28. Zaki J. Empathy: A motivated account. Psychol Bull. 2014;140(6):1608-1647. https://doi.org/10.1037/a0037679

29. Zhang S, Qian Y, Yao Z, Ni Z, Zhang Y, et al. From approach to avoidance: How AI agent cognitive and affective empathy elicits the uncanny valley effect. Telemat Inform. 2025;101:102313. https://doi.org/10.1016/j.tele.2025.102313

30. Cao S, Fu D, Yang X, Wermter S, Liu X, Wu H. Pain recognition and pain empathy from a human-centered AI perspective. iScience. 2024;27(8):110570. https://doi.org/10.1016/j.isci.2024.110570

31. Birnbaum GE, Reis HT. Responsiveness: Affective interdependence in close relationships. In: Mikulincer M, Shaver PR, editors. Mechanisms of social connection: From brain to group. Washington (DC): American Psychological Association; 2014. p. 255-271. https://psycnet.apa.org/doi/10.1037/14250-015

32. Kahn PH Jr, Kanda T, Ishiguro H, Gill BT, Shen S, Gary HE, Ruckert JH. Will people keep the secret of a humanoid robot? Psychological intimacy in HRI. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction; 2015 Mar 2-5; Portland (OR), USA. p. 173-180. https://doi.org/10.1145/2696454.2696486

33. Föyen LF, Zapel E, Lekander M, Hedman-Lagerlöf E, Linsäter E. Artificial intelligence vs. human expert: Licensed mental health clinicians' blinded evaluation of AI-generated and expert psychological advice on quality, empathy, and perceived authorship. Internet Interv. 2025;41:100841. https://doi.org/10.1016/j.invent.2025.100841

34. Ghosh AK, Rentzsch M, Ura S. Fostering Empathy in AI-Driven Co-Creation of CAD Models. Procedia CIRP. 2025;134:741-746. https://doi.org/10.1016/j.procir.2025.02.206

35. Inzlicht M, Cameron CD, D'Cruz J, Bloom P. In praise of empathic AI. Trends Cogn Sci. 2024;28(2):89-92. https://doi.org/10.1016/j.tics.2023.12.003

36. Rubin M, Arnon H, Huppert JD, Perry A. Considering the Role of Human Empathy in AI-Driven Therapy. JMIR Ment Health. 2024;11:e56529. https://doi.org/10.2196/56529

37. Council for Accreditation of Counseling and Related Educational Programs (CACREP). 2016 CACREP Standards. [cited 2026 Apr 18]. https://www.cacrep.org/for-programs/2016-cacrep-standards/

38. Chen D, Chauhan K, Parsa R, Liu ZA, Liu FF, Mak E, Eng L, Hannon BL, Croke J, Hope A, Fallah-Rad N. Patient perceptions of empathy in physician and artificial intelligence chatbot responses to patient questions about cancer. NPJ Digit Med. 2025;8(1):275. https://doi.org/10.1038/s41746-025-01671-6

39. Knudsen HK, Roman PM, Abraham AJ. Quality of clinical supervision and counselor emotional exhaustion: The potential mediating roles of organizational and occupational commitment. J Subst Abuse Treat. 2013;44(5):528-533. https://doi.org/10.1016/j.jsat.2012.12.003

40. Gustavsson M, Bivall AC. The challenges clinical supervisors experience when supervising students in the workplace. High Educ Skills Work-based Learn. 2025;15(7):127-138. https://doi.org/10.1108/HESWBL-11-2024-0347

41. McStay A, Bakir V. Soft law for unintentional empathy: Addressing the governance gap in emotion-recognition AI technologies. J Responsible Technol. 2025;23:100126. https://doi.org/10.1016/j.jrt.2025.100126

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx