

Alter egos alter engagement: perspective-taking can improve disclosure quantity and depth to AI chatbots in promoting mental wellbeing

Christopher You^{1,*}, Rashi Ghosh¹, Melissa Vilaro², Roshan Venkatakrishnan³, Rohith Venkatakrishnan³, Andrew Maxim¹, Xuening Peng¹, Danish Tamboli¹, and Benjamin Lok¹

¹ *Virtual Experiences Research Group, University of Florida, Department of Computer & Information Science & Engineering, Gainesville, FL, USA*

² *University of Florida, Institute of Food & Agricultural Sciences, Gainesville, FL, USA*

³ *University of Central Florida, Department of Computer Science, Orlando, FL, USA*

Correspondence*:
Christopher You
christopheryou@ufl.edu

ABSTRACT

Introduction: Emotionally intelligent AI chatbots are increasingly used to support college students' mental wellbeing. Yet, adoption remains limited, as users often hesitate to open up due to emotional barriers and vulnerability. Improving chatbot design may reduce some barriers, but users still bear the emotional burden of opening up and overcoming vulnerability. This study explores whether *perspective-taking* can support user disclosure by addressing underlying psychological barriers.

Methods: In this between-subjects study, 96 students engaged in a brief reflective conversation with an embodied AI chatbot. **Perspective-Taking** participants defined and imagined a designated other's perspective and responded from that viewpoint. **Control** participants provided self-information and responded from their own perspective. Disclosure was measured by quantity (word count) and depth (information, thoughts, and feelings). Additional immediate measures captured readiness, intentions for mental wellbeing, and attitudes toward the chatbot and intervention.

Results: **Perspective-Taking** participants disclosed significantly greater quantity, overall depth, thoughts depth, and frequencies of high disclosures of thoughts and information. Both groups showed significant improvements in readiness and intention to address mental wellbeing, with no difference in improvement magnitude. However, **Control** participants reported significantly lower (better) skepticism towards the intervention and greater increases in willingness to engage with AI chatbots comparatively.

Discussion: This study highlights how perspective-taking and distancing may facilitate greater disclosure to AI chatbots supporting mental wellbeing. We explore the nature of these disclosures and how perspective-taking may drive readiness and enrich the substance of disclosures. These

findings suggest a way for chatbots to evoke deeper reflection and effective support while potentially reducing the need to share sensitive personal self-information directly with generative AI systems.

Keywords: artificial intelligence, chatbot, mental wellbeing, perspective-taking, disclosure, emotional expression

1 INTRODUCTION

Amid the excitement and rigor of college life, many students encounter mental health challenges that can feel overwhelming and isolating. Recent surveys reveal that over 60% of U.S. college students report experiencing at least one mental health-related issue (e.g., stress, anxiety, depressive symptoms) during their education (Lipson et al., 2022). The most recent Healthy Minds dataset (2023–2024) paints an even starker picture: 78% of students currently indicate some level of need with emotional or wellbeing challenges, yet only 54% had ever reached out to professional counseling, with only 36% doing so in the prior year. Although recent trends may reflect more positivity, available early data reported a median delay of *11 years* between the onset of symptoms and initial treatment among a general U.S. population (Wang et al., 2004). In response to the rising demand and persistent barriers, research has increasingly turned to digital mental wellbeing support through emotionally intelligent AI chatbots. While other telehealth modalities also aim to expand access, such chatbots offer unique advantages by mitigating time, availability, and location barriers through their asynchronous nature. Numerous studies suggest that emotionally intelligent AI chatbots can even provide interim support for depressive symptoms (Gardiner et al., 2017; Fitzpatrick et al., 2017; Bickmore et al., 2010a; Schroeder et al., 2018; Abd-Alrazaq et al., 2020). At the same time, recent systematic and meta reviews also highlight limited effectiveness and inconsistent results in addressing mental health concerns (Feng et al., 2025; Limpanopparat et al., 2024). Hence, another promising application lies in chatbots' abilities to empower users to proactively manage their wellbeing or seek additional support from professionals (Casu et al., 2024). This approach is motivated by evidence suggesting that replacing human care with automated systems in therapeutic contexts can leave users feeling discomfort and reluctance in deeper engagements (You et al., 2022; Brown and Halpern, 2021; Zhang and Wang, 2024; Khawaja and Bélisle-Pipon, 2023). Such self-empowering chatbots have been leveraged to support goal-setting (Aggarwal et al., 2023), adhere to medication goals (Bickmore et al., 2010b), drive engagement with online therapy (Yasukawa et al., 2024), promote smoking abstinence (Olano-Espinosa et al., 2022), or increase efficacy in addressing eating disorders (Fitzsimmons-Craft et al., 2022; Shah et al., 2022). Rapid advancements of large language models also further improve emotionally intelligent AI chatbots' abilities to overcome obstacles to promote self-wellbeing (Lawrence et al., 2024).

Despite promising developments, meaningful engagement with chatbots for wellbeing is far from guaranteed. Recent reports indicate that U.S. college student adoption of chatbot mental health services remains limited, and attitudes are significantly more negative compared to traditional services (Rackoff et al., 2025). Lingering reluctance may result in limited self-disclosure, shallow interaction patterns, and reduced ability to provide meaningful support (Nguyen et al., 2012; Hill et al., 2015; Carmichael et al., 2022; Crutzen et al., 2014; Coghlan et al., 2023). Users may also abandon chatbot interactions due to technical issues, a perceived lack of human emotion or empathy, or doubts in the chatbot's ability to provide meaningful support (You et al., 2022; Saadati et al., 2023). Achieving meaningful engagement with chatbots often requires users to self-disclose content that might not otherwise be disclosed due to vulnerability or discomfort. If the goal is to help users feel safe enough to share, then we must also address the psychological mechanisms that underlie emotional risk itself. To do so, many approaches aim to enhance chatbots to normalize stigmas (Cui et al., 2024; Branley-Bell et al., 2023) or to become more accommodating in their conversations (e.g., more empathetic, human-like, acceptable) (Zhai and Wibowo, 2022; Inkster et al., 2018; Edalat et al., 2025; Grové, 2021; Ennis et al., 2023; Denecke et al., 2021; Chin et al., 2023). However, the emotional burden still largely falls on users: users must choose to open up, risk being vulnerable, and overcome deeply personal inhibitions. Additional challenges arise with respect to privacy and ethical concerns in engaging with AI. Uncertainty in data storage, access,

and confidentiality may impede disclosures (Vaidyam et al., 2019). Moreover, ethical concerns around propagating prejudice due to algorithm bias and limited capabilities for responding to crises raise critical questions about the safety, fairness, and reliability of AI-driven mental health interventions (Huo et al., 2025; Boucher et al., 2021). Even if such scenarios are ‘safe’, deeper consideration for the user may be needed to foster disclosure in such environments.

In an effort to mediate user reluctance in disclosing to chatbots, we propose employing *perspective-taking*. Taking another’s perspective allows one to “discern the thoughts, feelings, and motivations of [others]” (Gehlbach et al., 2012) and offers emotional distance to reflect on distressing experiences (Kross and Ayduk, 2011; Liberman et al., 2007; Gehlbach et al., 2012; Boland et al., 2021). Given that users often adopt altered identities in digital contexts (Bargh et al., 2002; Bullingham and Vasconcelos, 2013), interactions with chatbots may offer a unique opportunity to explore identity through perspective taking. We conceptualize taken perspectives in relation to the “self” and “other”, where “other” refers to external entities, such as strangers, friends, family, or even hypothetical entities (Liberman and Trope, 2014; Liberman et al., 2007). The ability to take perspectives is one formed in early childhood that is considered critical to empathetic capability (Piaget, 2013; Kohlberg, 1976). Empathy is often attributed to the ability to take perspectives (Batson et al., 1997a): we *share* in other people’s emotions (Eisenberg et al., 2006) and *reconstruct* their mental states for ourselves (Hogan, 1969; Davis and Love, 2017). Though seemingly intuitive, perspective-taking relies on concrete knowledge to ground inferences about others’ actions (Davis et al., 2004) and is facilitated by greater self–other overlap (Ames et al., 2008). As a result, less informed perspectives (i.e., distal constructs) can result in *abstractions*, or the employment of general heuristics and social rules to estimate behavior (Liberman et al., 2007; Trope and Liberman, 2010). Pertinent examples include abstract syntax in speech (e.g., more adjectives than descriptive verbs) (Semin and Fiedler, 1988; Liberman et al., 2007), higher-level terminology in descriptions (Vallacher and Wegner, 1989; Liberman and Trope, 1998; Hampson et al., 1986), or more polite, indirect language (Brown and Levinson, 1987; Semin and Fiedler, 1988).

The primary motivation for perspective-taking in this study is its demonstrated impact on user behavior, attitudes, and outcomes. Perspective-taking can lead to improved prosocial behaviors and intentions of change (Batson et al., 2002; Van Loon et al., 2018; Galinsky and Moskowitz, 2000; Uhl-Haedicke et al., 2019). Pahl & Bauer found that briefly adopting the perspective of a young woman affected by environmental changes increased participants’ engagement with environmental materials and enhanced their intentions toward environmental action (Pahl and Bauer, 2013). Perspective-taking with outgroups has been shown to reduce aggression, increase empathy, and diminish stereotyping and bias (Richardson et al., 1994; Shechtman and Tanus, 2006; Galinsky and Ku, 2004; Vescio et al., 2003; Seinfeld et al., 2018; Shaffer et al., 2019). Perspective-taking may also lead to improved outcomes for the self (Blatt et al., 2010; Boland et al., 2021; Taylor and Brown, 1988; Hodges et al., 2010). Boland et al. found that adopting the perspectives of past selves or others, by receiving or offering compassion, can reduce emotional discomfort and enhance self-compassion (Boland et al., 2021). Perspective-taking is shaped by factors like altruistic concern and egotistic motivation (Batson, 2014; Cialdini et al., 1987), yet a compelling effect may stem from the merging of self and other (Batson et al., 1997c; Goldstein and Cialdini, 2007; Ames et al., 2008; Hodges et al., 2010). Aron and Aron describe this as incorporating others into the self, ultimately viewing them as extensions of oneself (Aron and Aron, 1986). Perspective-takers are thought to internalize the insights, thoughts, and emotions of others (Cialdini et al., 1997; Davis et al., 1996). Perspective-takers may project their own traits onto others (e.g., “I liked this movie; therefore, my friend will too”) (Liberman et al., 2007; Galinsky and Moskowitz, 2000). They may also adopt traits of others, as seen when taking a professor’s perspective increased self-ratings of intelligence (Galinsky et al., 2008). These effects intensify

when individuals internalize others' experiences, emotions, and attributes as part of their self-concept (De Cremer, 2004; Batson et al., 1997c; Krebs, 1975).

Similar to prior literature (Pahl and Bauer, 2013; Uhl-Haedicke et al., 2019; Blatt et al., 2010), this study investigates perspective-taking as a means to promote behavioral engagement, rather than investigating underlying mechanisms of overlap or attitudes towards the taken other. Specifically, This study examines how perspective-taking can enhance disclosure in conversations with emotionally intelligent AI chatbots. Our utilization of an emotionally intelligent AI chatbot entails an embodied conversational agent (ECA)-guided reflective conversation for addressing ambivalence to change, where theory-driven, AI-generated empathetic expressions of dialogue are adapted and delivered based on individual user disclosures (see Section 2.2.2). Although current approaches have demonstrated that AI systems can detect and express emotion with potential for higher sophistication (Rupp et al., 2025; Gonzalez-Acosta et al., 2025; Edalat et al., 2025; Valderrama and Sheoran, 2025; Coelho et al., 2025), the present study's integration of an emotionally intelligent AI conversation serves as a *platform* to empirically evaluate perspective-taking. The findings of this work arise from a between-participants study that recruited primarily STEM students from the University of Florida who were randomized into **Perspective-Taking** and **Control** conditions. **Perspective-Taking** participants took the perspective of a self-defined, known other and engaged in the AI-guided conversation fully from the other's perspective; **Control** participants completed identical tasks with a self-framing and engaged fully from their self-perspective. This study investigated perspective-taking's effects on the following measures: disclosure (word *quantity* and categorical *depth*), readiness to address mental wellbeing, and attitudes toward the intervention and chatbot. Our main hypotheses predicted **Perspective-Taking** would enhance engagement in the forms of greater disclosure quantities and depths in comparison to the **Control**:

H1_{DisclosureQuantity}: **Perspective-Taking** participants will exhibit greater quantities of disclosure compared to **Control** participants.

H2_{DisclosureDepth}: **Perspective-Taking** participants will exhibit greater depths of disclosure compared to **Control** participants.

Secondary to disclosure, we investigated how taking an other's perspective impacted readiness and attitudes. We hypothesized **Perspective-Taking** participants' readiness to address mental wellbeing would significantly improve overall from Pre- to Post-measure.

H3_{ReadinessOverall}: **Perspective-Taking** participants will exhibit improved readiness after the reflective conversation on wellbeing.

Though it was not fully expected that perspective-taking would outperform a self-perspective in a conversation dictated by disclosure and self-reflection, we deferred the remaining outcome hypothesis in favor of the **Perspective-Taking** (experimental) condition:

H4_{ReadinessComparison}: **Perspective-Taking** participants will exhibit greater improvements on readiness compared to **Control** participants.

H5_{AttitudesIntervention}: **Perspective-Taking** participants will exhibit more positive attitudes towards the present wellbeing intervention compared to **Control** participants.

H6_{AttitudesChatbots}: **Perspective-Taking** participants will exhibit greater improvements on attitudes towards AI wellbeing chatbots compared to **Control** participants.

2 MATERIALS AND METHODS

Two main conditions were examined to assess the impact of perspective-taking: **Perspective-Taking** (perspective of an other) and **Control** (perspective of self). All participants completed two main intervention steps consisting of perspective-taking tasks and a reflective conversation. The primary difference is in the framing of the tasks themselves (see Table 1).

Condition	Perspective	Task Framing
Perspective-Taking	Other	<i>Imagine someone in your life who would benefit from discussing their mental health through an online intervention. Step into their perspective, consider their experiences, and engage in the tasks ahead... To start, type: "I am ready to play the role of [alias]."</i>
Control	Self	<i>Engage in this self-intake by reflecting on your own experiences and reasons for discussing your mental wellbeing. Consider your perspective, experiences, and engage in the tasks ahead... To start, type: "I am ready to begin the conversation."</i>

Table 1. Summary of the study conditions, their perspectives, and task framing. Both conditions completed the same steps; however, the framing of the tasks differed based on whether participants took an other's perspective (**Perspective-Taking**) or engaged as the self (**Control**). A high-level overview of the framing is provided, but see Section 2.1 for specific details.

2.1 Study design

Prior to the study start, participants selected between a male and female ECA for the remainder of the interaction. The ECA introduced itself and provided an overview of the study based on the participant's assigned condition. The study consisted of two main phases of interaction described in this section: *perspective-taking* and *reflective conversation*.

2.1.1 Perspective-taking phase

Participants in the **Perspective-Taking** condition were instructed to identify, describe, and imagine another's perspective during this phase. Rather than being given a fictional persona, participants chose a real person in their life, such as a friend or family member, who would benefit from the reflective conversation. The decision to have participants define a real, known other who might benefit from the interaction was based on prior research and considerations specific to this study. First, perspective-taking may fail due to insufficient information or a lack of reason to set aside egocentric bias and take perspectives (Epley and Caruso, 2012; Kavanagh et al., 2020), and greater proximity will increase the likelihood of adopting an other's perspective (Batson, 2014). Allowing participants to define a known other who might benefit from the conversation offers greater familiarity with the other's mental wellbeing and a meaningful reason to take their perspective. Second, establishing conversational depth was a high priority for the study's wellbeing aims. An other who is too distant from the participant may be difficult to portray, leading to more abstract, higher-level responses (Liberman et al., 2007). Given strong evidence that people can spontaneously take perspectives or empathize without prompting (Furlanetto et al., 2013; Tversky and

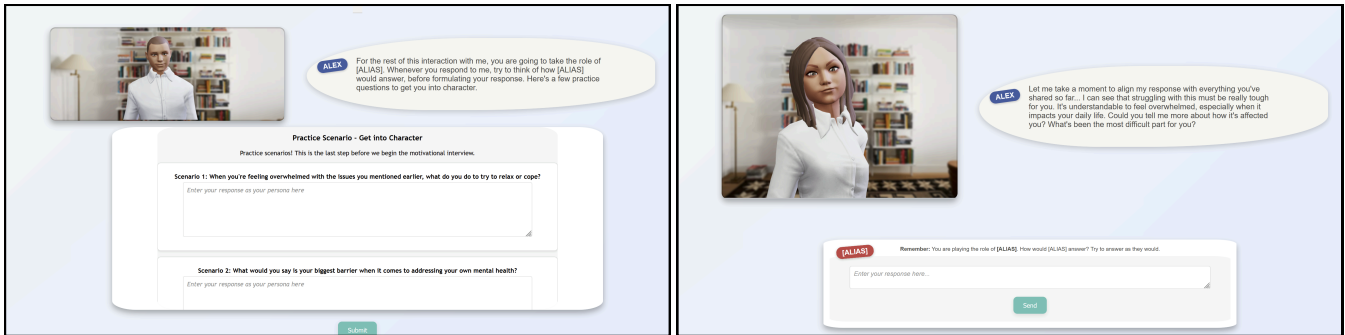


Figure 1. The intervention interfaces illustrating the male and female ECA. (Left) mock conversation scenario in perspective-taking phase and (Right) sample disclosure item Q2 in reflective conversation phase.

Hard, 2009; Davis, 2018; Batson et al., 1991), allowing participants to define their own target seemed suitable for supporting perspective-taking in this context.

To support effective perspective-taking, the task extended beyond the typical narrative and imaginative phases commonly used in prior studies (Pahl and Bauer, 2013; Batson et al., 1997a; Weyant, 2007). This study used a persona-crafting task called empathy mapping, which aligns with established perspective-taking processes (Davis et al., 2004). Given empathy's strong connection to perspective-taking (Batson et al., 1997b, 2002, 1991), empathy mapping was appropriate, as it helps participants understand others by viewing the world through their eyes and evokes empathy through persona design (Ferreira et al., 2015). Empathy mapping aimed to personify the other by capturing demographic information, personality traits, values, wellbeing concerns, goals, and brief imaginative descriptions of the other's life. These fields were further informed based on literature in empathy mapping and design of patient personas in eHealth interventions (Bland, 2012; Bartels et al., 2023; Ledel Solem et al., 2020).

The AI chatbot guided participants through the empathy mapping tasks, explaining the aims and requirements for each prompted item. After designing the persona, participants reviewed their other's persona and imagined their other's perspective and experiences. To deepen perspective-taking and ensure participants responded only from their other's viewpoint, the AI chatbot provided mock conversation prompts for first-person replies as the other. Afterwards, participants began the reflective conversation by entering the following phrase: "I am ready to play the role of [alias]" (where "alias" refers to the other's defined name). In the **Control** condition, the same empathy mapping and mock scenario tasks were delivered by the AI chatbot to maintain structural parity. The difference in conditions arises in the framing of the task as a personal intake in the **Control**, rather than perspective-taking.

2.1.2 Reflective conversation phase

The reflective conversation was designed to evoke disclosure from participants conducive to building their readiness to address their wellbeing. Therefore, the conversation was designed using principles from motivational interviewing, a client-centered approach to enhance readiness, resolve ambivalence, and encourage capability and autonomy (Hettema et al., 2005; Rollnick and Miller, 1995). To enable meaningful comparison of disclosure quantity and depth, a fixed set of conversation items replaced the open-ended format typical of motivational interviews. This design allowed participants to receive the same, verbatim open-ended questions across interactions. Furthermore, the AI chatbot was designed to convey empathy and emotion in response, which can influence disclosure attitudes (Ludwig et al., 2022; Aron et al., 1997). A structured conversation also afforded consistency in the quantity and depth of empathetic

expression from the AI chatbot. The rule-based reflective conversation script was designed by two authors MV, an expert in health communication trained in motivational interviewing, and CY, who received the standard full-day training in motivational interviewing (Söderlund et al., 2011). A formal pilot with (n = 58) participants was conducted while iteratively updating the conversation's script.

Four classifications of motivational interviewing strategies reviewed by Hardcastle et al. served as sub-phases for the present reflective conversation (Hardcastle et al., 2017). In line with motivational interviewing's open nature, the authors describe their classifications as identified themes rather than *strict* design requirements for conversations. Their classifications include motivational interviewing strategies for *engaging*, *focusing*, *evoking*, and *planning*. Although the conversation includes both closed- and open-ended items per Hardcastle et al., we focus here on the *nine* open-ended disclosure items (see Section 2.3.1). These nine disclosure items include strategies directly from three of the four classifications: **Engaging**: two "open-ended question" disclosure items, **Evoking**: five disclosure items on "troubleshooting" barriers to change, "looking forward" on future possibilities, "identifying past successes" in coping strategies, "exploring values" relating to the wellbeing concern or behavior, and "brainstorming" options to change, and **Planning**: two disclosure items of "considering change options" and "developing a change plan" towards one concrete, self-designed next step (see Table 2) (Hardcastle et al., 2017). In *Focusing*, participants were provided the opportunities to engage in a set of resources (NIMH & CDC) containing techniques for improving mental wellbeing through closed-ended responses.

Participants completed the reflective conversation by responding to the AI chatbot's nine disclosure items and additional closed-ended items across the four sub-phases. With each disclosure response from the participant, the AI chatbot employed a sequence of providing empathetic expression before delivering the ensuing prompt (see Section 2.2.2 for implementation details). To ensure proper study completion, static interface messages reminded participants to engage from the defined perspective, and the AI chatbot's first question requested participants to provide their defined alias for conversation. Participants who provided an alias differing from their taken perspective were removed from analyses. All participants completed the reflective conversation's disclosure and closed-ended items across four sub-phases. **Perspective-Taking** participants responded from the designated other's perspective, while **Control** participants responded from their own. Following the conversation, participants engaged in a short transitional phase to return to their own perspective, labeled self-reflection in the **Control** group. During this transition, participants were told to momentarily pause to reorient to their own perspective and experiences and self-reflect on the conversation. **Perspective-Taking** participants were reminded to complete post-surveys from their own perspective, as in the pre-survey.

2.2 AI chatbot and empathetic expression protocol

This section describes the architecture for the AI chatbot and the design for empathetic expression in responses. The AI chatbot interaction is built on a Node.js framework, which is commonly used to build and deploy web applications. The study is deployed asynchronously over the web, where participants were required to complete the study on a desktop or laptop web-enabled device.

2.2.1 AI chatbot

The generation of ECAs, their verbal responses, and their corresponding non-verbal behaviors are described. ECAs were employed as evidence has indicated they can provide a level of human touch and foster greater willingness to disclose (Pickard et al., 2016; Loveys et al., 2022; Lucas et al., 2014). Each ECA is designed via ReadyPlayerMe, a free-to-use online tool to generate 3D models that can be rigged,

rendered, and utilized on the web using a Three.js library. ReadyPlayerMe provides integrated blendshapes to the model, which support the non-verbal behaviors of animation and lip-syncing. A male and female ECA options were generated (see Figure 1) and included in the pilot tests with the (n = 58) participants to broadly check for any negative sentiments in design choices. Upon accessing the study's webpage, the ReadyPlayerMe-exported ECA model is loaded and rendered on users' devices.

When a participant interacts with the ECA, verbal responses containing text and audio are generated statically or dynamically using the rule-based conversation script and LLMs. To generate the verbal response, OpenAI's Completions (4o-mini) and Text-To-Speech (tts-1) models are employed (male voice: *echo*; female voice: *shimmer*). The conversation script indicates how text and audio responses should be *statically* or *dynamically* generated. Static verbal responses are pre-generated to control the interactions so that participants receive identical responses when necessary. Static verbal responses from the ECA include the nine disclosure items in the reflective conversation or the empathy mapping items in the perspective-taking phase. Both conditions followed the same rule-based conversation script during the reflective conversation, producing identical static verbal responses. However, the perspective-taking phase used two distinct scripts due to differences in the empathy mapping framing. In contrast, dynamic verbal responses are generated in real-time based on individual participant queries across both conditions. Dynamic verbal responses largely pertain to the empathetic expressions delivered in the *reflective conversation*. Where the script calls for a dynamic verbal response, an empathetic expression strategy guides the LLM (see Section 2.2.2). The ECA immediately responds with a verbal backchannel (e.g., "Thanks for being open. I'm working on generating something thoughtful based on what you've shared.") to acknowledge input and mask LLM response delays. The system generates dynamic verbal responses and queues them to deliver after the verbal backchannel finishes.

The ECAs perform non-verbal behaviors such as animations and lip-syncing that correspond to their verbal responses, implemented using the open-source repository *TalkingHead*¹. For animations, the ECA employs template behaviors when idling or speaking (e.g., standing straight, leaning to the side, gestures). When the ECA is idle, a sequence of randomized idle poses with a generic breathing animation is rendered. When the ECA is speaking (i.e., when a verbal response is delivered), additional animations were integrated alongside the randomized talking poses (e.g., a wave when the ECA introduces itself). For lip-syncing, transcriptions and timestamps of the spoken audio are derived from the verbal responses. The transcription is processed to extract individual phonemes, which are mapped to corresponding visemes. These visemes are coded in the lip-sync system using Oculus Lipsync and *TalkingHead*. Timestamps show when to apply visemes to the ECA's facial blendshapes during playback to simulate natural speech. In sum, a typical conversation turn will entail: receiving user input, delivering verbal backchannels, generating appropriate (dynamic) verbal response for empathetic expression, retrieving subsequent (static) verbal response for disclosure or closed-ended item, delivering entire verbal response, animating non-verbal behaviors, and synchronizing lip movements to verbal response.

2.2.2 Empathetic expressions of dialogue

While numerous articles explore opportunities to detect and convey emotion accordingly (Rupp et al., 2025; Gonzalez-Acosta et al., 2025; Edalat et al., 2025; Valderrama and Sheoran, 2025; Coelho et al., 2025), the present work primarily focuses on how emotionally intelligent systems can be further enhanced by psychological theories of perspective-taking and self-distancing. However, establishing emotional intelligence from the AI chatbot remains critical to the study's mental wellbeing design and in

¹ <https://github.com/met4citizen/TalkingHead>

Conversation Phase	Disclosure Items	Disclosure Strategies	Empathetic Expression Strategies
<i>Engaging</i>	Q1	Open-ended question	Offer emotional support and reflective statements
	Q2	Open-ended question	Offer emotional support and reframing
<i>Focusing</i>	N/A	N/A	N/A
<i>Evoking</i>	Q3	Troubleshooting	Coming alongside and normalizing
	Q4	Looking forward	Affirmations and emphasize autonomy
	Q5	Identify past successes	Reframing and affirmations
	Q6	Values exploration	Offer emotional support and reflective statements
	Q7	Brainstorming	Reframing and agreement with a twist
<i>Planning</i>	Q8	Considering change options	Permission to provide information and advice and emphasize autonomy
	Q9	Develop a change plan	Support change/persistence and emphasize autonomy

Table 2. The specific strategies employed for the conversation's nine disclosure items and empathetic expressions. The disclosure strategies refer to the nine open-input items analyzed, and the empathetic expression strategies illustrate how prompts were designed (Hardcastle et al., 2017).

understanding how perspective-taking can enhance such chatbots. Thus, we designed the AI chatbot to convey empathetic expression strategically to individual participant disclosures. In health communications, there are *opportunities* (when) empathy must be conveyed and corresponding *expressions* or representations (what) of empathy (Bylund and Makoul, 2005, 2002). With AI chatbots, frameworks suggest that similar processes of recognition and communication can be administered (Lin et al., 2023; Yalçın, 2020). Focusing on perspective-taking, this study aims to streamline the process using a rule-based, structured approach with LLMs. The rules determine *when* to prompt during the nine open-ended disclosure items, while the LLMs use motivational interviewing strategies to generate *what* to say through empathetic expressions.

The AI chatbot's empathetic expressions are based on Hardcastle et al.'s motivational interviewing strategy classifications, which also guided the design of the disclosure items (Hardcastle et al., 2017). The present study's empathetic expression strategies are primarily derived from relational strategies within the four classifications. The nature of the previous disclosure item primarily guides the empathetic expression used in the *reflective conversation* script. Open-ended questions like Q1 and Q2 engage participants by asking them to describe a mental wellbeing concern or goal and its impact on them. Hardcastle et al.'s strategies of offering emotional support, summarization/reflective statements, and reframing help participants feel heard and encourage reflection. In questions that help the participant *plan*, like in Q8 and Q9, it is more important to emphasize autonomy in the participant's choice and support their change and persistence (Hettema et al., 2005). Table 2 illustrates the specific disclosure and empathetic expression strategies utilized in the present study, as listed directly from Hardcastle et al.'s classifications (Hardcastle et al.,

2017). For each disclosure, emotional dialogue expressions in dynamic verbal responses are generated by identifying the relevant theory-driven strategy, adapting to user disclosures and conversation history, and prompting the AI model accordingly. By anchoring each empathetic expression in an established theoretical classification, the present study provides a controlled and interpretable environment to assess the impact of perspective-taking in conversations within AI chatbots that are intelligent to user disclosures with their expressions of emotion.

2.3 Measures

To address the hypotheses, three primary constructs were investigated: disclosure, readiness, and attitudes.

2.3.1 Disclosure

Based on prior literature (Carpenter and Greene, 2015; Lee et al., 2020), disclosure is assessed through measures of **quantity** and **depth**. For $H1_{\text{DisclosureQuantity}}$, LIWC-22 was used to capture *word counts* across the nine disclosure items to determine if the disclosed quantity of words was altered by the perspective-taking manipulation (Boyd et al., 2022). To supplement analyses on disclosure quantity, we calculated an *abstractness* score (1 - 5) across participants' nine disclosure responses, using the Linguistic Category Model (LCM) (Semin and Fiedler, 1991; Semin et al., 2002). Seih et al.'s generated LIWC-22 dictionary and their described process for using the *TreeTagger*² tool was used to capture frequencies for LCM (Seih et al., 2017; Johnson-Grey et al., 2020).

For $H2_{\text{DisclosureDepth}}$, qualitative analysis was performed using the process and categories defined by Barak & Gluck-Ofri to code each of the nine participant responses in terms of *information*, *thoughts*, and *feelings* (Barak and Gluck-Ofri, 2007). Each response was segmented into distinct statements and categorized as follows: *information*, when the writer shared personal details, experiences, or factual content; *thoughts*, when they expressed personal opinions or reflections; and *feelings*, when they conveyed emotional or affective responses. Within each category, one of three levels of depth is assigned: 1 - no disclosure about the user in the category altogether, 2 - a disclosure about the user but in general or mild expressions, 3 - a disclosure about the user in personally revealing, intimate, or deep expressions. Therefore, each response will have resulted in a score (1 - 3) for all categories of information, thoughts, and feelings. An *overall depth* score for the amount of disclosure is obtained by combining the levels of information, thoughts, and feelings for each response (Barak and Gluck-Ofri, 2007). Sample responses categorized as depth levels 1, 2, and 3 for each category can be found in Table 3.

A total of 96 participants properly completed the entire intervention, but disclosure analysis includes 55 participants' responses to the disclosure items: **Control** ($n = 29$ participants \times 9 items = 261 items) and **Perspective-Taking** ($n = 26$ participants \times 9 items = 234 items). Technical errors early in data logging prevented the capture of conversation logs for the outstanding participants. The resulting disclosure analysis includes a robust set of ($n = 493$ items \times 3 codes = **1479**) codes, after validating responses and omitting ($n = 2$) responses due to invalid input. Authors AM, DT, and XP served as three independent coders with condition- and participant-anonymized, shuffled versions of the conversation transcripts. Each author had prior experience in qualitative methods and received training on the process by Barak & Gluck-Ofri before individually coding the same 30% subset of the data ($n = 615$ codes). Kendall's W indicated the three coders statistically significantly agreed in their assessments, $W = .856, p < .001$. Disputes within responses were settled as a group, and each coder individually coded a third of the remaining data. See Table 4 for quantities of depth at each level across participants' nine disclosure items.

² <https://www.ims.uni-stuttgart.de/en/research/resources/tools/treetagger/>

	Information	Thoughts	Feelings
1: No disclosure	[P33] My girlfriend is the best	[P44] Projects take a long time to complete	[P46] I have used breathing techniques and meditation. also talking to people and not isolating myself [has worked so far].
2: Low disclosure	[P38] Just letting my thoughts roam [has not been] effective sometimes as it leads me to start off on one thing and by the end I will have thought through 20 different things and I forget how I even got there.	[P40] My friends [inspire me to take action]. I want to be the best version of myself for them to be a good friend.	[P9] Today, I am feeling okay... I don't feel as bad.
3: High disclosure	[P47] I tried therapy, but I didn't like it so I stopped. I refuse to take medication to try and help my mental health. So far, nothing I've done has worked for me so far.	[P11] I remember when he was young and loved spending time with me and talking to me, and now it just seems like he hates me. I wish we could be close like before!	[P3] When I'm alone, I often feel these dark thoughts. I'm very angry, sad, and confused about the whole situation

Table 3. Sample responses illustrating codes of information, thoughts, and feelings from our study population. Each statement will always receive three codes; therefore, statements shown in this table may have received different scores for their non-represented categories (e.g., [P46] was rated as (depth = 1) no disclosure of feelings and (depth = 3) high disclosure of information).

2.3.2 Readiness

To assess readiness for wellbeing change, readiness is assessed through measures of **stage** of readiness, **composite** readiness score, and **intentions** to address mental wellbeing. For stage and composite, we collected responses to the *Readiness-to-Change Questionnaire* (Heather and Hönekopp, 2008). This questionnaire is grounded in the Transtheoretical Model of Change (TTM) (Prochaska and Velicer, 1997), a structured and theoretical framework commonly used in health interventions and digital health (Zhang et al., 2024; Ferron and Massa, 2013) to conceptualize behavior change as a progression through distinct stages (Smith, 2020). Computational modeling of TTM has demonstrated its validity in classifying users into these stages (Attrey et al., 2026), and TTM-based digital interventions have shown efficacy in promoting behavioral change (Nghaimesh and Abd Ali, 2024). When combined with empathetic communication strategies in chatbots, TTM-based assessments can enhance responsiveness to users' psychological needs (Lin et al., 2023). We assess an individual's readiness to change *stage* across three stages: **Pre-Contemplation**, **Contemplation**, and **Action** (Rollnick et al., 1992). The stage measure indicates whether the person is not yet considering change (PC), thinking about making a change (C), or actively working toward change (A). The *composite* readiness measure is produced by the following equation: C + A - PC. Additionally, one single item was adapted from prior work to assess participant *intention* to address their mental wellbeing, Pre and Post (Yang et al., 2010; Gayet-Ageron et al., 2020). **H3**_{ReadinessOverall} investigates within-condition changes from Pre to Post for the measures of stage, composite readiness, and

intent. $H4_{\text{ReadinessComparison}}$ investigates between-condition changes from Pre to Post for the measures of stage, composite readiness, and intent.

2.3.3 Attitudes

The attitudinal metrics include a questionnaire on participant attitudes towards the present study's wellbeing chatbot intervention (**skepticism, confidence, technologization threat, anonymity**) and a single-item measure on **willingness** to engage with AI *chatbots* for mental wellbeing. For $H5_{\text{AttitudesIntervention}}$, attitudes were measured through an adaptation of the Attitude towards Psychological Online Interventions (APOI) Questionnaire (Schröder et al., 2015). The scale comprises four dimensions: *skepticism* and perception of risks, *confidence* in effectiveness, *technologization* threat, and *anonymity* benefits. For $H6_{\text{AttitudesChatbots}}$, attitudes of willingness to engage with AI chatbots was measured Pre and Post through a single item similar to the prior intention metric.

2.4 Procedure

We conducted a between-participant study with the described system with undergraduates at the University of Florida. Participants selected a time to participate in the study through one of the university's research recruitment platforms, which provides course credit to students as compensation for research studies. After giving informed consent, participants completed the pre-survey measures of readiness listed in Section 2.3. Participants were then randomized into one of **Perspective-Taking** or **Control**. Each participant completed the intervention steps of *perspective-taking* and *reflective conversation* as described in Section 2.1. Concluding the intervention steps and reflection in their self-perspective, participants completed the post-survey measures of readiness and attitudes described in Section 2.3, as well as demographics. Participants were debriefed on the study concerning how their anonymized data would be used and were subsequently granted course credit for their participation.

2.5 Participants

An a priori power analysis using G*Power was conducted for a mixed-design ANOVA with 2 groups (between-subjects factor) and 2 time points (within-subjects factor). Assuming an $\alpha = 0.05$, power $(1-\beta) = 0.95$, a medium effect size of $f = 0.20$, a correlation among repeated measures = 0.5, and sphericity met, we yield the minimum accepted sample size of $N \geq 84$. This study was approved by the University of Florida Institutional Review Board, and all participants provided written informed consent. To account for dropout and errors in completion, a total of 99 participants were recruited via the research recruitment platform and completed the entirety of the Procedure in Section 2.4. Three ($n = 3$) **Perspective-Taking** participants were excluded from analyses for introducing themselves as an alias deviant from their defined perspective's alias (see Section 2.1.2). The final analysis included 96 participants, with ($n = 48$) participants each in the **Control** and **Perspective-Taking** conditions.

Participants ranged in age from 18 to 41 years ($M = 21.8, SD = 2.98$). Gender identities included 67% male, 29% female, and 4% non-binary or unreported. Demographics were 52% White, 36% Asian or Pacific Islander, 4% Black or African American, 4% mixed, and 4% unreported, with 16% also identifying as Hispanic or Latino. In terms of education, all participants were students at the University of Florida, with 79% attending as undergraduate students and the remainder as graduate students. As the employed recruitment platform provides compensation for computer science-related courses, breakdowns of majors largely pertain to STEM: 71% computer science-related, 22% engineering, 5% mathematics or education-related, and 2% unlisted.

3 RESULTS

After collection and coding, data pre-processing was conducted using Python (3.12.2). Statistical analyses were primarily performed using IBM SPSS Statistics (Version 30). Descriptive statistics were computed to summarize the key variables across conditions. The significance level was set at $p < 0.05$, and assumptions for each test (e.g., normality tests via Shapiro-Wilk) were evaluated before conducting the analyses. Assumptions for independent samples t-tests and ANCOVAs, using Pre as a covariate, were tested and revealed violations of normality in measures, $p < 0.05$. Therefore, Mann-Whitney U tests and aligned rank transform (ART) ANOVAs, with Condition (Control and Perspective-Taking) and Time (Pre and Post) as factors, were conducted for each measure. Effect sizes were calculated and listed via rank-biserial correlation (r), partial eta squared η_p^2 , and Cramer's V for the corresponding non-parametric tests and Chi-square tests. Post hoc pairwise comparisons for ART ANOVAs were performed using ART-C with a Holm correction across the six pairwise post hoc comparisons to control the familywise error rate (Holm, 1979; Wobbrock et al., 2011). Due to a lack of support for ART ANOVAs in SPSS (see Section 3.2 for analysis), ART ANOVAs were analyzed in R (4.5.0) using ARTool (Wobbrock et al., 2011).

3.1 Disclosure

3.1.1 Quantity

Word Count. Mann-Whitney U test found a significant difference between conditions in word counts (averaged across the nine disclosure items), $U = 523, z = 2.45, p = 0.014, r = 0.33$. Word counts were significantly higher in **Perspective-Taking** (Mdn = 20.2, M = 21.9, SD = 11.7) compared to **Control** (Mdn = 10.4, M = 14.7, SD = 10.3) (see Figure 2).

Abstractness. Mann-Whitney U test revealed no significant difference in abstractness via LCM scores between conditions, $U = 380, z = 0.051, p = 0.960$. Descriptives for abstractness (1 = concrete, 5 = abstract) are included for reference: **Perspective-Taking** (Mdn = 3.22, M = 3.27, SD = 0.243) and **Control** (Mdn = 3.22, M = 3.24, SD = 0.179).

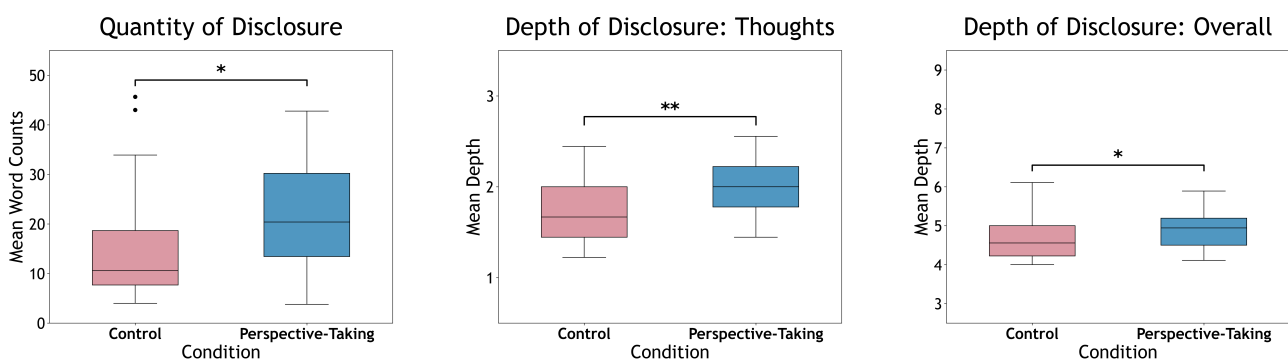


Figure 2. Box plots with medians for disclosure for **Control** and **Perspective-Taking** in terms of (Left) quantities, (Middle) depth of thoughts, and (Right) depth overall. Quantity, depth of thoughts, and depth overall refer to means for word counts, depth (intimacy) of thoughts, and depth (intimacy) of overall content, respectively, across the nine disclosure items, with significance illustrated (** < 0.01 , * < 0.05).

Depth of disclosure	Control						Perspective-Taking					
	Information		Thoughts		Feelings		Information		Thoughts		Feelings	
	n	%	n	%	n	%	n	%	n	%	n	%
1: No disclosure	110	42.3	135	51.9^a	246	94.6	108	46.4	95	40.8^a	226	97.0
2: Low disclosure	83	31.9^b	66	25.4	6	2.3	45	19.3^b	50	21.4	1	0.4
3: High disclosure	67	25.8^c	59	22.7^d	8	3.1	80	34.3^c	88	37.8^d	6	2.6
Total Counts	260	100	260	100	260	100	233	100	233	100	233	100

Table 4. Table illustrates the frequencies (n and %) of depth codes for each condition in terms of the categories of *Information*, *Thoughts*, and *Feelings*. Significantly different proportions are illustrated by matching **bold superscripts: a, b, c, d** (e.g., superscript **d** refers to a significantly greater proportion of high disclosure (depth = 3) of *Thoughts* in **Perspective-Taking** compared to **Control**).

3.1.2 Depth

In addition to Mann-Whitney U tests, Chi-square tests for homogeneity were employed to assess frequencies of depth 1, 2, or 3 for each category of information, thoughts, and feelings. Post hoc pairwise comparisons for Chi-squares were conducted using z-tests with a Bonferroni correction.

Information. Mann-Whitney U test revealed no significant difference in depth of information disclosure between conditions, $U = 433$, $z = 0.952$, $p = 0.341$.

Chi-square and post hoc tests indicated a significantly greater proportion of high disclosures (depth = 3) for **Perspective-Taking** compared to **Control**, $\chi^2(2) = 11.0$, $p < 0.01$, $V = 0.15$. In turn, a significantly lower proportion of low disclosures (depth = 2) was found for **Perspective-Taking** compared to **Control**. See Table 4 for coded depth frequencies and differences in Information.

Thoughts. Mann-Whitney U test found a significant difference in depth of thoughts disclosure between conditions, $U = 549$, $z = 2.92$, $p < 0.01$, $r = 0.39$. Depth of thoughts were significantly higher in **Perspective-Taking** (Mdn = 2.00, $M = 1.97$, $SD = 0.318$) compared to **Control** (Mdn = 1.67, $M = 1.71$, $SD = 0.320$) (see Figure 2).

Chi-square and post hoc tests indicated a significantly greater proportion of high disclosures (depth = 3) for **Perspective-Taking** compared to **Control**, $\chi^2(2) = 13.4$, $p < 0.001$, $V = 0.16$. In turn, a significantly lower proportion of no disclosures (depth = 1) was found for **Perspective-Taking** compared to **Control**. See Table 4 for coded depth frequencies and differences in Thoughts.

Feelings. Mann-Whitney U test revealed no significant difference in depth of feelings disclosure between conditions, $U = 333$, $z = -0.938$, $p = 0.348$.

There was a heavy skew in scores with no disclosure of feelings (depth = 1). Fisher's exact test was conducted due to an inadequate sample size for the chi-square test of homogeneity (Cochran, 1954). The distributions of feelings depth scores were not significantly different between conditions, $p = 0.224$.

Overall Depth. Mann-Whitney U test found a significant difference in overall disclosure depth between conditions, $U = 504$, $z = 2.15$, $p = 0.032$, $r = 0.29$. Overall depth was significantly higher in **Perspective-Taking** (Mdn = 4.94, $M = 4.90$, $SD = 0.519$) compared to **Control** (Mdn = 4.56, $M = 4.63$, $SD = 0.530$) (see Figure 2).

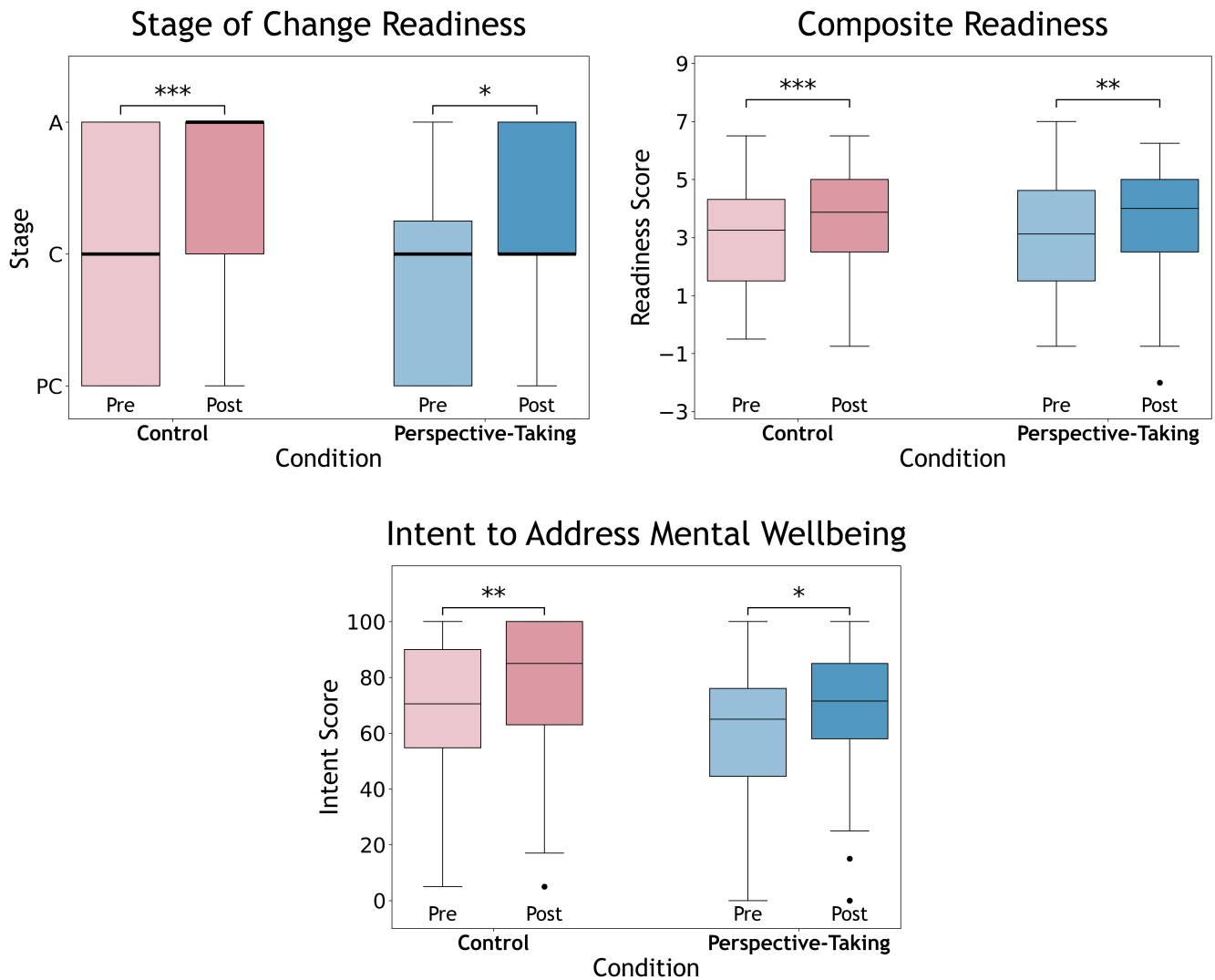


Figure 3. Box plots of Pre- and Post-readiness measures for **Perspective-Taking** and **Control** with medians: (Top-Left) stage of readiness for **Pre-Contemplation**, **Contemplation**, and **Action**, (Top-Right) composite readiness scores, and (Bottom) intent to address wellbeing. Significance within conditions from Pre to Post illustrated (*** < 0.001, ** < 0.01, * < 0.05). No significant effects of Condition × Time interaction.

3.2 Readiness

Stage. The ART ANOVA revealed a significant main effect of Time, $F(1, 94) = 17.5, p < 0.001, \eta_p^2 = 0.157$, indicating an overall improvement in stage of readiness from Pre to Post across conditions. Post hoc analyses revealed a significant improvement in stage from Pre to Post for **Control** only ($t(94) = 3.58, p < 0.01, r = 0.35$). No significant main effect of Condition or Condition × Time interaction was found, suggesting that the magnitudes of improvement over time did not differ significantly. Separate analysis on the deltas from Pre-stage to Post-stage also found no significant difference between conditions, $p > 0.05$ (see Figure 3).

Composite. Similar to *Stage*, the ART ANOVA revealed a significant main effect of Time, $F(1, 94) = 27.6, p < 0.001, \eta_p^2 = 0.227$, indicating an overall improvement in composite readiness scores from Pre to Post across conditions. Post hoc analyses revealed significant increases in composite readiness from

Pre to Post for each condition: **Perspective-Taking** ($t(94) = 3.45, p < 0.01, r = 0.34$) and **Control** ($t(94) = 4.06, p < 0.001, r = 0.39$). No significant main effect of Condition or Condition \times Time interaction was found, suggesting that the magnitudes of improvement over time did not differ significantly. Separate analysis on the deltas from Pre-composite readiness to Post-composite readiness also found no significant difference between conditions, $p > 0.05$ (see Figure 3).

Intention. ART ANOVA revealed a significant main effect of Condition, $F(1, 93) = 6.43, p = 0.013, \eta_p^2 = 0.065$. Post hoc comparisons for Condition revealed significantly higher overall intentions in the **Control** compared to **Perspective-Taking**, $p = 0.013, r = 0.25$. A significant main effect of Time was also observed, $F(1, 93) = 18.9, p < 0.001, \eta_p^2 = 0.169$, indicating an overall improvement in intention to address mental wellbeing from Pre to Post across conditions. Post hoc analyses revealed significant increases in intentions from Pre to Post for each condition: **Perspective-Taking** ($t(93) = 2.85, p = 0.022, r = 0.28$) and **Control** ($t(93) = 3.48, p < 0.01, r = 0.34$). No significant effect of Condition \times Time interaction was found, suggesting that the magnitudes of improvement over time did not differ significantly. Separate analysis on the deltas from Pre-intention to Post-intention also found no significant difference between conditions, $p > 0.05$ (see Figure 3).

3.3 Attitudes

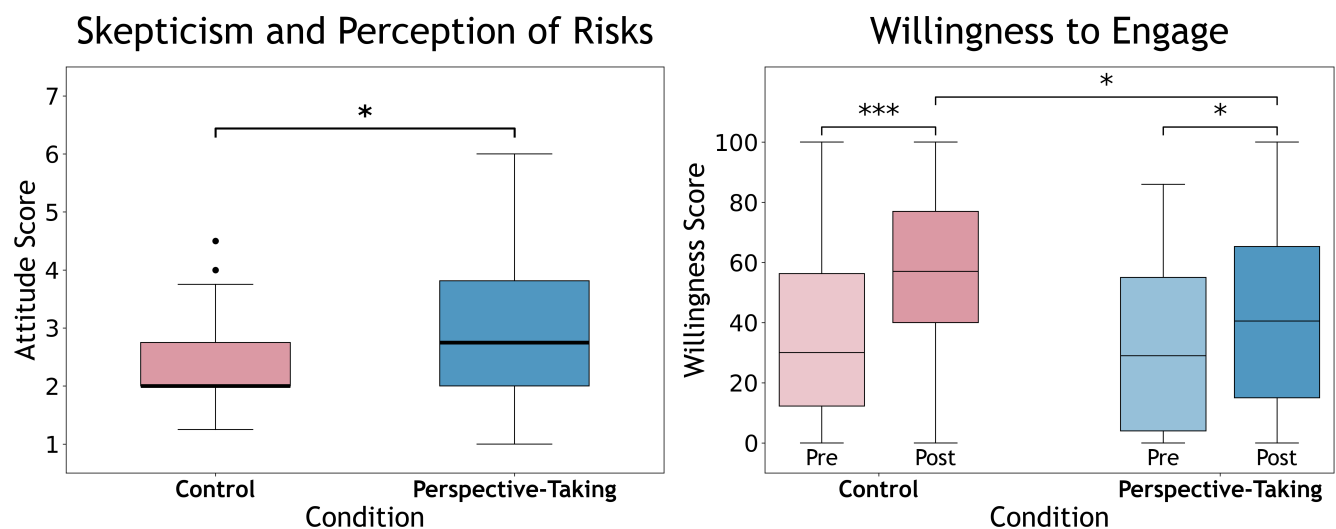


Figure 4. Box plots of attitudinal measures for **Perspective-Taking** and **Control** with medians: (Left) skepticism and perception of risks, and (Right) Pre- and Post-willingness to engage with AI chatbots for mental wellbeing. Significance in box plot between conditions and Pre to Post differences illustrated (*** < 0.001 , * < 0.05). There was a significant interaction effect of Condition \times Time in willingness to engage in favor of **Control**.

Skepticism and perception of risks. Mann-Whitney U test found a significant difference in skepticism and perception of risks between conditions, $U = 1558, z = 3.02, p < 0.01, r = 0.31$. Skepticism and perception of risks were significantly higher (worse) in **Perspective-Taking** (Mdn = 2.75, M = 3.09, SD = 1.33) compared to **Control** (Mdn = 2.00, M = 2.31, SD = 0.733) (see Figure 4).

Confidence in effectiveness. Mann-Whitney U test revealed no significant difference in confidence in effectiveness between conditions, $U = 961, z = -1.42, p = 0.155$.

Technologization threat. Mann-Whitney U test revealed no significant difference in technologization threat between conditions, $U = 1122, z = -.221, p = 0.825$.

Anonymity Benefits. Mann-Whitney U test revealed no significant difference in anonymity benefits between conditions, $U = 1161, z = 0.063, p = 0.950$.

Willingness to Engage. ART ANOVA revealed a significant main effect of Time, $F(1, 93) = 45.9, p < 0.001, \eta_p^2 = 0.331$, indicating an overall improvement in willingness to engage with AI chatbots for mental wellbeing from Pre to Post across conditions. Post hoc analyses revealed significant increases in willingness from Pre to Post for each condition: **Perspective-Taking** ($t(93) = 3.09, p = 0.010, r = 0.31$) and **Control** ($t(93) = 5.72, p < 0.001, r = 0.51$). A significant Condition \times Time interaction was also observed, $F(1, 93) = 5.23, p = 0.024, \eta_p^2 = 0.053$, suggesting that the effect of time differed between conditions. Additionally, the **Control** reported significantly greater willingness to engage with AI wellbeing chatbots at Post compared to **Perspective-Taking**, $p = .032, r = 0.22$ (see Figure 4).

4 DISCUSSION

Our results suggest that perspective-taking can significantly alter the ways in which users disclose to chatbots. In line with hypotheses $H1_{\text{DisclosureQuantity}}$ and $H2_{\text{DisclosureDepth}}$, **Perspective-Taking** participants disclosed significantly greater word quantities, depth of thoughts, and overall depth than **Control** participants. Perspective-taking also resulted in more frequent high-depth (level 3) disclosures in both information and thoughts compared to the control. Results also showed significant improvement in all readiness measures across both conditions, supporting our hypothesis $H3_{\text{ReadinessOverall}}$. Improvements in readiness did not support our deferred choice of **Perspective-Taking** for $H4_{\text{ReadinessComparison}}$, but surprisingly, we also found no interaction between Condition and Time, nor any difference in the rate of change (deltas) across readiness measures. The promising effects are tempered by attitudes: **Perspective-Taking** participants showed significantly greater skepticism and a less pronounced increase in willingness to engage with wellbeing chatbots than **Control** participants, contrary to $H5_{\text{AttitudesIntervention}}$ and $H6_{\text{AttitudesChatbot}}$. We interpret the findings observed on disclosure and the implications of this work for wellbeing chatbots accordingly.

4.1 Interpreting effects on disclosure

To contextualize the effects of perspective-taking on disclosure, we provide interpretations of the improved disclosure, consider the nature of the disclosures, and identify limits to our disclosure findings.

Perspective-taking significantly improved the quantity and depth of participants' disclosures (see Figure 2). Our findings echo prior work showing that perspective-taking can shift engagement behavior in applied contexts, now extended to chatbot-mediated disclosures (Williams et al., 2024; Ahn et al., 2013; Pahl and Bauer, 2013). Such literature suggests perspective-takers often align their behavior with their expectations of the other's imagined actions, which can override intrinsic behavioral constraints (Batson et al., 1997a, 2002; Batson, 2014; Kavanagh et al., 2020; Davis et al., 2004). In the present study, we suggest that the change in disclosure behavior stems from similar effects and indicates greater substance within these disclosures, rather than an abstract increase in verbosity. Despite prior claims that distal constructs and psychological distance promotes abstraction (Liberman et al., 2007; Liberman and Trope, 2008), we observed no such increase in abstract language among perspective-taking disclosures. The observed improvement was also not limited to quantity, as the depth of thoughts and overall disclosures were significantly greater when perspective-taking. This would indicate that **Perspective-Taking** participants illustrate a greater disclosure of personal and intimate information (Altman and Taylor, 1973; Carpenter and Greene, 2015). While quantity and depth can often relate, lower quantity disclosures can still result in higher depths (Barak and Gluck-Ofri, 2007), and their correlations are not necessarily positive (Tolstedt and Stokes, 1984). The observed findings of improved disclosure depths further suggest that perspective-taking fostered greater substantive content from users, rather than a simple inflation of abstract or verbose wording.

While disclosure quantity and depth seem to have meaningfully improved, it is worthwhile to discuss the nature of the disclosures produced in the **Perspective-Taking** condition. A natural question that arises is in the self-relevance of these disclosures, as they were uttered wholly from the perspective of the designated other. Even if the disclosures do not directly mirror personal information or thoughts, prior literature notes that individuals often project self-relevant traits onto imagined others (Liberman et al., 2007; Trope and Liberman, 2010). Although the present study did not directly measure overlap, the study design of self-designation for an other, imagined in the first person, aimed to promote overlap and afford successful perspective-taking (Davis et al., 2004; Myers et al., 2014). Our earlier discussion on a lack of abstractness differences would also align with such literature (Epley and Caruso, 2012; Trope and Liberman,

2010; Liberman and Trope, 2008). Even further, the results also demonstrate that **Perspective-Taking** participants' readiness improved, despite disclosing from an other's perspective. The readiness gains may reflect previously documented merging effects (Ames et al., 2008; Goldstein and Cialdini, 2007; Aron and Aron, 1986). Together, the study design and findings suggest plausibility that **Perspective-Taking** participants projected some properties of the self (albeit, to a likely lesser extent than **Control**) in their estimations of the other, resulting in the improved readiness outcomes. We ultimately still characterize our findings as an improvement in *disclosure*, rather than *self-disclosure*, recognizing the limitations in determining the extent to which the **Perspective-Taking** disclosures directly pertain to participants' selves.

The claims on disclosure are with respect to multiple dimensions: quantity via word counts and abstractions, and depth via information, thoughts, and overall. However, the disclosure of *feelings* remains an area for deeper investigation. Within the present study, few disclosures pertained to participant emotions or feelings, regardless of condition. Roughly 96% of the responses across both conditions were assigned (depth = 1) no disclosure of feelings. Since both conditions disclosed little emotional content, this may seemingly be explained by the fact that each of the nine disclosure items directly requested disclosure of information (e.g., experiences, background) or thoughts (e.g., opinions, goals, plans). While our reflective conversation did assess participants' feelings and emotions, such sentiments were primarily captured in the closed-ended items, which could not be included in the analysis of the disclosure items. As a result, the responses to the nine disclosure items almost exclusively pertained to participants' direct histories, experiences, thoughts, and opinions. However, the lack of emotional expression may also reflect a broader limitation of perspective-taking itself, which may enhance depth and thought but not necessarily encourage the disclosure of affective content. This possibility invites further investigation into whether perspective-taking facilitates cognitive but not emotional forms of disclosure.

4.2 AI Chatbots to promote mental wellbeing

In light of findings on readiness, we discuss how perspective-taking led to such effects, implications for wellbeing chatbot interactions, and design considerations based on user attitudes.

While both conditions experienced significantly improved readiness outcomes, our findings suggest that the degree of improvement between conditions was not significantly different (see Figure 3). Though hypotheses were deferred in favor of **Perspective-Taking**, it would also reason that speaking from one's own perspective (**Control**) should naturally afford disclosures more pertinent to the self, as well as more self-tailored expressions of emotion from the chatbot. We provide several possible explanations of how **Perspective-Taking** disclosures may have led to seemingly comparable user outcomes. First, the improved readiness observed in **Perspective-Taking** likely reflects previously discussed mechanisms of activated motivations to change and blurred boundaries of helping the self or an other (Batson, 2014; Cialdini et al., 1987, 1997; Davis et al., 1996; Pahl and Bauer, 2013). And while potential self-other overlap effects may also account for such improvements, another plausible possible factor is the therapeutic effect of written emotional experiences, which has been shown to enhance wellbeing (Pennebaker and Francis, 1996; Pennebaker, 1997). Written emotional disclosures may cover topics such as emotional experiences, future aspirations, or past successes, akin to topics in our reflective conversation. Prior work has found that writing forms of emotional disclosures can help facilitate relief for anxiety and depressive symptoms (Graf et al., 2008; Konig et al., 2014; King, 2001), with improvements in self-esteem (O'Connor et al., 2011) and even physical health symptoms (Radcliffe et al., 2007; Greenberg and Stone, 1992; Frisina et al., 2004). Interestingly, some studies have investigated the effects of written emotional disclosures from non-self perspectives. Greenberg et al. found that writing emotional disclosures from an unexperienced, imagined

perspective led to improvements in health symptoms and lower immediate reports of depression, fatigue, and avoidance (Greenberg et al., 1996). King et al. also found that writing from a distance perspective in the form of a hypothetical (ideal) self could elicit similar health effects in comparison to writing about self-experiences (King, 2001). Our findings echo prior work illustrating that writing emotional disclosures, regardless of perspective, may produce positive effects on the self. Finally, it is also worth briefly mentioning the potential role of observed enhancements in disclosures in **Perspective-Taking**. The findings suggest that **Perspective-Taking** participants wrote more quantity with greater depth than the **Control**, which may have further contributed to overall improvements and lack of significant differences between conditions. No formal analysis was able to be conducted on the relationship between disclosures and readiness, but such investigations could further elucidate how chatbots can promote self-outcomes.

Based on our findings and discussion, it would appear that direct self-disclosures may not be a strict requirement to promote user outcomes with chatbots. This may have numerous implications for pathways to promote interactions with emotional intelligence AI. In chatbot conversations addressing highly stigmatized topics (e.g., severe health issues, sexual health, or mental health), users may limit their disclosures due to shame or fear of judgment (Branley-Bell et al., 2023; Miles et al., 2021; Cui et al., 2024). If our findings hold, a distanced perspective may be able to be leveraged to overcome such stigmas and draw deeper disclosures for self-benefit. A similar domain that applies such techniques is within therapeutic role-plays, which has demonstrated that *imaginative* scenarios can be employed for self-understanding, improvement, and behavior (Corsini, 2017; Moreno, 1934, 1987). The present study also supports preliminary findings suggesting that such engagements may be suitable for human-computer simulations (Matthews et al., 2014). Another implication of our work is towards the ethical, safe usage of AI chatbots, especially with regard to user data privacy and security. Though chatbots have been shown to be a promising opportunity for health outcomes (Gardiner et al., 2017; Fitzpatrick et al., 2017; Abd-Alrazaq et al., 2020), concerns arise in the employment of *generative AI* for wellbeing. Generative AIs can run the risk of memorizing or reproducing data, which poses further considerations in digital health conversations that pertain to protected or sensitive health information (Feretzakis et al., 2024). Generative AI also faces broad technology risks associated with compromised data and leaks (Chen and Esmailzadeh, 2024). While cybersecurity safeguards and processing data in a de-identified state can provide a layer of security, even de-identified information can potentially be re-identified with real persons (Feretzakis et al., 2024). Perspective-taking may offer a potential mitigation strategy for these risks. By encouraging distanced disclosure, users may still benefit from reflective engagement without exposing identifiable or sensitive information. Such efforts align with recent ethical recommendations that emphasize a need to minimize data exposure in AI-mediated mental health contexts (Luxton and Hudlicka, 2022; Cabrera et al., 2023). Given skepticism findings suggesting that the present intervention was perceived as less relevant to perspective-takers, distanced perspectives may be able to help buffer against negative effects that arise from AI hallucinations, since inaccurate or misleading responses may not necessarily be interpreted as personally relevant or diagnostic (Blease and Rodman, 2024). In this way, our study contributes to ongoing discussions about how to design AI systems that are both effective and ethically responsible in sensitive domains.

The promising findings on disclosure and readiness should be interpreted in light of the more complex pattern observed in attitudes. Perspective-taking may be able to promote disclosure behaviors, but it may come at a cost of perceived personal applicability of the chatbot's support. Worth noting is that the skepticism measure was employed to gauge perceptions of the intervention's ability to provide *effective, personal* support, and the resulting attitudes would be consistent with expectations that **Perspective-Taking** would include less self-relevant disclosures and/or support compared to **Control**. The findings of **H5** and **H6** are also aligned with established literature on self-distancing and construal-level theory that

illustrate that distance reduces egocentric experiences with stimuli and leads to less self-relevant appraisals (Ayduk and Kross, 2010; Liberman et al., 2007). In other words, when individuals adopt another person's perspective, they may feel less directly connected to the experience and perceive it as less personally relevant or useful, even if it encourages thoughtful reflection. Given the role that attitudes may play in one's decision to engage with such digital interventions (Marangunić and Granić, 2015), the divergence in attitudes warrants additional considerations. The confounding effect raises practical concerns for the design of supportive AI systems that leverage psychological distance. A perceived lack of belief that the AI can support the user may lead to diminished future engagements (Li et al., 2024), even if the intervention effectively prompts deeper disclosure in the present. Users may also be resistant to advice from the AI chatbot, despite its potential effectiveness due to distancing effects and relevance (Park et al., 2024). If perspective-taking prompts deeper reflection but undermines one's attitudes towards using such systems, its standalone use may be insufficient. Perspective-taking and distancing theories could enhance wellbeing chatbot engagements, but may need to be complemented with strategies that restore personal resonance to foster congruently positive attitudes. Future work could investigate practices in therapeutic contexts where distancing is spontaneous rather than longitudinal (Ayduk and Kross, 2010) or patients switch between immersed and distanced perspectives (Barbosa et al., 2017).

4.3 Limitations and future work

There are limitations to this study that help contextualize its findings and identify avenues for future research. The focus of this work was on driving disclosure and wellbeing among university-attending populations with AI; as such, recruitments were made through a University of Florida student research platform. The resulting population consisted primarily of STEM students, which limits the generalizability of the results to broader student populations. STEM students may have more familiarity with AI compared to general populations and differing mental health concerns that may alter their usage of such systems (Lillywhite and Wolbring, 2024). Furthermore, the present work involved a large qualitative corpus of 1,479 codes for participant disclosures, but successful capture of the lost data may have allowed greater ability to analyze relationships between disclosures and outcomes. The lack of emotional disclosures from this structured, reflective conversation also limits the ability to understand how such methods can elicit affective engagement. Understanding these relationships could further clarify the role of disclosure as a mediating factor in AI chatbot engagements. A few study design limitations are mentioned for future work to help validate research with perspective-taking and emotionally intelligence AI chatbots. While the present perspective-taking intervention appeared effective, the absence of a placebo condition limits our ability to isolate the effects of intervention content from outstanding engagement effects. The results of this work are interpretable within single-session mental wellbeing conversations. As a result, the present outcomes are confined to immediate effects on disclosure and readiness to address mental wellbeing, but longitudinal effects on disclosure or actual changes in healthy behavior remain an unexplored area for future work. The intervention also relied on a carefully structured set of tasks to elicit the perspective within an asynchronous environment. Wellbeing interventions may struggle to incorporate such specific tasks within their contexts. Although prior research indicates that perspective-taking can happen more spontaneously (Furlanetto et al., 2013; Tversky and Hard, 2009), future integrations may need to consider alternative methodologies to integrate distancing practically, especially with respect to the attitudinal findings on skepticism. Several related areas for investigation outside of the current scope of work include: perceptions of the chatbot's expression of emotion, participants' attitudes toward their designated, or self–other overlap. Future work should continue to research the noted limitations, as well as opportunities to employ potential implications of the present work.

5 CONCLUSION

AI chatbots continue to act as a medium towards reducing barriers for mental support, especially when supplemented with emotional intelligence. However, the capabilities of such chatbots and their resulting outcomes can be limited without meaningful engagements and disclosures from users. A conversation with little-to-no depth may only elicit a surface-level understanding of the wellbeing concerns and needs of an individual. Similar to support from a counselor, a friend, or loved one, a chatbot's capability to appropriately assist and empathize may increase when provided with a greater quantity and depth of context. The findings of this study illustrate that perspective-taking may be able to enhance disclosure to AI chatbots for wellbeing. Specifically, our results suggest that perspective-taking led participants to share significantly greater disclosure in forms of word quantity and depth across multiple categories, with limited evidence of abstractions beyond what was seen in our control. Furthermore, the AI chatbot intervention seemingly helped all participants improve their readiness and intentions to address mental wellbeing, and perspective-taking did not seemingly diminish the gain in participants' improvements. In light of prior literature, our findings may suggest that meaningful disclosure to chatbots to improve mental wellbeing readiness may not necessarily require direct self-disclosures. In doing so, we describe implications for how perspective-taking and distancing theories may further enhance disclosure to chatbots in sensitive contexts or in pursuit of minimizing the disclosure of sensitive self-information. Future work should continue to investigate how greater disclosure can be evoked to meaningfully foster user outcomes with emotionally intelligent AI chatbots based on the limitations and emergent gaps identified in the present study.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving humans were approved by the University of Florida Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CY: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Software, Writing - original draft, Writing - review & editing. RG: Formal analysis, Methodology, Writing - review & editing. MV: Conceptualization, Methodology, Supervision, Writing - review & editing. RnV: Methodology, Validation, Writing - review & editing. RhV: Methodology, Validation, Writing - review & editing. AM: Formal analysis, Software, Writing - review & editing. DT: Formal analysis, Software, Writing - review & editing. XP: Formal analysis, Software, Writing - review & editing. BL: Project administration, Resources, Supervision, Validation, Writing - review & editing.

FUNDING

The author(s) declare no financial support was received for the research, authorship, and/or publication of this article.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

GENERATIVE AI STATEMENT

The author(s) declare that Generative AI was used in the creation of this manuscript. AI technology was used to improve the language, grammar, and readability of the manuscript.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

REFERENCES

- Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., and Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *Journal of medical Internet research* 22, e16021
- Aggarwal, A., Tam, C. C., Wu, D., Li, X., and Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. *Journal of medical Internet research* 25, e40789
- Ahn, S. J., Le, A. M. T., and Bailenson, J. (2013). The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychology* 16, 7–38
- Altman, I. and Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. (Holt, Rinehart & Winston)
- Ames, D. L., Jenkins, A. C., Banaji, M. R., and Mitchell, J. P. (2008). Taking another person's perspective increases self-referential neural processing. *Psychological science* 19, 642–644
- Aron, A. and Aron, E. N. (1986). *Love and the expansion of self: Understanding attraction and satisfaction*. (Hemisphere Publishing Corp/Harper & Row Publishers)
- Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., and Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and social psychology bulletin* 23, 363–377
- Attrey, L., Dua, S., Kaushik, R., Anand, S., and Agarwal, A. (2026). Modeling the transtheoretical model for health behavior stage analysis: Tool development and testing. In *Handbook of Deep Learning Models for Healthcare Data Processing* (CRC Press). 144–157
- Ayduk, Ö. and Kross, E. (2010). From a distance: implications of spontaneous self-distancing for adaptive self-reflection. *Journal of personality and social psychology* 98, 809
- Barak, A. and Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior* 10, 407–417
- Barbosa, E., Amendoeira, M., Ferreira, T., Teixeira, A. S., Pinto-Gouveia, J., and Salgado, J. (2017). Immersion and distancing across the therapeutic process: relationship to symptoms and emotional arousal. *Research in Psychotherapy: Psychopathology, Process, and Outcome* 20, 258
- Bargh, J. A., McKenna, K. Y., and Fitzsimons, G. M. (2002). Can you see the real me? activation and expression of the “true self” on the internet. *Journal of social issues* 58, 33–48
- Bartels, S. L., Taygar, A. S., Johnsson, S. I., Petersson, S., Flink, I., Boersma, K., et al. (2023). Using personas in the development of ehealth interventions for chronic pain: A scoping review and narrative synthesis. *Internet Interventions* 32, 100619
- Batson, C. D. (2014). *The altruism question: Toward a social-psychological answer* (Psychology Press)
- Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., and Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of personality and social psychology* 61, 413
- Batson, C. D., Chang, J., Orr, R., and Rowland, J. (2002). Empathy, attitudes, and action: Can feeling for a member of a stigmatized group motivate one to help the group? *Personality and Social Psychology Bulletin* 28, 1656–1666
- Batson, C. D., Early, S., and Salvarani, G. (1997a). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and social psychology bulletin* 23, 751–758
- Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., et al. (1997b). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of personality and social psychology* 72, 105

- Batson, C. D., Sager, K., Garst, E., Kang, M., Rubchinsky, K., and Dawson, K. (1997c). Is empathy-induced helping due to self–other merging? *Journal of personality and social psychology* 73, 495
- Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., and O’Donnell, J. (2010a). Response to a relational agent by hospital patients with depressive symptoms. *Interacting with computers* 22, 289–298
- Bickmore, T. W., Puskar, K., Schlenk, E. A., Pfeifer, L. M., and Sereika, S. M. (2010b). Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers* 22, 276–288
- Bland, D. (2012). Agile coaching tip—what is an empathy map. Available in <http://www.bigvisible.com/2012/06/what-is-an-empathy-map>
- Blatt, B., LeLacheur, S. F., Galinsky, A. D., Simmens, S. J., and Greenberg, L. (2010). Does perspective-taking increase patient satisfaction in medical encounters? *Academic Medicine* 85, 1445–1452
- Blease, C. and Rodman, A. (2024). Generative artificial intelligence in mental healthcare: an ethical evaluation. *Current Treatment Options in Psychiatry* 12, 5
- Boland, L., Campbell, D., Fazekas, M., Kitagawa, W., MacIver, L., Rzczkowska, K., et al. (2021). An experimental investigation of the effects of perspective-taking on emotional discomfort, cognitive fusion and self-compassion. *Journal of contextual behavioral science* 20, 27–34
- Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., et al. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. *Expert review of medical devices* 18, 37–49
- Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. (2022). The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin* 10, 1–47
- Branley-Bell, D., Brown, R., Coventry, L., and Sillence, E. (2023). Chatbots for embarrassing and stigmatizing conditions: could chatbots encourage users to seek medical advice? *Frontiers in Communication* 8, 1275127
- Brown, J. E. and Halpern, J. (2021). Ai chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. *SSM-Mental Health* 1, 100017
- Brown, P. and Levinson, S. C. (1987). *Politeness: Some universals in language usage*, vol. 4 (Cambridge university press)
- Bullingham, L. and Vasconcelos, A. C. (2013). ‘the presentation of self in the online world’: Goffman and the study of online identities. *Journal of information science* 39, 101–112
- Bylund, C. L. and Makoul, G. (2002). Empathic communication and gender in the physician–patient encounter. *Patient education and counseling* 48, 207–216
- Bylund, C. L. and Makoul, G. (2005). Examining empathy in medical encounters: an observational study using the empathic communication coding system. *Health communication* 18, 123–140
- Cabrera, J., Loyola, M. S., Magaña, I., and Rojas, R. (2023). Ethical dilemmas, mental health, artificial intelligence, and llm-based chatbots. In *International Work-Conference on Bioinformatics and Biomedical Engineering* (Springer), 313–326
- Carmichael, L., Poirier, S.-M., Coursaris, C. K., Léger, P.-M., and Sénécal, S. (2022). Users’ information disclosure behaviors during interactions with chatbots: the effect of information disclosure nudges. *Applied Sciences* 12, 12660
- Carpenter, A. and Greene, K. (2015). Social penetration theory. *The international encyclopedia of interpersonal communication*, 1–4
- Casu, M., Triscari, S., Battiato, S., Guarnera, L., and Caponnetto, P. (2024). Ai chatbots for mental health: A scoping review of effectiveness, feasibility, and applications. *Appl. Sci* 14, 5889

- Chen, Y. and Esmailzadeh, P. (2024). Generative ai in medical practice: in-depth exploration of privacy and security challenges. *Journal of Medical Internet Research* 26, e53008
- Chin, H., Song, H., Baek, G., Shin, M., Jung, C., Cha, M., et al. (2023). The potential of chatbots for emotional support and promoting mental well-being in different cultures: mixed methods study. *Journal of Medical Internet Research* 25, e51712
- Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., and Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of personality and social psychology* 73, 481
- Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fultz, J., and Beaman, A. L. (1987). Empathy-based helping: Is it selflessly or selfishly motivated? *Journal of personality and social psychology* 52, 749
- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10, 417–451
- Coelho, J., Pécune, F., Micoulaud-Franchi, J.-A., Bioulac, B., and Philip, P. (2025). Promoting mental health in the age of new digital tools: balancing challenges and opportunities of social media, chatbots, and wearables. *Frontiers in Digital Health* 7, 1560580
- Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., and D’Alfonso, S. (2023). To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital health* 9, 20552076231183542
- Corsini, R. (2017). *Role playing in psychotherapy* (Routledge)
- Crutzen, R., Bosma, H., Havas, J., and Feron, F. (2014). What can we learn from a failed trial: insight into non-participation in a chat-based intervention trial for adolescents with psychosocial problems. *BMC research notes* 7, 1–7
- Cui, Y., Lee, Y.-J., Jamieson, J., Yamashita, N., and Lee, Y.-C. (2024). Exploring effects of chatbot’s interpretation and self-disclosure on mental illness stigma. *Proceedings of the ACM on Human-Computer Interaction* 8, 1–33
- Davis, J. L. and Love, T. P. (2017). Self-in-self, mind-in-mind, heart-in-heart: The future of role-taking, perspective taking, and empathy. In *Advances in group processes* (Emerald Publishing Limited). 151–174
- Davis, M. H. (2018). *Empathy: A social psychological approach* (Routledge)
- Davis, M. H., Conklin, L., Smith, A., and Luce, C. (1996). Effect of perspective taking on the cognitive representation of persons: a merging of self and other. *Journal of personality and social psychology* 70, 713
- Davis, M. H., Soderlund, T., Cole, J., Gadol, E., Kute, M., Myers, M., et al. (2004). Cognitions associated with attempts to empathize: How do we imagine the perspective of another? *Personality and Social Psychology Bulletin* 30, 1625–1635
- De Cremer, D. (2004). The closer we are, the more we are alike: The effect of self-other merging on depersonalized self-perception. *Current Psychology* 22, 316–324
- Denecke, K., Abd-Alrazaq, A., and Househ, M. (2021). Artificial intelligence for chatbots in mental health: opportunities and challenges. *Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges*, 115–128
- Edalat, A., Hu, R., Patel, Z., Polydorou, N., Ryan, F., and Nicholls, D. (2025). Self-initiated humour protocol: A pilot study with an ai agent. *Frontiers in Digital Health* 7, 1530131
- Eisenberg, N., Spinrad, T., and Sadovsky, A. (2006). Empathy-related responding in children. In *Handbook of moral development* (Psychology Press). 535–568
- Ennis, E., O’Neill, S., Mulvenna, M., and Bond, R. (2023). Chatbots supporting mental health and wellbeing of children and young people; applications, acceptability and usability. In *European Conference on Mental Health*

- Epley, N. and Caruso, E. M. (2012). Perspective taking: Misstepping into others' shoes. In *Handbook of imagination and mental simulation* (Psychology Press). 295–309
- Feng, Y., Hang, Y., Wu, W., Song, X., Xiao, X., Dong, F., et al. (2025). Effectiveness of ai-driven conversational agents in improving mental health among young people: Systematic review and meta-analysis. *Journal of medical Internet research* 27, e69639
- Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., and Verykios, V. S. (2024). Privacy-preserving techniques in generative ai and large language models: A narrative review. *Information* 15, 697
- Ferreira, B., Silva, W., Oliveira, E., and Conte, T. (2015). Designing personas with empathy map. In *SEKE*. vol. 152
- Ferron, M. and Massa, P. (2013). Transtheoretical model for designing technologies supporting an active lifestyle. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*. 1–8
- Fitzpatrick, K. K., Darcy, A., and Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health* 4, e7785
- Fitzsimmons-Craft, E. E., Chan, W. W., Smith, A. C., Firebaugh, M.-L., Fowler, L. A., Topooco, N., et al. (2022). Effectiveness of a chatbot for eating disorders prevention: a randomized clinical trial. *International Journal of Eating Disorders* 55, 343–353
- Frisina, P. G., Borod, J. C., and Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of nervous and mental disease* 192, 629–634
- Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., and Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in human neuroscience* 7, 455
- Galinsky, A. D. and Ku, G. (2004). The effects of perspective-taking on prejudice: The moderating role of self-evaluation. *Personality and social psychology bulletin* 30, 594–604
- Galinsky, A. D. and Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology* 78, 708
- Galinsky, A. D., Wang, C. S., and Ku, G. (2008). Perspective-takers behave more stereotypically. *Journal of personality and social psychology* 95, 404
- Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., et al. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient education and counseling* 100, 1720–1729
- Gayet-Ageron, A., Rudaz, S., and Perneger, T. (2020). Study design factors influencing patients' willingness to participate in clinical research: a randomised vignette-based study. *BMC medical research methodology* 20, 1–8
- Gehlbach, H., Brinkworth, M. E., and Wang, M.-T. (2012). The social perspective taking process: What motivates individuals to take another's perspective? *Teachers College Record* 114, 1–29
- Goldstein, N. J. and Cialdini, R. B. (2007). The spyglass self: a model of vicarious self-perception. *Journal of personality and social psychology* 92, 402
- Gonzalez-Acosta, A. M., Vargas-Treviño, M., Batres-Mendoza, P., Guerra-Hernandez, E. I., Gutierrez-Gutierrez, J., Cano-Perez, J. L., et al. (2025). The first look: a biometric analysis of emotion recognition using key facial features. *Frontiers in Computer Science* 7, 1554320

- Graf, M. C., Gaudiano, B. A., and Geller, P. A. (2008). Written emotional disclosure: A controlled study of the benefits of expressive writing homework in outpatient psychotherapy. *Psychotherapy Research* 18, 389–399
- Greenberg, M. A. and Stone, A. A. (1992). Emotional disclosure about traumas and its relation to health: effects of previous disclosure and trauma severity. *Journal of personality and social psychology* 63, 75
- Greenberg, M. A., Wortman, C. B., and Stone, A. A. (1996). Emotional expression and physical health: Revising traumatic memories or fostering self-regulation? *Journal of personality and social psychology* 71, 588
- Grové, C. (2021). Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in psychiatry* 11, 606041
- Hampson, S. E., John, O. P., and Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implications. *Journal of Personality and Social Psychology* 51, 37
- Hardcastle, S. J., Fortier, M., Blake, N., and Hagger, M. S. (2017). Identifying content-based and relational techniques to change behaviour in motivational interviewing. *Health psychology review* 11, 1–16
- Heather, N. and Hönekopp, J. (2008). A revised edition of the readiness to change questionnaire [treatment version]. *Addiction Research & Theory* 16, 421–433
- Hettema, J., Steele, J., and Miller, W. R. (2005). Motivational interviewing. *Annu. Rev. Clin. Psychol.* 1, 91–111
- Hill, J., Ford, W. R., and Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in human behavior* 49, 245–250
- Hodges, S. D., Clark, B. A., and Myers, M. W. (2010). Better living through perspective taking. In *Positive psychology as social change* (Springer). 193–218
- Hogan, R. (1969). Development of an empathy scale. *Journal of consulting and clinical psychology* 33, 307
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* , 65–70
- Huo, B., Boyle, A., Marfo, N., Tangamornsuksan, W., Steen, J. P., McKechnie, T., et al. (2025). Large language models for chatbot health advice studies: a systematic review. *JAMA Network Open* 8, e2457879–e2457879
- Inkster, B., Sarda, S., Subramanian, V., et al. (2018). An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth* 6, e12106
- Johnson-Grey, K. M., Boghrati, R., Wakslak, C. J., and Dehghani, M. (2020). Measuring abstract mind-sets through syntax: Automating the linguistic category model. *Social Psychological and Personality Science* 11, 217–225
- Kavanagh, D., Barnes-Holmes, Y., and Barnes-Holmes, D. (2020). The study of perspective-taking: Contributions from mainstream psychology and behavior analysis. *The Psychological Record* 70, 581–604
- Khawaja, Z. and Bélisle-Pipon, J.-C. (2023). Your robot therapist is not your therapist: understanding the role of ai-powered mental health chatbots. *Frontiers in Digital Health* 5, 1278186
- King, L. A. (2001). The health benefits of writing about life goals. *Personality and social psychology bulletin* 27, 798–807

- Kohlberg, L. (1976). Moral stages and moralization: The cognitive-development approach. *Moral development and behavior: Theory research and social issues*, 31–53
- Konig, A., Eonta, A., Dyal, S. R., and Vrana, S. R. (2014). Enhancing the benefits of written emotional disclosure through response training. *Behavior therapy* 45, 344–357
- Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social psychology* 32, 1134
- Kross, E. and Ayduk, O. (2011). Making meaning out of negative experiences by self-distancing. *Current directions in psychological science* 20, 187–191
- Lawrence, H. R., Schneider, R. A., Rubin, S. B., Matarić, M. J., McDuff, D. J., and Bell, M. J. (2024). The opportunities and risks of large language models in mental health. *JMIR Mental Health* 11, e59479
- Ledel Solem, I. K., Varsi, C., Eide, H., Kristjansdottir, O. B., Børøsund, E., Schreurs, K. M., et al. (2020). A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of epio. *Journal of medical Internet research* 22, e15889
- Lee, Y.-C., Yamashita, N., and Huang, Y. (2020). Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. *Proceedings of the ACM on Human-Computer Interaction* 4, 1–27
- Li, L., Peng, W., and Rheu, M. M. (2024). Factors predicting intentions of adoption and continued use of artificial intelligence chatbots for mental health: examining the role of utaut model, stigma, privacy concerns, and artificial intelligence hesitancy. *Telemedicine and e-Health* 30, 722–730
- Lieberman, N. and Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of personality and social psychology* 75, 5
- Lieberman, N. and Trope, Y. (2008). The psychology of transcending the here and now. *Science* 322, 1201–1205
- Lieberman, N. and Trope, Y. (2014). Traversing psychological distance. *Trends in cognitive sciences* 18, 364–369
- Lieberman, N., Trope, Y., Stephan, E., et al. (2007). Psychological distance. *Social psychology: Handbook of basic principles* 2, 353–383
- Lillywhite, B. and Wolbring, G. (2024). Auditing the impact of artificial intelligence on the ability to have a good life: using well-being measures as a tool to investigate the views of undergraduate stem students. *Ai & Society* 39, 1427–1442
- Limpanopparat, S., Gibson, E., and Harris, A. (2024). User engagement, attitudes, and the effectiveness of chatbots as a mental health intervention: A systematic review. *Computers in Human Behavior: Artificial Humans*, 100081
- Lin, S., Lin, L., Hou, C., Chen, B., Li, J., and Ni, S. (2023). Empathy-based communication framework for chatbots: a mental health chatbot application and evaluation. In *Proceedings of the 11th International Conference on Human-Agent Interaction*. 264–272
- Lipson, S. K., Zhou, S., Abelson, S., Heinze, J., Jirsa, M., Morigney, J., et al. (2022). Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. *Journal of affective disorders* 306, 138–147
- Loveys, K., Hiko, C., Sagar, M., Zhang, X., and Broadbent, E. (2022). “i felt her company”: A qualitative study on factors affecting closeness and emotional support seeking with an embodied conversational agent. *International Journal of Human-Computer Studies* 160, 102771
- Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. (2014). It’s only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37, 94–100

- Ludwig, V. U., Berry, B., Cai, J. Y., Chen, N. M., Crone, D. L., and Platt, M. L. (2022). The impact of disclosing emotions on ratings of interpersonal closeness, warmth, competence, and leadership ability. *Frontiers in Psychology* 13, 989826
- Luxton, D. D. and Hudlicka, E. (2022). Intelligent virtual agents in behavioral and mental healthcare: Ethics and application considerations. In *Artificial intelligence in brain and mental health: philosophical, ethical & policy issues* (Springer). 41–55
- Marangunić, N. and Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society* 14, 81–95
- Matthews, M., Gay, G., and Doherty, G. (2014). Taking part: role-play in the design of therapeutic systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 643–652
- Miles, O., West, R., and Nadarzynski, T. (2021). Health chatbots acceptability moderated by perceived stigma and severity: a cross-sectional survey. *Digital health* 7, 20552076211063012
- Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations.
- Moreno, J. L. (1987). *The essential Moreno: Writings on psychodrama, group method, and spontaneity* (Springer Publishing Company)
- Myers, M. W., Laurent, S. M., and Hodges, S. D. (2014). Perspective taking instructions and self-other overlap: Different motives for helping. *Motivation and Emotion* 38, 224–234
- Nghaimesh, M. O. and Abd Ali, M. B. H. (2024). Efficacy of stage-matched intervention based on the transtheoretical model of behavior change in enhancing high school students' decisional balance of digital gaming behavior: A randomized controlled trial. *Central Asian Journal of Medical and Natural Science* 5, 221–229
- Nguyen, M., Bin, Y. S., and Campbell, A. (2012). Comparing online and offline self-disclosure: A systematic review. *Cyberpsychology, Behavior, and Social Networking* 15, 103–111
- O'Connor, D. B., Hurling, R., Hendrickx, H., Osborne, G., Hall, J., Walklet, E., et al. (2011). Effects of written emotional disclosure on implicit self-esteem and body image. *British Journal of Health Psychology* 16, 488–501
- Olano-Espinosa, E., Avila-Tomas, J. F., Minue-Lorenzo, C., Matilla-Pardo, B., Serrano, M. E. S., Martinez-Suberviola, F. J., et al. (2022). Effectiveness of a conversational chatbot (dejal@ bot) for the adult population to quit smoking: pragmatic, multicenter, controlled, randomized clinical trial in primary care. *JMIR mHealth and uHealth* 10, e34273
- Pahl, S. and Bauer, J. (2013). Overcoming the distance: Perspective taking with future humans improves environmental engagement. *Environment and Behavior* 45, 155–169
- Park, G., Chung, J., and Lee, S. (2024). Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention. *Current Psychology* 43, 4352–4363
- Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological science* 8, 162–166
- Pennebaker, J. W. and Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & emotion* 10, 601–626
- Piaget, J. (2013). *The moral judgment of the child* (Routledge)
- Pickard, M. D., Roster, C. A., and Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior* 65, 23–30
- Prochaska, J. O. and Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American journal of health promotion* 12, 38–48

- Rackoff, G. N., Zhang, Z. Z., and Newman, M. G. (2025). Chatbot-delivered mental health support: Attitudes and utilization in a sample of us college students. *Digital Health* 11, 20552076241313401
- Radcliffe, A. M., Lumley, M. A., Kendall, J., Stevenson, J. K., and Beltran, J. (2007). Written emotional disclosure: Testing whether social disclosure matters. *Journal of social and clinical psychology* 26, 362–384
- Richardson, D. R., Hammock, G. S., Smith, S. M., Gardner, W., and Signo, M. (1994). Empathy as a cognitive inhibitor of interpersonal aggression. *Aggressive behavior* 20, 275–289
- Rollnick, S., Heather, N., Gold, R., and Hall, W. (1992). Development of a short ‘readiness to change’ questionnaire for use in brief, opportunistic interventions among excessive drinkers. *British journal of addiction* 87, 743–754
- Rollnick, S. and Miller, W. R. (1995). What is motivational interviewing? *Behavioural and cognitive Psychotherapy* 23, 325–334
- Rupp, L. H., Kumar, A., Sadeghi, M., Schindler-Gmelch, L., Keinert, M., Eskofier, B. M., et al. (2025). Stress can be detected during emotion-evoking smartphone use: a pilot study using machine learning. *Frontiers in Digital Health* 7, 1578917
- Saadati, S. A., Saadati, S. M., et al. (2023). The role of chatbots in mental health interventions: user experiences. *AI Tech Behav Soc Sci* 1, 19–25
- Schröder, J., Sautier, L., Kriston, L., Berger, T., Meyer, B., Späth, C., et al. (2015). Development of a questionnaire measuring attitudes towards psychological online interventions—the apoi. *Journal of affective disorders* 187, 136–141
- Schroeder, J., Wilkes, C., Rowan, K., Toledo, A., Paradiso, A., Czerwinski, M., et al. (2018). Pocket skills: A conversational mobile web app to support dialectical behavioral therapy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–15
- Seih, Y.-T., Beier, S., and Pennebaker, J. W. (2017). Development and examination of the linguistic category model in a computerized text analysis method. *Journal of Language and Social Psychology* 36, 343–355
- Seinfeld, S., Arroyo-Palacios, J., Iruetagoiena, G., Hortensius, R., Zapata, L. E., Borland, D., et al. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Scientific reports* 8, 2692
- Semin, G. R. and Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of personality and Social Psychology* 54, 558
- Semin, G. R. and Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European review of social psychology* 2, 1–30
- Semin, G. R., Görts, C. A., Nandram, S., and Semin-Goossens, A. (2002). Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition & Emotion* 16, 11–28
- Shaffer, V. A., Bohanek, J., Focella, E. S., Horstman, H., and Saffran, L. (2019). Encouraging perspective taking: Using narrative writing to induce empathy for others engaging in negative health behaviors. *PloS one* 14, e0224046
- Shah, J., DePietro, B., D’Adamo, L., Firebaugh, M.-L., Laing, O., Fowler, L. A., et al. (2022). Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. *International Journal of Eating Disorders* 55, 1229–1244
- Shechtman, Z. and Tanus, H. (2006). Counseling groups for arab adolescents in an intergroup conflict in israel: Report of an outcome study. *Peace and Conflict* 12, 119–137
- Smith, M. (2020). The transtheoretical model, stages of change and motivational interviewing. In *A Comprehensive Guide to Addiction Theory and Counseling Techniques* (Routledge). 148–160

- Söderlund, L. L., Madson, M. B., Rubak, S., and Nilsen, P. (2011). A systematic review of motivational interviewing training for general health care practitioners. *Patient education and counseling* 84, 16–26
- Taylor, S. E. and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin* 103, 193
- Tolstedt, B. E. and Stokes, J. P. (1984). Self-disclosure, intimacy, and the depenetration process. *Journal of Personality and Social Psychology* 46, 84
- Trope, Y. and Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review* 117, 440
- Tversky, B. and Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition* 110, 124–129
- Uhl-Haedicke, I., Klackl, J., Muehlberger, C., and Jonas, E. (2019). Turning restriction into change: imagine-self perspective taking fosters advocacy of a mandatory proenvironmental initiative. *Frontiers in Psychology* 10, 2657
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., and Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry* 64, 456–464
- Valderrama, C. E. and Sheoran, A. (2025). Identifying relevant eeg channels for subject-independent emotion recognition using attention network layers. *Frontiers in Psychiatry* 16, 1494369
- Vallacher, R. R. and Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology* 57, 660
- Van Loon, A., Bailenson, J., Zaki, J., Bostick, J., and Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PloS one* 13, e0202442
- Vescio, T. K., Sechrist, G. B., and Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European journal of social psychology* 33, 455–472
- Wang, P. S., Berglund, P. A., Olfson, M., and Kessler, R. C. (2004). Delays in initial treatment contact after first onset of a mental disorder. *Health services research* 39, 393–416
- Weyant, J. M. (2007). Perspective taking as a means of reducing negative stereotyping of individuals who speak english as a second language 1. *Journal of Applied Social Psychology* 37, 703–716
- Williams, C., Rauwolf, P., Boulter, M., and Parkinson, J. A. (2024). Closing the gap: how psychological distance influences willingness to engage in risky covid behavior. *Behavioral Sciences* 14, 449
- Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146
- Yalçın, Ö. N. (2020). Empathy framework for embodied conversational agents. *Cognitive Systems Research* 59, 123–132
- Yang, Z. J., McComas, K., Gay, G., Leonard, J. P., Dannenberg, A. J., and Dillon, H. (2010). From information processing to behavioral intentions: Exploring cancer patients' motivations for clinical trial enrollment. *Patient Education and Counseling* 79, 231–238
- Yasukawa, S., Tanaka, T., Yamane, K., Kano, R., Sakata, M., Noma, H., et al. (2024). A chatbot to improve adherence to internet-based cognitive-behavioural therapy among workers with subthreshold depression: a randomised controlled trial. *BMJ Ment Health* 27
- You, C., Ghosh, R., Maxim, A., Stuart, J., Cooks, E., and Lok, B. (2022). How does a virtual human earn your trust? guidelines to improve willingness to self-disclose to intelligent virtual agents. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. 1–8

- Zhai, C. and Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: the case of second language acquisition. *Heliyon* 8
- Zhang, B., Kalampakorn, S., Powwattana, A., Sillabutra, J., and Liu, G. (2024). A transtheoretical model-based online intervention to improve medication adherence for chinese adults newly diagnosed with type 2 diabetes: a mixed-method study. *Journal of Primary Care & Community Health* 15, 21501319241263657
- Zhang, Z. and Wang, J. (2024). Can ai replace psychotherapists? exploring the future of mental health care. *Frontiers in Psychiatry* 15, 1444382