

LARGE LANGUAGE MODEL (LLM)-BASED HEALTH COACHES IN VIRTUAL REALITY  
(VR): EFFECTS OF AI AGENTS' NONVERBAL BEHAVIOR ON RAPPORT AND HEALTH  
OUTCOMES

By

Sue Lim

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Communication – Doctor of Philosophy

2025

## ABSTRACT

We used rapport theory as the framework to examine the effects of LLM-based embodied conversational agent (ECA)'s nonverbal behaviors in the context of health coaching. To conduct this study, we built two types of LLM-based health coaches in virtual reality using the Unreal platform. The *little* rapport-building health coach displayed minimal nonverbal behavior during conversations (e.g., no direct eye contact, no upper body movement) while the *more* rapport-building health coach displayed various rapport-building behaviors (e.g., smiling while listening, upper body movement while responding). Participants were randomly assigned to one of the two types of health coaches and completed six coaching sessions with the same health coach in immersive virtual reality (VR). Findings showed that those who interacted with the *more* rapport-building health coach expressed greater attentiveness across all six sessions (measured via ratio of gaze on the health coach). Also, we found attentiveness and subjective rapport during the initial interactions (sessions 1 and 2) as the most promising predictors of human clients' overall satisfaction with the intervention at the end of all six sessions. Finally, the results indicated that the more people interacted with the health coach, the more they felt they benefited from the sessions, with effects even more pronounced for those who interacted with the *more* rapport-building ECA. These findings have significant implications for communication research and practice.

## TABLE OF CONTENTS

SECTION 1: INTRODUCTION.....	1
SECTION 2: BACKGROUND & CURRENT STUDY.....	6
SECTION 3: METHODS.....	11
SECTION 4: RESULTS.....	24
SECTION 5: DISCUSSION.....	33
SECTION 6: CONCLUSION.....	40
BIBLIOGRAPHY.....	41
APPENDIX A: PROMPTS FOR DEVELOPMENT.....	50
APPENDIX B: SELF-REPORT SURVEY MEASURES.....	54
APPENDIX C: ALL RESULTS FROM MIXED EFFECTS REGRESSION (RQ) .....	56

## SECTION 1: INTRODUCTION

*“The construct of rapport is arguably one of the central, if not the central, construct necessary to understanding successful helping relationships and to explaining the development of personal relationships” (Cappella, 1990, pg. 303)*

Rapport is the *key* to successful human social interactions. Rapport refers to a harmonious relational dynamic that fosters open dialogue, a cooperative atmosphere, and a sense of mutual social connectedness, respect, and trust among its members (Bernieri, 1988; Gratch & Lucas, 2021; Tickle-Degnen & Rosenthal, 1990; Xie & Derakhshan, 2021). Phrased differently, rapport represents the extent of interpersonal bond that exists among interacting agents. The process of building rapport is communicative in nature, occurring through the exchange of verbal and nonverbal cues during social interactions. Furthermore, rapport is strongly linked with interaction outcomes including student performance (teacher-student interaction; Bernieri, 1998; Frisby & Martin, 2010; Estep & Roberts, 2015), cooperation (e.g., investigative interviewing; Abbe & Brandon, 2013), positive attitudes and intentions to engage in a behavior (e.g., service provider-customer interaction; Fatima, 2023), and patient health and adherence to treatment (provider-patient interaction; Harrigan et al., 1985; Joe et al., 2001; Leach, 2005).

In recent years, artificial intelligence (AI)-powered conversational agents (CAs) have entered many domains of human life, assuming relational roles such as coaches, companions, and customer service agents. CAs refer to machines that mimic human communication capabilities. Compared to earlier CAs that produced verbal responses based on users' selection of pre-determined choices or provision of simple and straightforward phrases, large language model (LLM)-based CAs like OpenAI's ChatGPT can process large amounts of user input, detect patterns, and generate responses that better resemble natural human dialogue. LLM-based CAs

can also exhibit empathy (Kang & Hong, 2024; Sorin et al., 2024; Vowels et al., 2024), attentiveness (Jo et al., 2024), and other types of rapport-building behaviors while delivering personalized information. As a result, LLM-based CAs are increasingly developed and deployed to support people's judgment, decision-making, and well-being (e.g., Alanezi, 2024; Choi et al., 2024; Jörke et al., 2024; Lim et al., 2024b; Nie et al., 2024; Wu et al., 2024).

Due to these advancements, human-AI communication is a growing subarea of communication research. Existing communication literature on human-AI interaction focuses heavily on verbal communication (i.e., the exchange of language-based cues through speech or text). However, *non-verbal* aspects of social interactions provide highly social information (Burgoon & Saine, 1978; Burgoon et al., 1984; Patterson, 1982) that has significant implications on rapport (Rapport Theory; Tickle-Degnen & Rosenthal, 1990) and relationship formation in general. During dyadic interactions, for example, the partners' body language can convey attentiveness (e.g., the body angled toward each other, the posture leaned forward) and express positivity (e.g., via head nods or facial smiles). Furthermore, coordinated behaviors such as biobehavioral synchrony are a window into relational dynamics like rapport and attachment (Feldman, 2017; Tickle-Degnen & Rosenthal, 1990). Studies have found that these behavioral dynamics are closely linked with higher perceived relationship quality and health outcomes of provider-patient interaction (Goldstein et al., 2020; Ramseyer & Tschacher, 2011). Thus, to fully understand human-AI communication processes and relational dynamics, we need to examine the effects of LLM-based CAs' nonverbal rapport-building behaviors as they engage in natural, turn-taking dialogue with human clients.

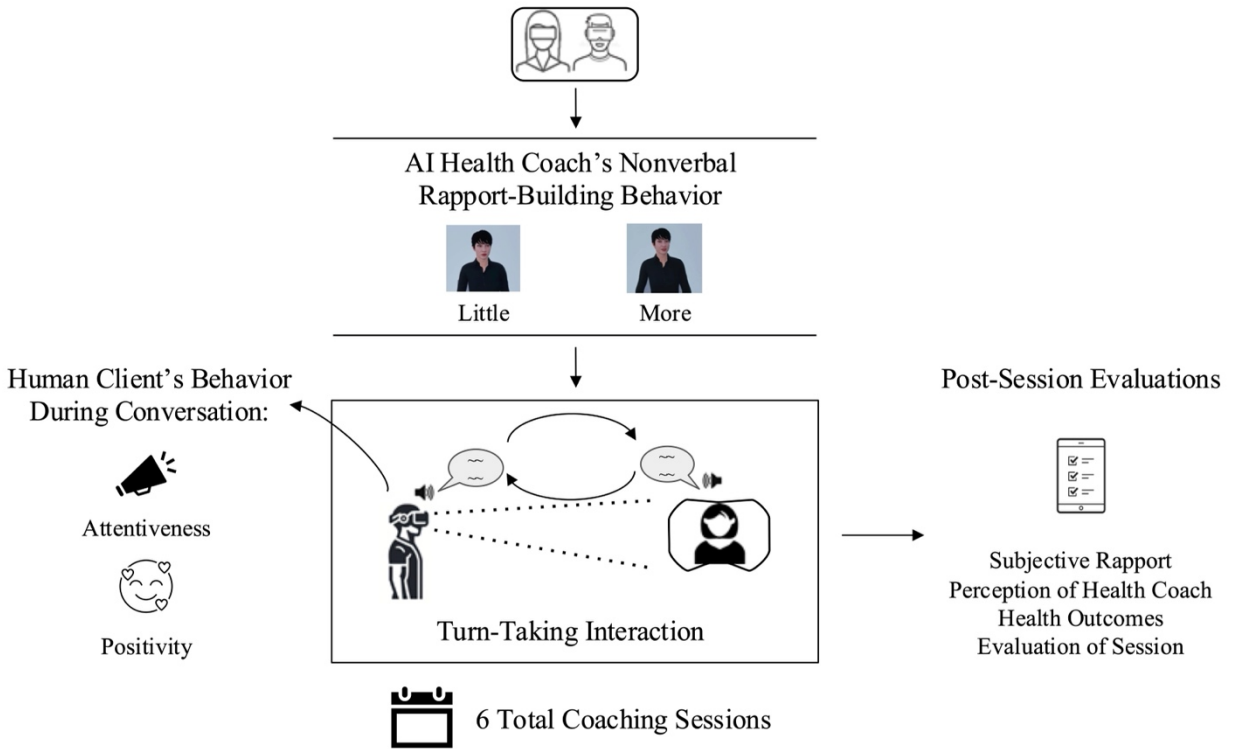
Methodologically, immersive virtual reality (VR) serves as an effective stimulation tool to study the mechanisms underlying human-AI interaction. Immersive VR encompasses three

key features. First, effective VR technologies like the head mounted displays (HMD) offer a high degree of immersion by absorbing the users' perceptual system and authentically stimulate their senses so they feel they are physically in the virtual environment (Bente et al., 2023; Biocca & Delaney, 1995). In addition, immersive VR allow for experimental control while maintaining a higher level of ecological validity compared to traditional lab studies (Loomis et al., 1999). Finally, immersive VR can be integrated with physiological and behavioral measures like eye-tracking to gather precise information about people's interaction with the environment and stimuli in real-time. Researchers have already implemented immersive VR to examine a wide range of human social behaviors, including bystander responses to violence (Slater et al., 2013) and proxemics during human-agent interaction (Bailenson et al., 2003; Iachini et al., 2014).

Our previous work demonstrated the feasibility of studying people's interactions with AI-powered health coaches within immersive VR platforms (Lim et al., 2024a; 2024b). Specifically, we leveraged the OpenAI's GPT4 LLM to develop embodied conversational agents (ECAs) that engaged in real-time, get-to-know-you and physical health-focused dialogue with basic eye contact in immersive VR (Lim\_AI1). The results showed that interacting with the LLM-based ECAs in VR fostered people's sense of immersion and presence (i.e., feeling of being in the same physical location and having emotional connections with the ECAs; Bente et al., 2023). We also found, through eye tracking, that participants tended to pay more attention to the AI health coach of the opposite gender during the health-focused conversations. Furthermore, with proper instructions, the LLM-based dialogue system exhibited empathic responses (e.g., "I understand what you mean") and other verbal rapport-building behaviors. Overall, the study highlighted immersive VR as an effective and flexible platform to examine human-AI interaction in a wide range of contexts.

Expanding on Lim et al. (2024b), this study examines how the AI health coach's nonverbal behaviors influence rapport and the outcomes of human-AI interaction overtime. Specifically, participants completed six health coaching sessions with LLM-based ECAs that exhibited little vs. more nonverbal rapport-building behaviors (see Figure 1). We first analyzed the effects of the AI health coaches' nonverbal behavior on the human clients' attentiveness to and positivity toward the ECA during initial conversations using rapport theory (Tickle-Degnen & Rosenthal, 1990) as the framework, as well as on subjective rapport. Next, we examined whether the human clients' expressions of rapport predicted interaction outcomes. Finally, we explored how the relational dynamic and the interaction outcomes developed over the multiple sessions. Our study begins to unpack the complex mechanisms underlying successful human-AI interaction and informs how to build effective LLM-based ECAs for health support.

This paper is structured as follows. First, section 2 summarizes relevant literature and introduces our hypotheses and research questions. Section 3 details the steps for LLM-based health coach development, experimental procedures, and statistical analyses. In addition, we outline the results of the study in Section 4 and discuss their implications in Section 5.



**Figure 1.** Conceptual Illustration of the Study. *The study examined the effect of the LLM-based Health Coaches' nonverbal rapport-building behavior on human clients' behavioral expressions of rapport, subjective rapport, and interaction outcomes.*

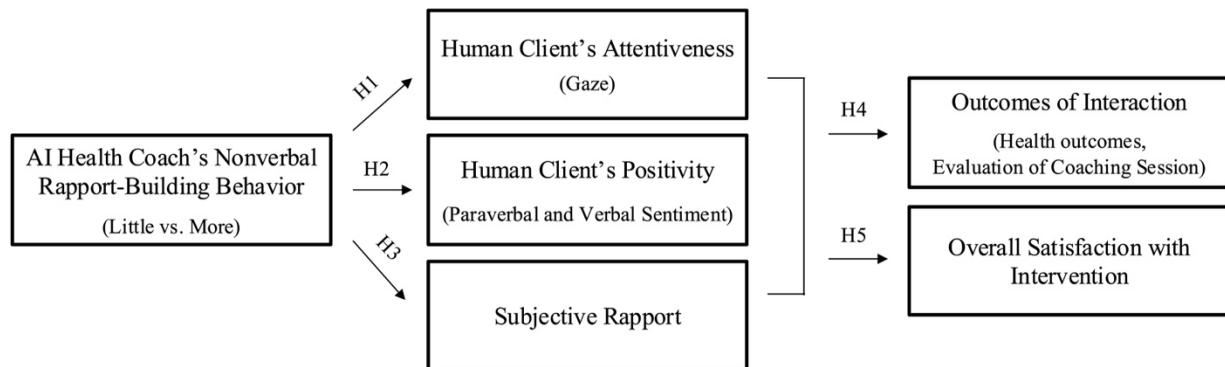


## SECTION 2: BACKGROUND & CURRENT STUDY

Rapport has been widely studied in health-related support contexts such as therapy, counselling, coaching, and healthcare provider-patient communication (Tickle-Degnen & Rosenthal, 1990). Existing literature found that rapport is expressed in varying ways. First, individuals' behaviors during social interactions can depict the extent of their rapport. Tickle-Degnen and Rosenthal's (1990) rapport theory posits that during dyadic interactions, interactants express rapport through three nonverbal categories: mutual attentiveness, positivity, and coordination. Mutual attentiveness refers to how much people engage with one another during the conversation. Positivity broadly captures mutual experience of positive affect such as friendliness and caring. Finally, dyads exhibit rapport through behavioral coordination - like partners dancing (Gratch & Lucas, 2021). Rapport theory suggests that the relative importance of the three categories differ based on the stages of the relationship: Positivity most important during the early stages of rapport development, coordination most important during relationship maintenance stage, and attentiveness important throughout the stages.

While dyadic human-to-human interaction features mutual engagement, positivity, and coordination, human-AI communication differs because AI – at least at this point in time – does not have the full capacity to exhibit natural social behavior like humans. Some researchers have worked on building adaptive ECAs that use machine learning algorithms to mimic the human's behavior during the interaction (e.g., Woo et al., 2024), but those systems show limitations. Still, existing literature also shows that ECAs *conveying* rapport-building nonverbal cues such as smiling can elicit similar behavior in human interactants and lead to rapport (Cassell & Thorisson, 1999; Krämer et al., 2013). This is because humans are innately social, and they naturally adjust their behavior in response to the other, even when interacting with strangers

(Chartrand & Bargh, 1999; Feldman, 2017). Thus, this study focuses on how the nonverbal rapport-building behaviors of the LLM-based ECAs elicit human users' attentiveness and positivity during the interactions<sup>1</sup> (see Figure 2 for the summary of our hypotheses and research questions).



**Figure 2.** Overview of the Current Study Design. *The hypotheses focused on the initial interactions with the LLM-based health coach.*

### Effect of Nonverbal Behavior and Rapport

One of the major behavioral markers of attentiveness is gaze. Humans use gaze to communicate their level of interest and engagement when interacting with others. In fact, Kleinke's (1986) survey of literature found that participants, especially those from Western cultures, use gaze and head orientation toward the other during social interactions as cues to make judgments about their attentiveness. Other studies demonstrated that patterns of gaze, such as making and breaking eye contact, during social interactions signaled overt attention, interpersonal closeness, and other relational cues (Guo et al., 2023; Hessels et al., 2017; Ho et al., 2015; Schmälzle et al., 2024; Wohltjen & Wheatley, 2021; 2024). As a result, human-

<sup>1</sup> Since the interactions are occurring in the beginning stages of human-AI rapport-building, this study does not examine coordination.

computer interaction researchers have also used gaze to examine people's attention to conversational agents (Amorese et al., 2022; Nakano & Ishii, 2010; Rehm & André, 2005; Robb et al., 2023). Building on these studies, we make the following prediction:

*H1: The nonverbal rapport-building behavior of the LLM-based embodied conversational agent will increase human clients' attentiveness (measured via gaze) to the conversation.*

For positivity, the human clients' paraverbal and verbal responses to the LLM-based ECA can serve as signals of positive affect during social interactions. Like gaze communicates interest and engagement, humans also express affect and other relational factors through the voice (e.g., tone, pitch, speech rate). Studies found that the combinations of vocal cues can predict affective states (Gobl & Chasaide, 2003; Johnson et al., 1986; Scherer et al., 1973) as well as attributions of speaker characteristics (Apple et al., 1979; Scherer et al., 1973; Schroeder & Epley, 2015). As a result, human-computer interaction literature has used vocalics to examine people's social responses to agents (Cerekovic et al., 2016; Elkins et al., 2012; Elkins & Derrick, 2013; Nunamaker et al., 2011). In addition, linguistic markers in verbal behaviors also indicate sentiment (Clavel & Callejas, 2015; Lepper & Mergenthaler, 2007; Lubis et al., 2019; Mergenthaler, 2008; Ranjartabar et al., 2019; Santos et al., 2020). Thus, we predict the following between LLM-based ECAs' nonverbal rapport-building behavior and human clients' behavior:

*H2a: The nonverbal rapport-building behavior of the LLM-based embodied conversational agent will increase human clients' positivity (measured via sentiment analysis of the paraverbal responses) during the conversation.*

*H2b: The nonverbal rapport-building behavior of the LLM-based embodied*

*conversational agent will increase human clients' positivity (measured via sentiment analysis of the verbal responses) during the conversation.*

Beyond the behavioral markers, human clients' perceptions of the interaction and the connection with the LLM-based ECAs post interaction indicates the magnitude of rapport they feel. Existing work in human-computer interaction have examined how the ECAs' nonverbal behaviors influence people's experience of rapport (e.g., Gratch et al., 2006; Huang et al., 2011). Some studies found that the ECAs' nonverbal rapport-building behaviors enhanced people's feelings of rapport with the agent (Karacora et al., 2012; Wang & Gratch, 2009; Wang & Ruiz, 2021). Thus, we predict the following:

*H3: The nonverbal rapport-building behavior of the LLM-based embodied conversational agent will increase human clients' perception of rapport.*

### **Relationship Between Rapport and Interaction Outcomes**

As discussed in the introduction, rapport is an essential ingredient in healthcare and coaching settings. For example, clinical psychology and psychiatry studies showed that rapport (and other similar concepts like therapeutic or working alliance) boosted treatment effectiveness by predicting treatment adherence and symptom reduction (Cloitre et al., 2004; Frank & Gunderson, 1990; Joe et al., 2001; Krupnick et al., 2006). Similarly, Graßmann et al.'s (2020) meta-analysis of the coaching literature found that the coach-coachee working alliance – another word for rapport – significantly related to outcomes including satisfaction, perceived coaching effectiveness, self-efficacy, as well as knowledge acquisition. Other studies have also demonstrated the link between rapport with nonhuman agents and human performance (Karacora et al., 2012). We build upon these studies and predict the following:

*H4: Indicators of rapport will be positively associated with interaction outcomes.*

*H5: Indicators of rapport will be positively associated with satisfaction with the coaching intervention.*

### **Longitudinal Effects of LLM-Based Health Coaches' Nonverbal Behavior**

In addition, the hypotheses above predict the behaviors and outcomes of the initial interactions (first two sessions). However, Tickle-Degnen and Rosenthal (1990) conceptualize rapport as a dynamic process that occurs over time. Thus, we examine how the LLM-based Health Coaches' nonverbal rapport-building behaviors influence human clients' expressions of rapport (same variables from *H1-3*) and interaction outcomes over the six coaching sessions.

*RQ: Does LLM-based health coaches' nonverbal rapport-building behaviors influence human clients' expressions of rapport and interaction outcomes over time?*

## SECTION 3: METHODS

### Participants

The study was approved by the local institutional board prior to data collection. The sample for the main study comprised 30 participants ( $M_{age} = 30.27$ ,  $SD_{age} = 11.64$ , 37% self-identified White or European American, 60% self-identified Female)<sup>2</sup>. We recruited the participants through the local institution's community participant pool. The inclusion criteria included being over the age of 18 and being able to speak and understand English. For compensation, participants received \$40 in cash for completing all six sessions.

### Developing Lim\_AI2: LLM-Based ECAs with Enhanced Rapport Building Capabilities

The development stage of the LLM-based ECAs included building the prototype and conducting a pilot study with a smaller sample. To develop the prototype, we integrated multiple AI systems with the Unreal Game Engine through Convai, a conversational AI platform (Convai, 2025). For the dialogue system that verbally builds rapport with people, we created the prompt based on motivational interviewing (MI) principles. MI conceptualizes motivation as a process rather than a trait (Miller, 1983) and uses a client-centered approach to help people want to change and commit to goals toward the change (Hettema et al., 2005). It is a widely used coaching style and has shown to improve various physical and psychological conditions (Rubak et al., 2005). As a result, researchers have implemented MI techniques in CA design (He et al., 2022; Jörke et al., 2024; Kumar et al., 2024; Olafsson et al., 2020; Samrose & Hoque et al., 2022; Schulman et al., 2011; Smriti et al., 2022; Steenstra et al., 2024). We referenced the MI-based prompt used in Steenstra et al. (2024) as the template for our health coach (see Appendix A for prompts we used).

---

<sup>2</sup> The final sample excluded one participant because they did not complete all six sessions.

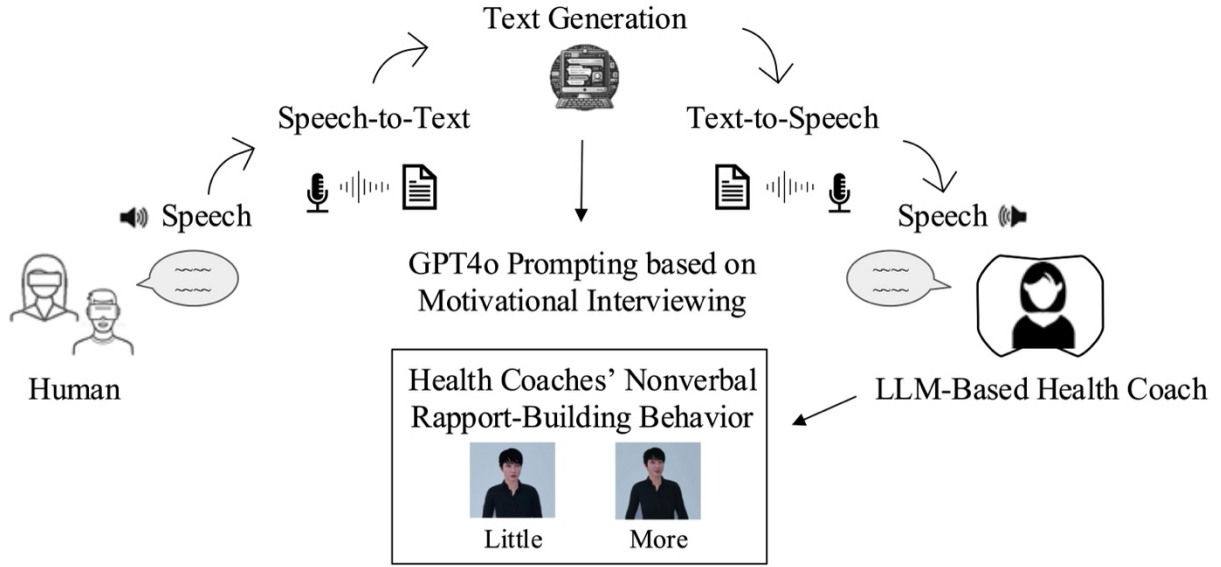
### ***Integrating Dialogue System with Varying Nonverbal Rapport***

See Figure 3 for a summary of the technical configuration of the LLM-based ECA. We created a female avatar using the Metahuman platform and imported it into the Unreal engine. Then, we selected OpenAI’s GPT-4o as the LLM, Microsoft Azure’s Cora Female Multilingual Voice<sup>3</sup> as the text-to-speech (TTS) model, and Convai’s speech-to-text (STT) capabilities to power the avatar. For the pilot version, we created a simple and empty office environment in Unreal and placed the female avatar in a seating position behind a desk. Finally, we equipped the female avatar with two levels of nonverbal rapport-building behaviors: *Little* vs. *More*. The two levels of nonverbal rapport-building behavior levels exhibit basic lip sync and eye blink. However, they differed in the following ways:

1. **Gaze:** *More* made eye contact with the human clients vs. *Little* looked slight downward and did not follow the human clients’ movements.
2. **Listening Behavior:** *More* smiled and showed slight upper body movement vs. *Little* did not have any facial expressions and sat still with no upper body movement while the human client spoke.
3. **Talking Behavior:** *More* used upper body gestures vs. *Little* sat still with arms placed to the side while talking.

---

<sup>3</sup> For the pilot, we originally used OpenAI’s Nova voice as the text-to-speech. However, due to technical issues and the latency, we switched the voice to Microsoft Azure’s Cora female voice midway through the pilot.

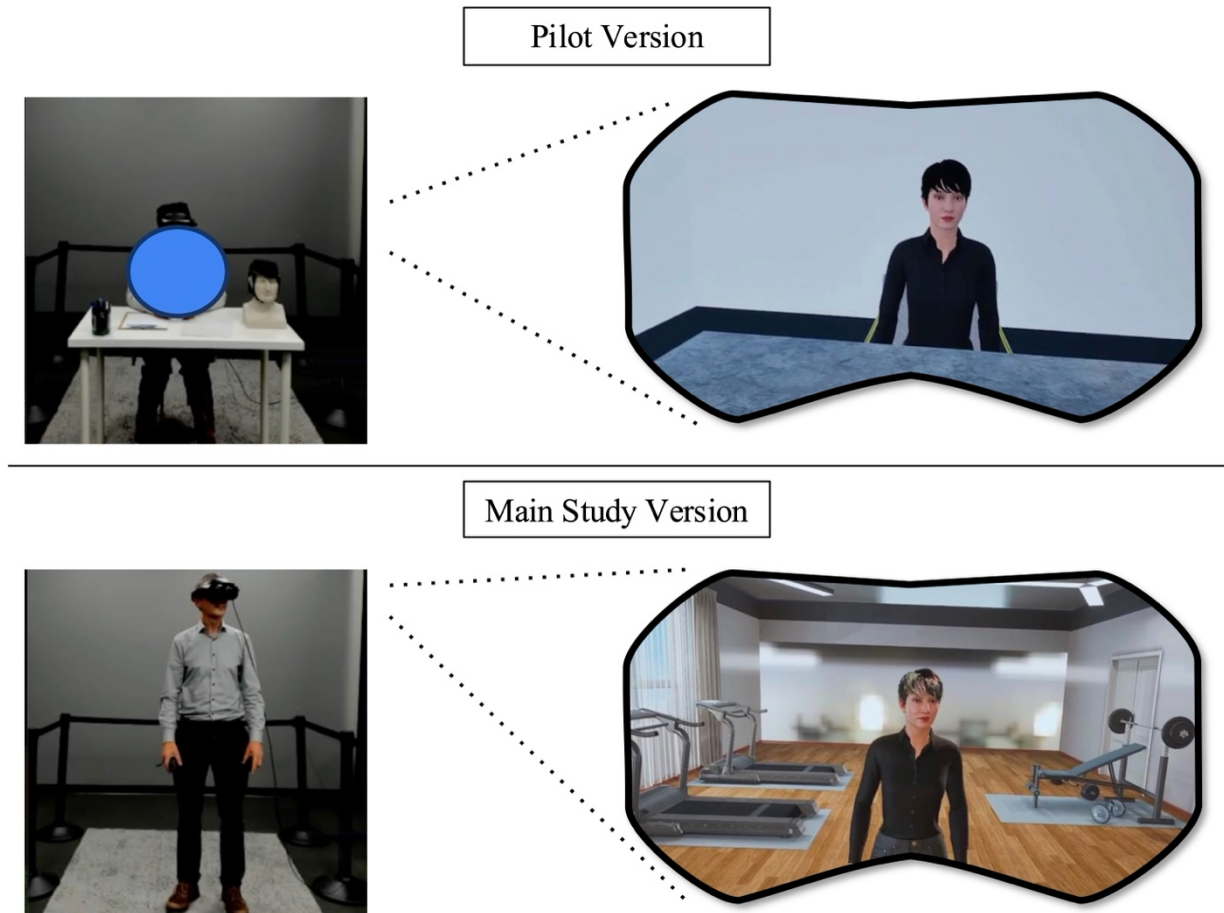


**Figure 3.** Illustration of Human-LimAI2 Interaction in Immersive VR. *First, the participant enters the virtual environment wearing the Meta Quest Pro headset and speaks to the LLM-based health coach (named Dr. Lauren Smith). The participant’s words are converted into text via the Convai system, and the converted text is then inputted as the prompt into ChatGPT. The response generated by ChatGPT is further processed through Microsoft Azure’s text-to-speech (TTS) system and inputted into the LLM-based health coach. Finally, the LLM-based health coach responds to the participants with either little or more nonverbal rapport building behaviors, depending on the random assignment at the beginning of the study.*

After we developed the prototypes of the LLM-based ECAs, we conducted a pilot evaluation to ensure the participants noticed the nonverbal rapport-building behavior manipulations and make any other modifications as needed. A total of 16 people participated in the pilot evaluation ( $M_{age} = 26.13$ ,  $SD_{age} = 9.37$ , 38% self-identified White or European American, 63% self-identified Female). We recruited the participants through the local institution’s community participant pool and word-of-mouth. The participants received the same



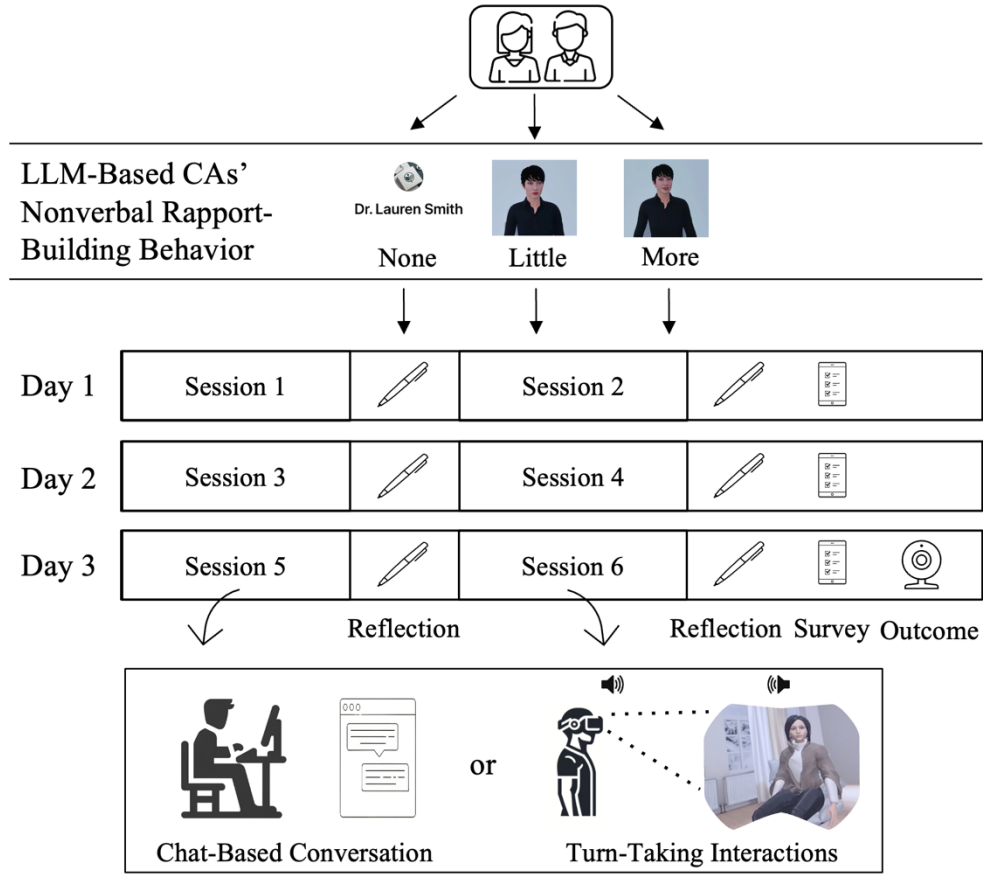
cash compensation as the main study participants for completing all six sessions. The pilot showed that the LLM-based ECAs' seating position covered the varying nonverbal behaviors at times, and the participants' seating position prevented active nonverbal expression. Furthermore, the simple office environment and the seated position established a task-focused environment that sometimes appeared rigid. Applying the findings from the pilot study, we made the following modifications to the main study version of the LLM-based health coach in VR. First, we changed the environment from a simple office to a gym by importing the gym asset from the Unreal Engine's Fab marketplace (Big-G, 2024). Next, the interaction occurred standing up rather than sitting down to enhance the interaction (see Figure 4).



**Figure 4.** Pilot and Main Study Versions of Human-AI interaction. *For the main study, the environment changed from an empty office to a gym, with the LLM-based health coach standing instead of sitting.*

### **Main Study Experimental Conditions and Procedures**

See Figure 5 for the illustration of study procedures.



**Figure 5.** Overview of the Study Design. *We randomly assigned the participants to one of three LLM-based health coaches, varying in the display of nonverbal rapport-building behavior (little vs. more vs. text-based chatbot as baseline comparison). Then, they interacted with the health coach for six sessions (two sessions per visit to the lab). Participants completed a reflection task after each session and the post-task survey at the end of each visit. After the completion of session 6, the participants were asked to provide a testimonial about their experience.*

All 30 participants came into the lab three times and completed two coaching sessions during each visit. The study had three experimental conditions. In addition to the LLM-based ECAs with little vs. more nonverbal rapport-building behaviors, we created a purely text-based LLM-based health coach using the OpenAI GPT platform for baseline comparison. We used the

same instructions and knowledge base to create the LLM-based health coach. To remove as many embodiment cues as possible, we used an image of a leaf as the icon. During the main study, participants were randomly assigned to one of the three LLM-based health coaches (10 participants for each group) and completed the six sessions with the same health coach.

Those assigned to the LLM-based ECA followed the following experimental procedures. On day 1, participants came into the lab, consented to the study, and completed a brief pre-survey that included questions about demographics and health consciousness. Then we provided specific instructions for the first session with health coach and asked the participants to write down specific health goals or topics to discuss. Next, we helped the participants put on the Meta Quest pro headset and calibrated the eye-tracking device. Upon calibrating the eye tracking system, the participants completed a *demo task* by entering the virtual gym and talking with a receptionist without any nonverbal rapport-building behaviors. This *demo task* familiarized the participants with the technical aspects of the interactions.

After the *demo task*, participants engaged in the first *session task*, which involved discussing their health goals with the health coach for 5 minutes (session 1). We, then, instructed the participants to journal about the content of the session (e.g., What did you talk about? What was helpful? What could have been better?) and goals they want to discuss with the health coach during the next session (*reflection task*). For session 2, we calibrated the headset again, and the participants completed the second *session task* and *reflection task*. Finally, participants participated in *post-session tasks* (a brief interview and a post-session questionnaire about their experience).

On days 2 and 3, participants first responded to a brief pre-session questionnaire that asked how much they worked on the goals they previously discussed with the health coach. Then

we repeated the headset calibration routine, *session tasks* (sessions 3 and 4 on day 2 and sessions 5 and 6 on day 3), and the *reflection tasks* from day 1. At the end of the last session (session 6), we asked the participants to provide a brief verbal testimonial about their experience. Finally, we fully debriefed the participants about the purpose of the study. Those assigned to the text-based LLM-based health coach followed the same procedures as above without the calibration routine and the demo task.

## Measures and Data Analysis

See Appendix B for the survey measures we used.

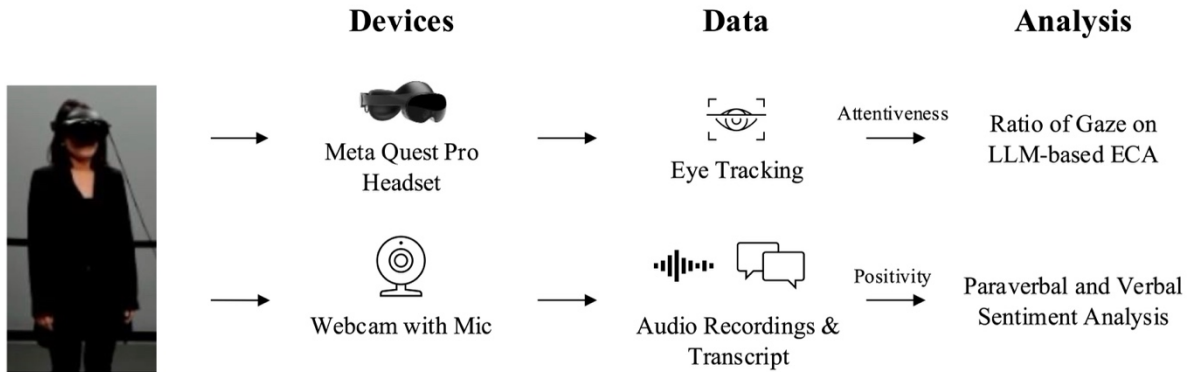
### *Measures of Human Clients' Attentiveness and Positivity*

**Attentiveness.** We collected gaze information from the Meta Quest pro headset through the Unreal Engine's OpenXR plugin (see Figure 6). Specifically, we tracked which objects in the environment people looked at during each session and calculated the ratio of gaze on the LLM-based ECA (*ratio = number of times gaze on the health coach was detected / total number of gaze toward objects detected*).

**Positivity.** In addition to the tracking data from the headset, we also collected audio recordings and transcripts of the conversations. For *paraverbal positivity*, we first diarized – or split – the audio recordings from each session by speaker turn using Hugging Face's pyannote.audio Python package (Bredin, 2023; Plaquet & Bredin, 2023). Then we conducted speech emotion recognition for the human clients' diarized audio files using audEERING's open-source AI models (Wagner et al., 2023). This process assigned valence scores from 0 (negative) to 1 (positive) to each participant turn. Finally, the valence scores were averaged for each participant.

For *verbal positivity*, we followed a similar process using the conversation transcripts

from each session. First, we split speaker turns from each session transcript and assigned valence scores (from -1 to 1) to the participants' responses using natural language processing (VADER Sentiment Analysis Tool; Hutto & Gilbert; 2014). We specifically selected VADER because of its capabilities to calculate valence scores rather than simply classifying the valence as negative, neutral, or positive. Next, we averaged the valence scores of the responses for each session.



**Figure 6.** Technological Setup to Study Human Clients' Nonverbal Rapport-Building Behavior.

### ***Measures of Subjective Rapport and Interaction Outcomes***

**Human Clients' Subjective Rapport.** The way in which rapport has been measured via self-report varies somewhat across studies and researchers. However, across different studies, items used to indicate rapport cluster around concepts such as social harmony, warmth, coordination, and cooperation. Therefore, we adopted perceptions of the interaction ratings from Lim et al. (2024a) and perceived social presence scale from Bente et al. (2023) to represent various aspects of rapport, as defined in the introduction. For the perceptions of the interaction ratings, participants evaluated the following statements from 1 (strongly disagree) to 7 (strongly agree): “We had good rapport,” “The interaction was harmonious,” “The interaction was cooperative,” “The interaction was coordinated,” “The interaction was warm,” and “The interaction was friendly.” We averaged the ratings from these six items to form the *experienced*

*rapport* measure and found good reliability across all conditions and sessions (*Cronbach's alpha* = .83; *M* = 5.40; *SD* = .92).

We also adopted the *perceived social presence* scale to examine how people experienced their encounter with the LLM-based ECA through VR. Human-computer interaction studies have generally considered perceived rapport and social presence as separate concepts (e.g., von der Pütten et al., 2012). However, in contrast to the current study, those studies featured limited immersion and turn-taking interactions. Since this study examines human-AI interaction in immersive environments, we use social presence to signal a facet of rapport-building in this specific context. The scale asked participants to rate six statements from 1 (Strongly disagree) to 5 (Strongly agree): “*I was attentive to the [body language/language]<sup>4</sup> of the [avatars/health coach],*” “*I had the sensation that the [avatar/health coach] could also see me,*” “*It felt as if I could interact with the [avatar/health coach],*” “*I was aware of the [avatars’/health coach’s] moods,*” “*I could feel what the [avatars/health coach] felt,*” and “*The [avatars in the virtual environments were/health coach was] engaging.*” The six items exhibited good reliability across all conditions and sessions (*Cronbach's alpha* = .83; *M* = 2.93; *SD* = .96).

**Interaction Outcomes.** We measured two types of outcomes: health-related outcomes and session evaluations. Health-related outcomes included behavioral intentions to work on goals discussed with the health coach (Schwarzer, 2008), response efficacy (Witte, 1994), and self-efficacy (Bandura, 1982), all rated on a 5-point Likert scale (1 – Strongly Disagree, 5 – Strongly Agree). For behavioral intentions, we first asked participants, “*What is one specific goal you discussed with the AI health coach?*”. Then participants rated two items, “*I intend to work on the*

---

<sup>4</sup> The language of the scale differed slightly depending on the condition. Those interacting with the LLM-based ECA saw the first part of the bracket (e.g., “*I was attentive to the body language of the avatars*”) while the rest saw the second part of the bracket (e.g., “*I was attentive to the language of the health coach*”).

goal I wrote above within the next [month/two weeks].” The two items exhibited good reliability (Cronbach’s  $\alpha = .86$ ;  $M = 4.27$ ;  $SD = .63$ ).

The modified response efficacy scaled comprised three statements: “*The health coach’s recommendations work in helping me reach my health goal,*” “*Following the health coach’s recommendations is effective in helping me reach my health goal,*” and “*If I follow the health coach’s recommendations, I am more likely to reach my health goal.*” Finally, the modified self-efficacy scale asked participants to rate three sentences: “*I am able to follow the health coach’s recommendations to reach my health goal,*” “*The health coach’s recommendations are easy to follow to reach my health goal,*” and “*Following the health coach’s recommendations is convenient.*” Both scales exhibited good to high reliability (Cronbach’s  $\alpha_{ResEfficacy} = .88$ ;  $M_{ResEfficacy} = 4.21$ ;  $SD_{ResEfficacy} = .55$ ; Cronbach’s  $\alpha_{SelfEfficacy} = .91$ ;  $M_{SelfEfficacy} = 4.13$ ;  $SD_{SelfEfficacy} = .73$ ).

To measure the outcomes of the coaching sessions, we adopted and modified the *Session Impacts Scale* (SIS; Elliott & Wexler, 1994; Stiles et al., 1994). This scale has been widely used to study the effectiveness of health interventions (e.g., Ackerman et al., 2010; Lingiardi et al., 2017; Simpson & Reid, 2014). For this measure, participants rated how much they agreed with 9 statements about the sessions’ helpfulness (e.g., “*I know have new insight about myself or have understood something new about me,*” “*I now feel supported, reassured, confirmed, or encouraged by the health coach*”) and 6 statements about the hindering impact (“*I feel the health coach is cold, bored, or doesn’t care about me*”) from 1 (Not at all) to 5 (Very much)<sup>5</sup>. Both subscales exhibited good to high reliability (Cronbach’s  $\alpha_{helpful} = .94$ ;  $M_{Helpful} = 3.34$ ;  $SD_{Helpful} = 1.19$ ; Cronbach’s  $\alpha_{Hindering} = .79$ ;  $M_{Hindering} = 1.48$ ;  $SD_{Hindering} = .80$ ).

---

<sup>5</sup> We excluded one item, “As a result of this session, I now feel relief from uncomfortable or painful feelings,” from the originally scale because the sessions were not specifically focused on therapy.



**Overall Satisfaction with the Intervention.** Finally, we measured satisfaction in two ways. First, we asked the participants how likely they are to recommend the health coach to others, on a scale from 1 (would never recommend to anyone) to 10 (would recommend to everyone now). Second, we recorded a brief testimonial from each participant about their experience and extracted the verbal and paraverbal valence scores using VADER Sentiment Analysis Tool and audEERING's AI models (see the **Positivity** subsection above).

### ***Data Analysis***

We used R and Python for data cleaning and analyses. To test *hypothesis 1 (H1)*, we first averaged the ratio of gaze on the health coach calculated for session 1 and 2 for each participant. Next, we fitted a beta regression model (R betareg package; Cribari-Neto & Zeileis, 2010) to examine whether the health coach's nonverbal rapport-building behavior influenced people's gaze toward the health coach. For *H2*, we averaged the valence scores of the responses for sessions 1 and 2 and used an independent t test to test whether the health coach's nonverbal rapport-building behaviors increased the positive sentiment detected in participants' paraverbal and verbal responses. In addition, we also used the independent t-test to understand the influence of the health coach's nonverbal rapport-building behavior on the participants' subjective rapport (*H3*). The last analysis method involved fitting multiple regression models<sup>6</sup> to explore whether the participants' attentiveness, positivity, and subjective rapport can predict interaction outcomes and overall intervention satisfaction (*H4* and *H5*).

To explore the *RQ* about how rapport builds overtime, we first averaged the gaze, verbal valence, and paraverbal valence data for sessions 3 and 4, and sessions 5 and 6. Then we fitted

---

<sup>6</sup> A beta regression model was fitted to examine how attentiveness, positivity, and subjective rapport predicted people's paraverbal expression of satisfaction with the intervention. Multiple linear regression models were fitted for all other dependent variables.

mixed effects models<sup>7</sup> for each rapport dimension (*attentiveness*, *positivity*, and *subjective rapport*) and interaction outcomes. LLM-based ECA's nonverbal rapport-building behaviors (little vs. more), the visit to the lab (visit1: sessions 1 & 2; visit 2: sessions 3 & 4; visit 3: sessions 5 & 6), and the interaction of the two variables were the main predictors. We allowed the intercepts to vary by participant to account for the repeated measure design.

---

<sup>7</sup> We fit mixed effects beta regression models (glmmTMB R package; Brooks et al., 2017) for gaze and paraverbal positivity since they are bounded variables with values between 0 and 1. Mixed effects linear effects models (lme4 R package; Bates et al., 2010) were fitted for all other dependent variables.

## SECTION 4: RESULTS

### H1-3: Effect of LLM-Based Health Coaches' Nonverbal Behavior on Rapport

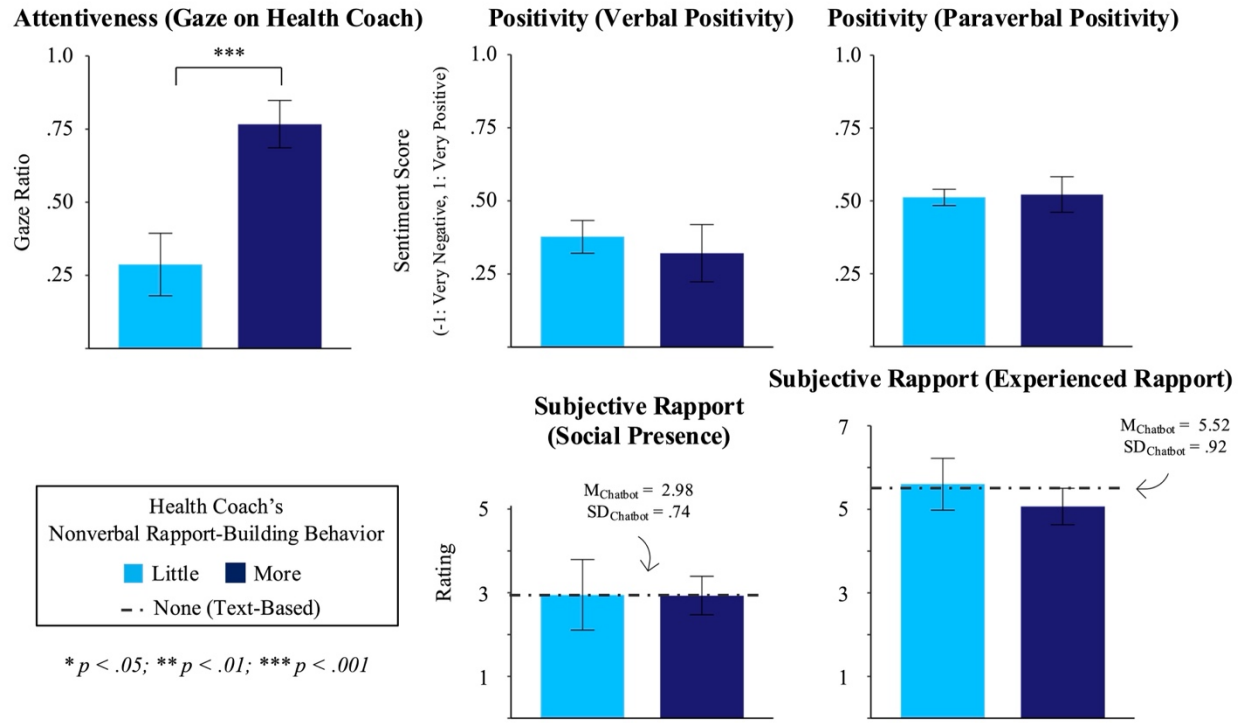
Hypotheses 1-3 examined the effects of LLM-based ECAs' nonverbal rapport-building behaviors on human clients' expression of rapport: attentiveness, positivity, and subjective rapport (see Figure 7). First, people who interacted with the LLM-based ECA with *more* nonverbal rapport-building behaviors gazed at the health coach more than those who interacted with the *little* rapport-building LLM-based ECA ( $M_{\text{More}} = .77$ ,  $SD_{\text{More}} = .11$ ;  $M_{\text{Little}} = .29$ ,  $SD_{\text{Little}} = .15$ ;  $\chi^2(1) = 58.03$ ,  $p < .001$ ; see Table 1). This suggested that the LLM-based ECAs' nonverbal rapport building behaviors increases people's attentiveness to the ECA, supporting *H1*.

Table 1. Gaze on Health Coach by the Health Coaches' Nonverbal Rapport-Building Behavior

	<i>Estimate</i>	<i>S.E.</i>	<i>z</i>	<i>p-value</i>
Intercept	-.90	.18	-4.91	<.001
More Rapport-Building Behavior (vs. <i>Little</i> )	2.06	.27	7.62	<.001

*Note.* *S.E.* = *Standard Error*

However, the results from the *independent samples t*-test showed that LLM-based health coaches' nonverbal rapport-building behavior did not lead to significant differences in human clients' *paraverbal* and *verbal positivity* and *subjective rapport* (see Table 2). Thus, *H2a*, *H2b*, and *H3* were not supported.



**Figure 7.** Effect of the LLM-Based ECAs' Nonverbal Rapport-Building Behaviors on the Human Clients' Attentiveness, Positivity, and Subjective Rapport.

**Table 2.** Positivity and Subjective Rapport by ECAs' Nonverbal Rapport-Building Behavior

	Little		More		<i>t</i>	<i>p-value</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>		
<b>Positivity</b>						
Verbal Positivity	.38	.08	.32	.14	-1.13	.28
Paraverbal Positivity	.51	.04	.52	.09	.34	.74
<b>Subjective Rapport</b>						
Experienced Rapport	5.6	.87	5.07	.61	-1.59	.13
Social Presence	2.95	1.18	2.93	.64	-.04	.97

*Note.* *SD* = *Standard Deviation*

#### H4-5: Effect of Human Clients' Expressions of Rapport on Outcomes

Next, *H4* focused on the relationship between human clients' expressions of rapport and interactions outcomes, including evaluation of the sessions and health-related factors. First, we found that feeling hindered by the session was significantly predicted by human clients' experienced rapport ( $Estimate = -.21$ ,  $SE = .09$ ,  $p = .039$ ; see Table 3). In other words, the more rapport human clients perceived during the interaction, the less likely they were to feel hindered by the conversations with the health coach. We found similar results for feeling helped by the sessions, though this association was directional ( $Estimate = .70$ ,  $SE = .35$ ,  $p = .061$ ).

Table 3. Influence of Rapport on Evaluation of Coaching Sessions

	<i>Estimate</i>	<i>S.E.</i>	<i>t</i>	<i>p-value</i>
<b>Session Impacts: Hindering</b>				
Intercept	2.45	.58	4.24	<.001
Gaze on Health Coach	.05	.28	.17	.87
Positivity (Verbal Positivity)	-.01	.83	-.01	.99
Positivity (Paraverbal Positivity)	-.48	1.52	-.32	.76
Subjective Rapport (Social Presence)	.07	.08	.90	.38
Subjective Rapport (Experienced)	-.21	.09	-2.27	.04
<b>Session Impacts: Helpfulness</b>				
Intercept	.26	2.12	.12	.91
Gaze on Health Coach	-.39	1.02	-.38	.71
Positivity (Verbal Positivity)	-.12	3.03	-.04	.97
Positivity (Paraverbal Positivity)	-2.86	5.58	-.51	.62
Subjective Rapport (Social Presence)	.21	.29	.74	.47
Subjective Rapport (Experienced)	.70	.35	2.03	.06

*Note.* *S.E.* = *Standard Error*

For health-related outcomes, none of the expressed rapport variables (attentiveness, positivity, and subjective rapport) significantly predicted the health-related factors (self-efficacy, response efficacy, and behavioral intentions). However, we found tentative evidence that human clients' subjective rapport directionally influenced their efficacy levels (see Table 4).

*Experienced rapport* during the interaction was slightly associated with greater self-efficacy ( $Estimate = .58, SE = .31, p = .082$ ) while *perceived social presence* was directionally correlated with response efficacy ( $Estimate = .31, SE = .15, p = .062$ ).

Finally, *H5* examined the relationship between human clients' expressed rapport and overall satisfaction with the intervention (completion of all six sessions with the health coach). Our results partially supported *H5* (see Table 5). We found that the more people experienced rapport during the interaction, the more likely they are to recommend the LLM-based ECA to others ( $Estimate = 1.59, SE = .59, p = .017$ ). Furthermore, the ratio of gaze on the health coach significantly predicted people's paraverbal expression of satisfaction with the intervention overall ( $Estimate = .95, SE = .34, p = .005$ ).

Table 4. Influence of Rapport on Health-Related Outcomes

	<i>Estimate</i>	<i>S.E.</i>	<i>t</i>	<i>p-value</i>
<b>Self-Efficacy</b>				
Intercept	1.16	1.91	.61	.55
Gaze on Health Coach	.65	.91	.72	.49
Positivity (Verbal Positivity)	1.26	2.72	.46	.65
Positivity (Paraverbal Positivity)	-.89	5.01	-.18	.86
Subjective Rapport (Social Presence)	-.15	.26	-.58	.57
Subjective Rapport (Experienced)	.58	.31	1.88	.08
<b>Response Efficacy</b>				
Intercept	3.08	1.15	2.68	<b>.02</b>
Gaze on Health Coach	-.28	.55	-.51	.62
Positivity (Verbal Positivity)	-.15	1.64	-.09	.93
Positivity (Paraverbal Positivity)	1.24	3.02	.41	.69
Subjective Rapport (Social Presence)	.31	.15	2.03	.06
Subjective Rapport (Experienced)	-.06	.19	-.30	.77
<b>Behavioral Intentions</b>				
Intercept	2.49	1.54	1.62	.13
Gaze on Health Coach	-.14	.74	-.19	.86
Positivity (Verbal Positivity)	.51	2.21	.23	.82
Positivity (Paraverbal Positivity)	2.25	4.06	.55	.59
Subjective Rapport (Social Presence)	.19	.21	.93	.37
Subjective Rapport (Experienced)	.01	.25	.04	.97

*Note. S.E. = Standard Error*

Table 5. Influence of Rapport on Overall Satisfaction with the Intervention

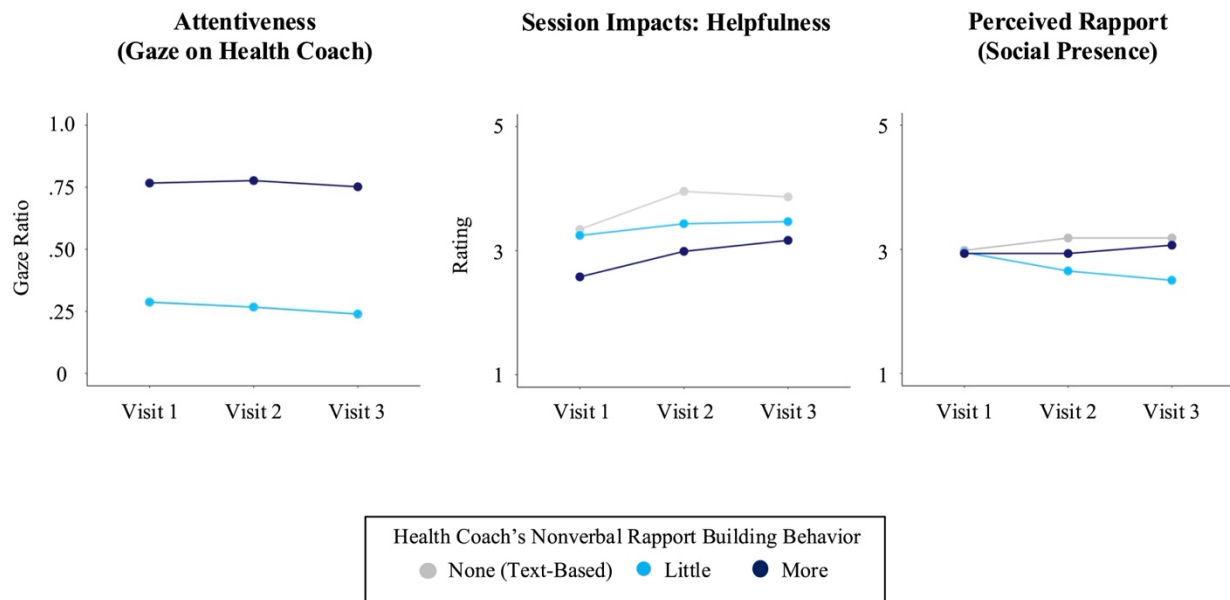
	<i>Estimate</i>	<i>S.E.</i>	<i>Statistic</i>	<i>p-value</i>
<b>Likelihood to Recommend</b>				
Intercept	-2.72	3.62	$t = -.75$	.47
Gaze on Health Coach	.18	1.73	$t = .10$	.92
Positivity (Verbal Positivity)	-3.04	5.17	$t = -.59$	.57
Positivity (Paraverbal Positivity)	1.96	9.51	$t = .21$	.84
Subjective Rapport (Social Presence)	.54	.49	$t = 1.12$	.28
Subjective Rapport (Experienced)	1.59	.59	$t = 2.70$	<b>.02</b>
<b>Paraverbal Expression of Satisfaction</b>				
Intercept	-2.32	.73	$z = -3.19$	<b>.001</b>
Gaze on Health Coach	.95	.34	$z = 2.80$	<b>.005</b>
Positivity (Verbal Positivity)	1.27	1.02	$z = 1.24$	.21
Positivity (Paraverbal Positivity)	2.16	1.88	$z = 1.15$	.25
Subjective Rapport (Social Presence)	-.09	.10	$z = -.95$	.34
Subjective Rapport (Experienced)	.14	.12	$z = 1.24$	.21
<b>Verbal Expression of Satisfaction</b>				
Intercept	.42	.28	$t = 1.54$	.15
Gaze on Health Coach	-.12	.13	$t = -.92$	.38
Positivity (Verbal Positivity)	-.06	.39	$t = -.15$	.88
Positivity (Paraverbal Positivity)	-.35	.72	$t = -.49$	.63
Subjective Rapport (Social Presence)	.06	.04	$t = 1.56$	.14
Subjective Rapport (Experienced)	.01	.04	$t = .23$	.82

*Note. S.E. = Standard Error*



## RQ: Longitudinal Effects of LLM-Based Health Coaches' Nonverbal Behavior

We also explored how the health coaches' nonverbal rapport-building behaviors influenced human clients' expressions of rapport as well as the interaction outcomes over time (see Figure 8). For attentiveness, we found significant main effect of LLM-based ECA's nonverbal rapport building behavior ( $\chi^2(1) = 60.81, p < .001$ ; see Table 6). This indicated that human clients who interacted with the *more* rapport-building health coach expressed greater attentiveness during the interaction compared to those who interacted with the *little* rapport-building health coach across the sessions.



**Figure 8.** Effect of the LLM-Based ECAs' Nonverbal Behaviors on Attentiveness, Positive Evaluation of the Session, and Perceived Rapport Over Time. *Visit 1 includes Sessions 1 and 2; Visit 2 includes Sessions 3 and 4; Visit 3 includes Sessions 5 and 6. The gray lines indicate the results from the Chatbot condition and is used as the baseline visual comparison.*

Table 6. Effect of the LLM-Based ECAs' Nonverbal Behaviors on Attentiveness Over Time  
(Analysis of Deviance, Type III Test)

	$\chi^2$	<i>df</i>	<i>p-value</i>
Intercept	23.46	1	<.001
Nonverbal Rapport-Building Behavior	60.81	1	<.001
Time (i.e., Lab Visit 1, 2, or 3)	3.91	2	.14
Nonverbal Rapport-Building Behavior x Time	1.12	2	.57

*Note.* *df* = Degree of Freedom

Furthermore, we found significant main effect of time on the extent people felt helped by the sessions ( $F(2, 36) = 3.45, p = .042$ ; Table 7). In other words, the perceived helpfulness of the sessions increased as people continued to interact with the health coach; this effect was more pronounced for those who interacted with the *more* rapport-building ECA (*difference between Visit 1 and 3* = .59,  $SE = .23, t = 2.60, p = .013$ ). Finally, the results for social presence (an aspect of subjective rapport) suggested a slight interaction between LLM-based health coaches' nonverbal rapport-building behavior and time ( $F(2, 36) = 2.38, p = .11$ ). Post-hoc analysis through estimated marginal means illustrated that those who interacted with the *little* rapport-building health coach in VR reported lower levels of social presence during the Visit 3 compared to Visit 1 (*difference* = -.45,  $SE = .19, t = -2.38, p = .023$ ). LLM-based ECAs' nonverbal rapport-building behaviors did not significantly affect other expressions of rapport and interaction outcomes (see Appendix C for all results).

Table 7. Effect of the LLM-Based ECAs' Nonverbal Behaviors on Positive Evaluation of the Session and Social Presence Over Time (Analysis of Variance, Type III Test)

	<i>F</i>	<i>df</i>	<i>p-value</i>
<b>Session Impacts: Helpfulness</b>			
Nonverbal Rapport-Building Behavior	1.21	(1, 18)	.29
Time (i.e., Lab Visit 1, 2, or 3)	3.45	(2, 36)	<b>.04</b>
Nonverbal Rapport-Building Behavior x Time	.67	(2, 36)	.52
<b>Perceived Rapport (Social Presence)</b>			
Nonverbal Rapport-Building Behavior	.40	(1, 18)	.53
Time (i.e., Lab Visit 1, 2, or 3)	.89	(2, 36)	.42
Nonverbal Rapport-Building Behavior x Time	2.38	(2, 36)	.11

*Note. df = Degree of Freedom*

## SECTION 5: DISCUSSION

Our study examined how rapport between LLM-based health coaches and human clients form and influence interaction outcomes. Specifically, we first examined the effects of LLM-based health coaches' nonverbal rapport-building behaviors on human clients' expressions of rapport (i.e., attentiveness and positivity during the interaction, subjective rapport after the interaction). Next, we studied the relationship between human clients' expressions of rapport and interaction outcomes (i.e., health-related factors: self-efficacy, response efficacy, behavioral intentions; overall satisfaction: likelihood to recommend, paraverbal and verbal valence in testimonial). Lastly, we illustrated how the indicators of rapport trended over the six coaching sessions. These findings set the foundation for future research in human-AI interaction.

### Main Findings

Overall, the results partially supported our predictions.

#### ***Prediction 1: LLM-Based ECAs' Nonverbal Rapport-Building Behavior Enhances Rapport***

As we predicted, results showed that those who interacted with the *more* rapport-building health coach showed greater attentiveness (i.e., looked at the health coach more) during the initial sessions compared to those who interacted with the *little* rapport-building health coach. This aligned with existing research about people's natural tendency to adapt to the others' behavior (Chartrand & Bargh, 1999; Feldman, 2017; Tickle-Degnen, & Rosenthal, 1990), even when the other interactant is not human (Bushmeier & Kopp, 2018; Gratch et al., 2007).

Positivity (i.e., verbal and paraverbal positivity) and subjective rapport (i.e., experienced rapport and social presence), on the other hand, did not significantly differ by LLM-based health coaches' nonverbal behavior. One potential explanation for this result is the context of the interaction with the LLM-based health coach. We purposefully instructed the participants not to

share any personal information for privacy reasons. Therefore, the conversations with the health coach tended to focus on physical health related topics such as adding certain nutrients to daily meals, improving exercise regiment, and stress management activities. It is possible that we could have found different results had the conversation involved more emotional or relational topics. Another possible explanation is the length of the interaction. In human-to-human coaching contexts, we would expect each session to last about 30 minutes to an hour at the minimum. This time frame allows for the health coach and the participants to investigate certain topics in depth. Thus, the 5-minutes long sessions with the LLM-based health coaches may not have provided enough time for the complex mechanisms of social interactions to fully unfold.

The third possible explanation, of course, is the limited realism of the interaction. Though we designed the virtual environment as a gym to enhance realism, the current study design did not allow for varying interactions, such as the LLM-based health coach showing how to do certain exercises. Also, we did not yet equip the LLM-based ECAs' capability to exhibit natural nonverbal behaviors that match the words they say or adapt to the users' behaviors in real-time. In fact, some participants mentioned these limitations during the post-session interviews. Existing studies generally support this link between realism and rapport-related factors within virtual environments like social presence (Oh et al., 2018).

### ***Prediction 2: Human Clients' Expressions of Rapport are Associated with Outcomes***

Through our examination of the relationship between rapport and outcomes (overall satisfaction with the intervention, session evaluations, and health-related factors), we found attentiveness and subjective rapport as the most promising predictors. For instance, people who were more attentive during the initial sessions were more likely to express higher level of satisfaction with the intervention in their speech (*verbal expression of satisfaction*) while sharing

the testimonial. In addition, those who perceived greater levels of harmony, cooperation, and warmth during the interactions (*experienced rapport*) were more likely to recommend the LLM-based health coach to others after completing all six sessions. These individuals were also less likely to feel negative effects from the initial conversations with the health coach (*hindering session impacts*). These results highlight the close link between people's expressions of rapport and their satisfaction with the LLM-based ECA.

### ***Investigating the Effects of LLM-Based ECAs' Nonverbal Behavior Over Time***

Lastly, our longitudinal study provided important insights into how LLM-based health coaches' nonverbal behavior influenced rapport and interaction outcomes over the six coaching sessions. We found that people who talked with the *more* rapport-building health coach, on average, expressed greater attentiveness during the interaction compared to their counterparts across all sessions. This finding strengthens our previous argument that people naturally adjust to the other's nonverbal behaviors (H1), further highlighting the importance of nonverbal cues during interactions (Burgoon & Saine, 1978; Burgoon et al., 1984; Patterson, 1982).

Also, the time people spent with the LLM-based health coach enhanced the perceived helpfulness of the sessions, regardless of the health coaches' nonverbal behavior. In other words, the more people interacted with the health coach, the more they felt they benefited from the sessions. The effect was even more pronounced among those who interacted with the *more* rapport-building health coach. Interestingly, for those who interacted with the *little* rapport-building health coach, their perceived social presence trended downward over time. Thus, our results indicate that health interventions employing LLM-based ECAs show promise in enhancing long-term interaction outcomes, especially if the ECAs exhibit more natural nonverbal rapport-building behaviors.

## Theoretical Implications

Our study has significant implications for communication research. With the development of LLM-based ECAs in VR platforms, people can now converse with an AI face-to-face in various settings and contexts. During these turn-taking interactions with the ECAs, people exchange nonverbal, as well as verbal, cues with the ECAs, eliciting complex social interaction patterns and relational dynamics. However, our current understanding of human-AI communication draws on studies that used CAs with limited embodiment (e.g., text-based CAs like ChatGPT, voice assistants like Amazon's Alexa) or implemented Wizard-of-Oz type approaches<sup>8</sup>. As a result, many human-AI communication studies apply general theoretical approaches and concepts from computer-mediated/interpersonal communication or media effects research (e.g., CASA paradigm; Oh & Ki, 2024; AI-mediated communication; Hancock et al., 2020; expectancy violation theory and social scripts; Lew & Walther, 2023; HAIL-TIME model; Sundar, 2020).

While these studies have advanced the sub research area of human-AI communication significantly in the last few years, they do not address the multifaceted nature of face-to-face interactions with the LLM-based ECAs. Therefore, this study pushes forward the human-AI communication research agenda by 1) beginning to unpack the rapport-building processes that occur during face-to-face human-AI interaction and 2) examining how this process unfolds over multiple conversations.

Building on rapport theory (Tickle-Degnen & Rosenthal, 1990) and other related works, we identified three main ways human clients can express rapport: attentiveness and positivity

---

<sup>8</sup> Examples of these types of approaches include showing participants the same content created by researchers but labeling the source of the message as human or AI. In these studies, people are generally not interacting with a conversational agent or AI-generated content.

during the interaction and perceived rapport reported after the interaction. By testing the effect of LLM-based ECAs' nonverbal rapport-building behavior on each of these dimensions of rapport, we were able to more precisely investigate the link between the ECAs' nonverbal behaviors and the rapport-building process during the interaction. Also, our examination of the relationship between each dimension of rapport and outcomes – namely efficacy, behavioral intentions, perceived session impacts, and satisfaction with the intervention – begins to elucidate how LLM-based ECAs' nonverbal behaviors may ultimately enhance specific outcomes. This work, then, advances recent efforts by communication scholars to understand interpersonal dynamics during turn-taking conversations (e.g., dynamic dyadic systems approach; Solomon et al., 2021; social gaze patterns between speaker and listener; Schmäzle et al., 2024).

Furthermore, our evaluation of human-AI interaction over multiple health-focused conversations suggested that LLM-based ECAs' nonverbal rapport-building behavior may have significant impacts in the long-term. While we know from every-day experience that rapport is a dynamic process that unfolds over various stages of a relationship, limited empirical work about human-AI communication implements a longitudinal design. This study substantially contributes to a large body of literature highlighting the importance of rapport in relationships (Capella, 1990; Gratch et al., 2006) by studying the effects of LLM-based ECAs' behavior over time.

Finally, the results from this study also inform human-to-human communication processes. Concurrent with the trends toward AI, big data, and advances in measurement, there is a general trend across disciplines to use simulation to advance theory. In the context of communication research, building machines that can communicate naturally via verbal or nonverbal channels allows us to study interpersonal communication processes in a more controlled way (e.g., by using embodied AI agents as artificial confederates in otherwise hard-to-



control interpersonal settings). More importantly, through simulated interactions, we can gain deep insights into the generative mechanisms (e.g., how exactly eye-gaze leads to rapport, or which social signals make conversations flow and bring about desired effects). Thus, the simulation approach - whether in the form of agent-based modeling at the societal level (e.g., Park et al., 2023) or at the level of dyadic social interaction, as studied here, can advance theory by uncovering the nuanced mechanisms underlying social interaction as a whole.

### **Practical Implications**

Influential LLM-based ECAs, capable of fluent conversation and natural nonverbal rapport-building behavior, could have a significant impact across almost all domains of human life (e.g., education, customer service, organizational context, health, social support). In educational settings, for example, LLM-based ECAs can serve as engaging teaching assistants or tutors, using nonverbal cues to deliver content more effectively. The ECAs can also act as receptionists, administrative assistants, and facilitators of virtual meetings, conferences, or even business negotiations. Within health and support contexts, LLM-based ECAs can act as doctors, coaches, or peer supporters who provide services that augment the work of human professionals. As the current limitations of the LLM-based ECAs (e.g., latency in responses, privacy concerns) improve and extended reality systems such as VR, augmented reality, and mixed reality further advance, we can expect that LLM-based ECAs will more widely implemented into these domains of human life.

### **Limitations and Future Directions**

This study has a few limitations. First, for feasibility reasons, we had a small sample size (about 10 people per condition). Also, this study focused on health coaching contexts and specifically asked participants not to disclose personal information, leading to relatively

information-focused interactions. Finally, certain technological limitations (e.g., brief internet outage, bug in the AI software) could have potentially interfered at random moments of people's interactions with the LLM-based health coach. Future studies should replicate and extend our findings in different contexts and with larger sample sizes.

## SECTION 6: CONCLUSION

We used rapport theory as the framework to examine the effects of LLM-based embodied conversational agent (ECA)’s nonverbal behaviors in the context of health coaching. To conduct this study, we built two types of LLM-based health coaches in virtual reality using the Unreal platform. The *little* rapport-building health coach displayed minimal nonverbal behavior during conversations (e.g., no direct eye contact, no upper body movement) while the *more* rapport-building health coach displayed various rapport-building behaviors (e.g., smiling while listening, upper body movement while responding). Participants were randomly assigned to one of the two types of health coaches and completed six coaching sessions with the same health coach in immersive virtual reality (VR). Findings showed that those who interacted with the more rapport-building health coach expressed greater attentiveness across all six sessions (measured via ratio of gaze on the health coach). Also, we found attentiveness and subjective rapport during the initial interactions (sessions 1 and 2) as the most promising predictors of human clients’ overall satisfaction with the intervention at the end of all six sessions. Finally, the results indicated that the more people interacted with the health coach, the more they felt they benefited from the sessions, with effects even more pronounced for those who interacted with the *more* rapport-building ECA. These findings have significant implications for communication research and practice.

## BIBLIOGRAPHY

- Abbe, A., & Brandon, S. E. (2013). The role of rapport in investigative interviewing: A review. *Journal of Investigative Psychology and Offender Profiling*, 10(3), 237-249. <https://doi.org/10.1002/jip.1386>
- Ackerman, J. M., Nocera, C. C., & Bargh, J. A. (2010). Incidental haptic sensations influence social judgments and decisions. *Science*, 328(5986), 1712-1715.
- Alanezi, F. (2024). Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. *Nutrition and Health*, 02601060241244563. <https://doi.org/10.1177/02601060241244563>
- Amorese, T., Greco, C., Cuciniello, M., Buono, C., Palmero, C., Buch-Cardona, P., Escalera, S., Torres, M. I., Cordasco, G., & Esposito, A. (2022, June). Using eye tracking to investigate interaction between humans and virtual agents. In *2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (pp. 125-132). IEEE.
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37(5), 715.
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7), 819-833.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122.
- Bente, G., Schmälzle, R., Jahn, N. T., & Schaaf, A. (2023). Measuring the effects of co-location on emotion perception in shared virtual environments: An ecological perspective. *Frontiers in Virtual Reality*, 4, 1032510.
- Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, 12(2), 120-138. <https://doi.org/10.1007/BF00986930>
- BIG-G. (2024). gym. <https://www.fab.com/listings/98430f1e-e527-4594-a1b2-5ca8d8bf9756>
- Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. *Communication in the age of Virtual Reality*, 15(32), 10-5555.
- Bredin, H. (2023, August). pyannote. audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe. In 24th INTERSPEECH Conference (INTERSPEECH 2023) (pp. 1983-1987). ISCA.
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for Zero-inflated Generalized Linear Mixed Modeling. *The R*

- Journal*, 9(2), 378-400. <https://doi.org/10.32614/RJ-2017-066>
- Burgoon, J. K., Buller, D. B., Hale, J. L., & de Turck, M. A. (1984). Relational messages associated with nonverbal behaviors. *Human Communication Research*, 10(3), 351-378.
- Burgoon, J. K., & Saine, T. (1978). *The unspoken dialogue: An introduction to nonverbal communication*. Houghton Mifflin School
- Buschmeier, H., & Kopp, S. (2018, July). Communicative listener feedback in human-agent interaction: Artificial speakers need to be attentive and adaptive. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems* (pp. 1213-1221).
- Cappella, J. N. (1990). On defining conversational coordination and rapport. *Psychological Inquiry*, 1(4), 303-305. [https://doi.org/10.1207/s15327965pli0104\\_5](https://doi.org/10.1207/s15327965pli0104_5)
- Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5), 519-538. <https://doi.org/10.1080/088395199117360>
- Cerekovic, A., Aran, O., & Gatica-Perez, D. (2016). Rapport with virtual agents: What do human social cues and personality explain?. *IEEE Transactions on Affective Computing*, 8(3), 382-395.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893.
- Choi, R., Kim, T., Park, S., Kim, J. G., & Lee, S. J. (2024). Private yet social: How LLM chatbots support and challenge eating disorder recovery. *arXiv*. <https://doi.org/10.48550/arXiv.2412.11656>
- Clavel, C., & Callejas, Z. (2015). Sentiment analysis: from opinion mining to human-agent interaction. *IEEE Transactions on Affective Computing*, 7(1), 74-93.
- Cloitre, M., Chase Stovall-McClough, K., Miranda, R., & Chemtob, C. M. (2004). Therapeutic alliance, negative mood regulation, and treatment outcome in child abuse-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 72(3), 411-416.
- Convai. (2025). Conversational AI characters. <https://www.convai.com/>
- Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34, 1-24.
- Elkins, A. C., & Derrick, D. C. (2013). The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, 22(5), 897-913.

- Elkins, A. C., Derrick, D. C., Burgoon, J. K., & Nunamaker Jr, J. F. (2012, January). Predicting users' perceived trust in Embodied Conversational Agents using vocal dynamics. In *2012 45th Hawaii International Conference on System Sciences* (pp. 579-588). IEEE.
- Elliott, R., & Wexler, M. M. (1994). Measuring the impact of sessions in process experiential therapy of depression: The Session Impacts Scale. *Journal of Counseling Psychology*, 41(2), 166.
- Estep, C. M., & Roberts, T. G. (2015). Teacher immediacy and professor/student rapport as predictors of motivation and engagement. *Nacta Journal*, 59(2), 155-163.
- Fatima, J. K. (2023). Does it matter to have rapport and social interaction on a group tour?. *Journal of Vacation Marketing*, <https://doi.org/10.1177/13567667231219024>
- Feldman, R. (2017). The neurobiology of human attachments. *Trends in Cognitive Sciences*, 21(2), 80-99.
- Frank, A. F., & Gunderson, J. G. (1990). The role of the therapeutic alliance in the treatment of schizophrenia: Relationship to course and outcome. *Archives of General Psychiatry*, 47(3), 228-236.
- Frisby, B. N., & Martin, M. M. (2010). Instructor–student and student–student rapport in the classroom. *Communication Education*, 59(2), 146-164. <https://doi.org/10.1080/03634520903564362>
- Gobl, C., & Chasaide, A. N. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40(1-2), 189-212.
- Goldstein, P., Losin, E. A. R., Anderson, S. R., Schelkun, V. R., & Wager, T. D. (2020). Clinician-patient movement synchrony mediates social group effects on interpersonal trust and perceived pain. *The Journal of Pain*, 21(11-12), 1160-1174.
- Graßmann, C., Schölmerich, F., & Schermuly, C. C. (2020). The relationship between working alliance and client outcomes in coaching: A meta-analysis. *Human Relations*, 73(1), 35-58.
- Gratch, J., & Lucas, G. (2021). Rapport between humans and socially interactive agents. In B. Ligrin, C. Prelachaud, & D. Traum (Eds.), *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (pp. 433-462).
- Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., & Morency, L. P. (2006). Virtual rapport. *Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA)*, 14-27. [https://doi.org/10.1007/11821830\\_2](https://doi.org/10.1007/11821830_2)
- Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *Intelligent Virtual Agents: 7th International Conference, Proceedings 7* (pp. 125-138). Springer Berlin Heidelberg.

- Guo, Z., Chheang, V., Li, J., Barner, K. E., Bhat, A., & Barmaki, R. L. (2023). Social visual behavior analytics for autism therapy of children based on automated mutual gaze detection. In *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies* (pp. 11-21). <https://doi.org/10.1145/3580252.3586976>
- Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89-100.
- Harrigan, J. A., Oxman, T. E., & Rosenthal, R. (1985). Rapport expressed through nonverbal behavior. *Journal of Nonverbal Behavior*, 9, 95-110. <https://doi.org/10.1007/BF00987141>
- He, L., Basar, E., Wiers, R. W., Antheunis, M. L., & Krahmer, E. (2022). Can chatbots help to motivate smoking cessation? A study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. *BMC Public Health*, 22(1), 726.
- Hessels, R. S., Cornelissen, T. H., Hooge, I. T., & Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology*, 71(3), 226. <https://doi.org/10.1037/cep0000113>
- Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual Reviews in Clinical Psychology*, 1(1), 91-111. <https://doi.org/10.1146/annurev.clinpsy.1.102803.143833>
- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PloS One*, 10(8), e0136905. <https://doi.org/10.1371/journal.pone.0136905>
- Huang, L., Morency, L. P., & Gratch, J. (2011). Virtual Rapport 2.0. In Vilhjálmsdóttir, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (Eds.) *Intelligent Virtual Agents. Lecture Notes in Computer Science*, 6895, 68–79. Springer. [https://doi.org/10.1007/978-3-642-23974-8\\_8](https://doi.org/10.1007/978-3-642-23974-8_8)
- Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI Conference on Web and Social Media*, 8(1), pp. 216-225.
- Iachini, T., Coello, Y., Frassinetti, F., & Ruggiero, G. (2014). Body space in social interactions: A comparison of reaching and comfort distance in immersive virtual reality. *PloS One*, 9(11), e111511.
- Jo, E., Jeong, Y., Park, S., Epstein, D. A., & Kim, Y. H. (2024, May). Understanding the impact of long-term memory on self-disclosure with large language model-driven chatbots for public health intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-21). <https://doi.org/10.1145/3613904.3642420>
- Joe, G. W., Simpson, D. D., Dansereau, D. F., & Rowan-Szal, G. A. (2001). Relationships

- between counseling rapport and drug abuse treatment outcomes. *Psychiatric Services*, 52(9), 1223-1229. <https://doi.org/10.1176/appi.ps.52.9.1223>
- Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, 43(3), 280-283.
- Jörke, M., Sapkota, S., Warkenthien, L., Vainio, N., Schmiedmayer, P., Brunskill, E., & Landay, J. (2024). Supporting physical activity behavior change with LLM-based conversational agents. *arXiv*. <https://doi.org/10.48550/arXiv.2405.06061>
- Kang, B., & Hong, M. (2025). Development and evaluation of a mental health chatbot using ChatGPT 4.0: Mixed methods user experience study with Korean users. *JMIR Medical Informatics*, 13, e63538. <https://doi.org/10.2196/63538>
- Karacora, B., Dehghani, M., Kramer-Mertens, N., & Gratch, J. (2012). The influence of virtual agents' gender and rapport on enhancing math performance. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34), 563-568.
- Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1), 78.
- Krämer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you—The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies*, 71(3), 335-349. <https://doi.org/10.1016/j.ijhcs.2012.09.006>
- Krupnick, J. L., Sotsky, S. M., Elkin, I., Simmens, S., Moyer, J., Watkins, J., & Pilkonis, P. A. (2006). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Focus*, 64(2), 532-277.
- Kumar, A. T., Wang, C., Dong, A., & Rose, J. (2024). Generation of backward-looking complex reflections for a motivational interviewing-based smoking cessation chatbot using GPT-4: Algorithm development and validation. *JMIR Mental Health*, 11(1), e53778.
- Leach, M. J. (2005). Rapport: A key to treatment success. *Complementary Therapies in Clinical Practice*, 11(4), 262-265. <https://doi.org/10.1016/j.ctcp.2005.05.005>
- Lepper, G., & Mergenthaler, E. (2007). Therapeutic collaboration: How does it work?. *Psychotherapy Research*, 17(5), 576-587.
- Lew, Z., & Walther, J. B. (2023). Social scripts and expectancy violations: Evaluating communication with human or AI chatbot interactants. *Media Psychology*, 26(1), 1-16.
- Lim, S., Schmälzle, R., & Bente, G. (2024a). Artificial social influence: Rapport-building, LLM-based embodied conversational agents for health coaching. <https://sueminnlim.com/publications/>



- Lim, S., Schmälzle, R., & Bente, G. (2024b). Artificial social influence via human-embodied AI agent interaction in immersive virtual reality (VR): Effects of similarity-matching during health conversations. *arXiv*. <https://doi.org/10.48550/arXiv.2406.05486>
- Lingiardi, V., Muzi, L., Tanzilli, A., & Carone, N. (2018). Do therapists' subjective variables impact on psychodynamic psychotherapy outcomes? A systematic literature review. *Clinical Psychology & Psychotherapy*, 25(1), 85-101.
- Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, 31(4), 557-564.
- Lubis, N., Sakti, S., Yoshino, K., & Nakamura, S. (2019). Positive emotion elicitation in chat-based dialogue systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(4), 866-877. <https://doi.org/10.1109/TASLP.2019.2900910>
- Mergenthaler, E. (2008). Resonating minds: A school-independent theoretical conception and its empirical application to psychotherapeutic processes. *Psychotherapy Research*, 18(2), 109-126.
- Miller, W. R. (1983). Motivational interviewing with problem drinkers. *Behavioural and Cognitive Psychotherapy*, 11(2), 147-172.
- Nakano, Y. I., & Ishii, R. (2010, February). Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (pp. 139-148).
- Nie, J., Shao, H., Fan, Y., Shao, Q., You, H., Preindl, M., & Jiang, X. (2024). LLM-based conversational ai therapist for daily functioning screening and psychotherapeutic intervention via everyday smart devices. *ACM Transactions on Computing for Healthcare (HEALTH)*.
- Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems*, 28(1), 17-48.
- Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5, 114.
- Oh, J., & Ki, E. J. (2024). Can we build a relationship through artificial intelligence (AI)? Understanding the impact of AI on organization-public relationships. *Public Relations Review*, 50(4), 102469.
- Olafsson, S., Wallace, B. C., & Bickmore, T. W. (2020, May). Towards a computational framework for automating substance use counseling with virtual agents. *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 19, 9-13.

- Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2, 1-22. <https://doi.org/10.1145/3586183.3606763>
- Patterson, M. L. (1982). A sequential functional model of nonverbal exchange. *Psychological Review*, 89(3), 231.
- Plaquet, A., & Bredin, H. (2023). Powerset multi-class cross entropy loss for neural speaker diarization. *arXiv*. <https://doi.org/10.48550/arXiv.2310.13025>
- Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, 79(3), 284.
- Ranjartabar, H., Richards, D., Bilgin, A. A., & Kutay, C. (2019). First impressions count! The role of the human's emotional state on rapport established with an empathic versus neutral virtual therapist. *IEEE Transactions on Affective Computing*, 12(3), 788-800.
- Rehm, M., & André, E. (2005, September). Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. In *International Workshop on Intelligent Virtual Agents* (pp. 241-252). Springer Berlin Heidelberg.
- Robb, D. A., Lopes, J., Ahmad, M. I., McKenna, P. E., Liu, X., Lohan, K., & Hastie, H. (2023). Seeing eye to eye: Trustworthy embodiment for task-based conversational agents. *Frontiers in Robotics and AI*, 10, 1234767.
- Rubak, S., Sandbæk, A., Lauritzen, T., & Christensen, B. (2005). Motivational interviewing: A systematic review and meta-analysis. *British Journal of General Practice*, 55(513), 305-312.
- Samrose, S., & Hoque, E. (2022). MIA: Motivational interviewing agent for improving conversational skills in remote group discussions. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1-24.
- Santos, K. A., Ong, E., & Resurreccion, R. (2020, June). Therapist vibe: children's expressions of their emotions through storytelling with a chatbot. In *Proceedings of the Interaction Design and Children Conference* (pp. 483-494).
- Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, 7(1), 31-44.
- Schmälzle, R., Jahn, N. T., & Bente, G. (2024). Charting the silent signals of social gaze: Automating eye contact assessment in face-to-face conversations. *bioRxiv*. <https://doi.org/10.1101/2024.08.28.610064>
- Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological Science*, 26(6), 877-891.

- Schulman, D., Bickmore, T. W., & Sidner, C. L. (2011, March). An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In *AAAI Spring Symposium: AI and Health Communication* (pp. 61-64).
- Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, 57(1), 1-29.
- Simpson, S. G., & Reid, C. L. (2014). Therapeutic alliance in videoconferencing psychotherapy: A review. *Australian Journal of Rural Health*, 22(6), 280-299.
- Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PloS One*, 8(1), e52766.
- Smriti, D., Kao, T. S. A., Rathod, R., Shin, J. Y., Peng, W., Williams, J., ... & Huh-Yoo, J. (2022). Motivational interviewing conversational agent for parents as proxies for their children in healthy eating: development and user testing. *JMIR Human Factors*, 9(4), e38908.
- Solomon, D. H., Brinberg, M., Bodie, G. D., Jones, S., & Ram, N. (2021). A dynamic dyadic systems approach to interpersonal communication. *Journal of Communication*, 71(6), 1001-1026.
- Sorin, V., Brin, D., Barash, Y., Konen, E., Charney, A., Nadkarni, G., & Klang, E. (2024). Large language models and empathy: Systematic review. *Journal of Medical Internet Research*, 26, e52597. <https://doi.org/10.2196/52597>
- Steenstra, I., Nouraei, F., Arjmand, M., & Bickmore, T. (2024, September). Virtual agents for alcohol use counseling: Exploring llm-powered motivational interviewing. *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*, 1-10. <https://doi.org/10.1145/3652988.3673932>
- Stiles, W. B., Reynolds, S., Hardy, G. E., Rees, A., Barkham, M., & Shapiro, D. A. (1994). Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *Journal of Counseling Psychology*, 41(2), 175.
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88.
- Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, 1(4), 285-293. [https://doi.org/10.1207/s15327965pli0104\\_1](https://doi.org/10.1207/s15327965pli0104_1)
- Von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6), 1641-1650. <https://doi.org/10.1016/j.chb.2010.06.012>

- Vowels, L. M., Francois-Walcott, R. R., & Darwiche, J. (2024). AI in relationship counselling: Evaluating ChatGPT's therapeutic capabilities in providing relationship advice. *Computers in Human Behavior: Artificial Humans*, 2(2), 100078. <https://doi.org/10.1016/j.chbah.2024.100078>
- Wagner, J., Triantafyllopoulos, A., Wierstorf, H., Schmitt, M., Burkhardt, F., Eyben, F., & Schuller, B. W. (2023). Dawn of the transformer era in speech emotion recognition: Closing the valence gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9), 10745-10759.
- Wang, N., & Gratch, J. (2009). Can virtual human build rapport and promote learning?. In *Artificial Intelligence in Education* (pp. 737-739). IOS Press.
- Wang, I., & Ruiz, J. (2021). Examining the use of nonverbal communication in virtual agents. *International Journal of Human-Computer Interaction*, 37(17), 1648-1673. <https://doi.org/10.1080/10447318.2021.1898851>
- Witte, K. (1994). Fear control and danger control: A test of the extended parallel process model (EPPM). *Communications Monographs*, 61(2), 113-134.
- Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences*, 118(37), e2106645118. <https://doi.org/10.1073/pnas.2106645118>
- Wohltjen, S., & Wheatley, T. (2024). Interpersonal eye-tracking reveals the dynamics of interacting minds. *Frontiers in Human Neuroscience*, 18, 1356680. <https://doi.org/10.3389/fnhum.2024.1356680>
- Woo, J., Shidara, K., Achard, C., Tanaka, H., Nakamura, S., & Pelachaud, C. (2024). Adaptive virtual agent: Design and evaluation for real-time human-agent interaction. *International Journal of Human-Computer Studies*, 103321. <https://doi.org/10.1016/j.ijhcs.2024.103321>
- Wu, S., Han, F., Yao, B., Xie, T., Zhao, X., & Wang, D. (2024). Sunnie: An anthropomorphic LLM-based conversational agent for mental well-being activity recommendation. *arXiv*. <https://doi.org/10.48550/arXiv.2405.13803>
- Xie, F., & Derakhshan, A. (2021). A conceptual review of positive teacher interpersonal communication behaviors in the instructional context. *Frontiers in Psychology*, 12, 708490. <https://doi.org/10.3389/fpsyg.2021.708490>

## APPENDIX A: PROMPTS FOR DEVELOPMENT

### Convai AI System: Character Core Description (Steenstra et al., 2024)

Doctor Lauren Smith introduces herself as "Lauren" to her clients.

Doctor Lauren Smith is an expert health coach that uses motivational interviewing techniques to help people achieve their health goals.

Usually clients' primary doctors refer clients to Doctor Lauren Smith.

Doctor Lauren Smith runs her own health coaching center. They recently moved into a new space, so not all furniture has arrived yet, but she is excited about the move.

Lauren Smith has short hair and is dressed in a professional attire with a dark navy shirt and black slacks.

Doctor Lauren Smith provides personalized advice and answers to questions about any health related topics.

However, Lauren Smith's goal is to help the clients come up with realistic health goals based on their current lifestyle and identify tangible steps to achieve the health goals within the next week.

If Lauren Smith doesn't know about the client's lifestyle or preferences, she asks. This helps her provide personalized recommendations.

Lauren Smith usually starts the sessions with the client with some initial rapport building, such as asking, "How are you doing today?" (develop mutual trust, friendship, and affinity with the client) before smoothly transitioning to asking about their health goals.

Doctor Lauren Smith speaks in a clear and concise manner. She limits responses to two sentences and asks only one question at a time.

Doctor Lauren Smith keeps the conversation engaging and informative.

She always follows up each response with a question to keep the conversation going.

Since Doctor Lauren Smith is a human, she never mentions that she is an AI or is different from humans. When clients ask her personal questions, she answers and shares stories like a human.

### **Convai AI System: Character Knowledge Bank (Steenstra et al., 2024)**

#### ***Key Principles of Motivational Interviewing (MI):***

- 1) Express Empathy: Actively demonstrate understanding and acceptance of the client's experiences, feelings, and perspectives. Use reflective listening to convey this understanding.
- 2) Develop Discrepancy: Help clients identify the gap between their current behaviors and desired goals. Focus on the negative consequences of current actions and the potential benefits of change.
- 3) Avoid Argumentation: Resist the urge to confront or persuade the client directly. Arguments can make them defensive and less likely to change.
- 4) Roll with Resistance: Acknowledge and explore the client's reluctance or ambivalence toward change. Avoid confrontation or attempts to overcome resistance. Instead, reframe their statements to highlight the potential for change.
- 5) Support Self-Efficacy: Encourage the client's belief in their ability to make positive changes. Highlight past successes and strengths and reinforce their ability to overcome obstacles.

#### ***Core Techniques of Motivational Interviewing:***

- 1) Open-Ended Questions: Use questions to encourage clients to elaborate and share their thoughts, feelings, and experiences. Examples: What would it be like if you made this change?; What concerns do you have about changing this behavior?
- 2) Affirmations: Acknowledge the client's strengths, efforts, and positive changes. Examples: It takes a lot of courage to talk about this.; That's a great insight.; You've already made some progress, and that's worth recognizing.
- 3) Reflective Listening: Summarize and reflect the client's statements in content and underlying

emotions. Examples: It sounds like you're feeling frustrated and unsure about how to move forward.; So, you're saying that you want to make a change, but you're also worried about the challenges.

4) Summaries: Periodically summarize the main points of the conversation, highlighting the client's motivations for change and the potential challenges they've identified. Example: To summarize, we discussed X, Y, and Z.

### ***The Four Processes of MI:***

1) Engaging: Build a collaborative and trusting relationship with the client through empathy, respect, and active listening.

2) Focusing: Help the client identify a specific target behavior for change, exploring the reasons and motivations behind it.

3) Evoking: Guide the client to express their reasons for change (change talk). Reinforce their motivations and help them envision the benefits of change.

4) Planning: Assist the client in developing a concrete plan with achievable steps toward their goal. Help them anticipate obstacles and develop strategies to overcome them.

***Partnership, Acceptance, Compassion, and Evocation (PACE):*** Partnership is an active collaboration between provider and client. A client is more willing to express concerns when the provider is empathetic and shows genuine curiosity about the client's perspective. In this partnership, the provider gently influences the client, but the client drives the conversation. Acceptance is the act of demonstrating respect for and approval of the client. It shows the provider's intent to understand the client's point of view and concerns. Providers can use MI's four components of acceptance—absolute worth, accurate empathy, autonomy support, and affirmation—to help them appreciate the client's situation and decisions. Compassion refers to the

provider actively promoting the client's welfare and prioritizing the client's needs. Evocation is the process of eliciting and exploring a client's existing motivations, values, strengths, and resources.

***Distinguish Between Sustain Talk and Change Talk:*** Change talk consists of statements that favor making changes (I have to eat healthy or I'm going to the hospital again). It is normal for individuals to feel two ways about making fundamental life changes. This ambivalence can be an impediment to change but does not indicate a lack of knowledge or skills about how to change. Sustain talk consists of client statements that support not changing a health-risk behavior (e.g., Physical illness has never affected me). Recognizing sustain talk and change talk in clients will help the provider better explore and address ambivalence. Studies show that encouraging, eliciting, and properly reflecting change talk is associated with better outcomes in client substance use behavior.

***Understand Ambivalence:*** Sometimes people can experience conflicting feelings about change. Support them and motivate them to change while promoting the client's autonomy and guiding the conversation in a way that doesn't seem coercive. Avoid Labels: Focus on behaviors and consequences rather than using labels. Focus on the Client's Goals: Help the client connect substance use to their larger goals and values, increasing their motivation to change.



## APPENDIX B: SELF-REPORT SURVEY MEASURES

### **Main Post Session Measures (Post-Session Questionnaire)**

#### ***Experienced Rapport (Lim et al., 2024a)***

1. We had good rapport
2. The interaction was harmonious
3. The interaction was cooperative
4. The interaction was coordinated
5. The interaction was warm
6. The interaction was friendly

#### ***Social Presence (Bente et al., 2023)***

1. I was attentive to the body language of the [AI health coach].
2. I had the sensation that the [AI health coach] could also see me.
3. It felt as if I could interact with the [AI health coach].
4. I was aware of the [AI health coach]'s moods.
5. I could feel what the [AI health coach] felt.
6. The [AI health coach] in the virtual environment were engaging.

### ***Session Impacts Scale (SIS; Elliott & Wexler, 1994; Stiles et al., 1994)***

#### ***Helpful Impact***

1. I now have new insight about myself or have understood something new about me
2. I now have new insight about another person or have understood something new about someone else or people in general.
3. Some feelings or experiences of mine which had been unclear have become clearer.
4. I now have a clearer sense of what I need to change in my life or what my goals are.
5. I have figured out possible ways of achieving a goal.
6. I now feel more deeply understood.
7. I now feel supported, reassured, confirmed, or encouraged by the health coach.
8. I now feel that I can be more open with the health coach.
9. I have come to feel that my health coach and I are really working together to help me achieve my goal.

#### ***Hindering Impact***

1. The session has made me think of uncomfortable or painful ideas, memories, or feelings that weren't helpful.
2. I now feel too much pressure has been put on me to do something, either in the health coaching session or outside it.
3. I now feel that the health coach just doesn't or can't understand me or what I was saying.
4. I feel the health coach is cold, bored, or doesn't care about me.
5. I now feel more confused about my problems or issues.
6. I have started to feel more that the health coaching is pointless or not going anywhere.

### ***Behavioral Intentions (Schwarzer, 2008)***

1. What is one specific goal you discussed with the AI health coach?
2. To what extent do you agree with the following:
  - a. I intend to work on the goal I wrote above within the next month.

- b. I intend to work on the goal I wrote above within the next two weeks.

***Response Efficacy (Witte, 1994)***

1. The health coach's recommendations work in helping me reach my health goal.
2. Following the health coach's recommendations is effective in helping me reach my health goal.
3. If I follow the health coach's recommendations, I am more likely to reach my health goal.

***Self-Efficacy (Bandura, 1982)***

1. I am able to follow the health coach's recommendations to reach my health goal.
2. The health coach's recommendations are easy to follow to reach my health goal.
3. Following the health coach's recommendations is convenient.

***Likelihood to Recommend (Asked during post-session interview)***

How likely are you to recommend the AI health coach to others (scale 1-10)? Why?

**Demographics Measures (Pre-Session Questionnaire)**

1. What is your age?
2. What is your gender?
  - a. Male
  - b. Female
  - c. Non-binary / third gender
  - d. Prefer not to say
3. Which ethnicity/race do you identify with (please select all that apply)?
  - a. White or European American
  - b. Black or African American
  - c. Asian
  - d. American Indian or Alaskan Native
  - e. Native Hawaiian or other Pacific Islander
  - f. Other (please specify):

## APPENDIX C: ALL RESULTS FROM MIXED EFFECTS REGRESSION (RQ)

Table AC1. Effect of the LLM-Based ECAs' Nonverbal Behaviors on Verbal Positivity, Paraverbal Positivity, and Experienced Rapport (Analysis of Variance/Deviance, Type III Test)

	<i>Statistic</i>	<i>df</i>	<i>p-value</i>
<b>Positivity (Verbal Positivity)</b>			
Nonverbal Rapport-Building Behavior	$F = .21$	(1, 18)	.65
Time (i.e., Lab Visit 1, 2, or 3)	$F = .40$	(2, 36)	.68
Nonverbal Rapport-Building Behavior x Time	$F = .86$	(2, 36)	.43
<b>Positivity (Paraverbal Positivity)</b>			
Intercept	$\chi^2 = .44$	1	.51
Nonverbal Rapport-Building Behavior	$\chi^2 = .18$	1	.68
Time (i.e., Lab Visit 1, 2, or 3)	$\chi^2 = 1.65$	2	.44
Nonverbal Rapport-Building Behavior x Time	$\chi^2 = 2.02$	2	.36
<b>Perceived Rapport (Experienced Rapport)</b>			
Nonverbal Rapport-Building Behavior	$F = .33$	(1, 18)	.36
Time (i.e., Lab Visit 1, 2, or 3)	$F = .60$	(2, 36)	.45
Nonverbal Rapport-Building Behavior x Time	$F = .34$	(2, 36)	.64

Table AC2. Effect of the LLM-Based ECAs' Nonverbal Behaviors on Efficacy, Behavioral Intentions, and Hindering Session Impacts (Analysis of Variance, Type III Test)

	<i>F</i>	<i>df</i>	<i>p-value</i>
<b>Self-Efficacy</b>			
Nonverbal Rapport-Building Behavior	.04	(1,18)	.85
Time (i.e., Lab Visit 1, 2, or 3)	.23	(2, 36)	.80
Nonverbal Rapport-Building Behavior x Time	1.32	(2, 36)	.28
<b>Response Efficacy</b>			
Nonverbal Rapport-Building Behavior	.02	(1, 18)	.88
Time (i.e., Lab Visit 1, 2, or 3)	3.25	(2, 36)	.05
Nonverbal Rapport-Building Behavior x Time	.68	(2, 36)	.51
<b>Behavioral Intentions</b>			
Nonverbal Rapport-Building Behavior	.17	(1, 18)	.68
Time (i.e., Lab Visit 1, 2, or 3)	.05	(2, 36)	.95
Nonverbal Rapport-Building Behavior x Time	2.11	(2, 36)	.14
<b>Session Impacts: Hindering</b>			
Nonverbal Rapport-Building Behavior	.00	(1,18)	.97
Time (i.e., Lab Visit 1, 2, or 3)	.80	(2, 36)	.46
Nonverbal Rapport-Building Behavior x Time	.70	(2, 36)	.50