

Please quote as: Meywirth, S., Janson, A. & Söllner, M. (2025). Designing for Trust: Integrating Self-referencing in Large Language Model-Based Health Coaching. In S. Chatterjee, J. vom Brocke & R. Anderson (eds.), Local Solutions for Global Challenges. DESRIST 2025. Lecture Notes in Computer Science (p./pp. 296–309), Cham: Springer. ISBN: 978-3-031-93975-4



Designing for Trust: Integrating Self-referencing in Large Language Model-Based Health Coaching

Sophia Meywirth¹ (✉) , Andreas Janson² , and Matthias Söllner¹

¹ University of Kassel, Kassel, Germany
{sophia.meywirth, soellner}@uni-kassel.de

² University of St. Gallen, St. Gallen, Switzerland
andreas.janson@unisg.ch

Abstract. This research applies the Design Science Research (DSR) methodology to investigate how self-referencing in Large Language Model (LLM)-based health coaching influences user trust and perceptions of anthropomorphism. We synthesized theory-driven design principles to guide the integration of self-referencing and demonstrated them in a vignette-based prototype. Through a single-factorial between-subjects experiment, analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM) and qualitative feedback, we identified a dual effect of self-referencing: while professional self-referencing enhances trust via increased anthropomorphism, overly personal references can directly undermine trust. Based on these findings, we refined our design principles to optimize trust-building in LLM-based coaching. Our contributions provide actionable design guidelines for creating more effective and trustworthy AI-driven health interventions, advancing the understanding of anthropomorphic design in digital coaching contexts.

Keywords: Health behavior change · Coaching · Large language models · Trust · Anthropomorphism

1 Introduction

The health and wellness industry has experienced significant growth in recent years, driven by a rising demand [1] from consumers who prioritize health, fitness, nutrition, and overall well-being. This trend is particularly influenced by a cultural shift towards preventive healthcare, largely advocated by younger generations like Gen Z. They favor proactive health management strategies, emphasizing dietary and fitness modifications to prevent chronic diseases [2, 3]. While the intent to adopt healthier lifestyles is evident, actual behavioral change remains challenging.

Digital interventions, particularly health coaching, have emerged as effective means to support individuals in achieving lifestyle changes through collaborative conversations and personalized guidance [4]. However, the scalability of such interventions is often

limited by the availability of human practitioners and the delayed interactions typical of traditional digital formats, which can diminish engagement [5–7].

With the recent advancements in large language models (LLMs), new possibilities have emerged. These technologies offer the potential for real-time, scalable, and highly personalized coaching interactions, mimicking human conversational capabilities and adapting responses to individual user needs. [6, 8] Preliminary applications of LLMs in fields like nutrition and physical activity coaching have demonstrated promising outcomes by providing dynamic, customized advice [9–12].

To fully leverage the benefits of LLMs in behavioral healthcare, their proper utilization is crucial [13]. Central to successful coaching is the establishment of user trust [8, 14]. A trusting relationship between the coach and coachee is essential, as it encourages users to disclose necessary health information [8] and adhere to the coach's advice [15]. In LLM-based health coaching, anthropomorphism – the extent to which users perceive human-like qualities in the system – can significantly influence trust. Studies suggest that incorporating human-like features in conversational agents makes interactions more natural and relatable, thereby fostering trust [16]. Precisely, anthropomorphic design elements, such as human-like behaviors and verbal social cues, can shape how trustworthy users perceive the system to be.

While prior research on anthropomorphic chatbot design has primarily focused on visual cues, such as the use of avatars and names, and stylistic elements of communication, such as tone and formality [17, 18], there has been limited exploration on how to design content-specific verbal cues in chatbot interactions. Content-specific cues – what is said rather than just how it is said – are particularly impactful in contexts that require ongoing engagement and rapport, such as health coaching [19]. In these settings, meaningful conversation beyond task-oriented dialogue can foster a supportive and trusted relationship between the user and the system [20]. One such underexplored content-specific cue is self-referencing, where the agent refers to itself to appear more human-like [21].

Given the complex role self-referencing plays in user perceptions of AI coaches, this paper investigates the nuanced design impacts of this anthropomorphic trait. We aim to explore the following design-based research question:

RQ: How should an LLM-based coaching intervention be designed to integrate self-referencing to foster trust?

To address our research question, we adopted a design science research methodology as outlined by Peffers et al. [22]. Our study aimed to establish initial design principles using a theory-driven and exploratory approach. We assessed these principles through a single-factorial between-subjects experiment, specifically examining the impact of self-referencing on perceptions of anthropomorphism and trust within LLM-based health coaching contexts. We employed partial least squares structural equation modeling (PLS-SEM) to analyze the contributions and effects of these human-like cues. Additionally, we integrated qualitative data from the experiments to further refine our findings and the evolving design principles. This comprehensive approach allowed us to explore both the positive and negative implications of self-referencing in enhancing the efficacy of digital health interventions.

2 Theoretical Background

2.1 Anthropomorphic Design

Anthropomorphism involves attributing human traits and characteristics to non-human entities like computers or systems [23–25]. Nass and Moon [25] introduced the “Computers are Social Actors” (CASA) paradigm, explaining that despite their lack of emotions, self-reference, and physical presence, people still apply social roles and expectations to computers, treating them as social beings. This tendency has been examined through two main theoretical lenses: one suggests humans automatically respond to lifelike or social cues without conscious thought, while the other proposes that people form mental models based on how systems mimic human behavior, influencing their interactions with these systems [24, 26].

Advances in AI since 2000 have significantly enhanced the capabilities of computers, enabling them to simulate emotions, use personal pronouns, and exhibit human-like behaviors [25]. This has transformed user interactions, particularly in the development of conversational agents that now incorporate a broad range of social cues, enhancing user engagement through more natural interactions [26, 27]. Feine et al. [21] detailed this in their taxonomy of social cues for conversational agents, highlighting differences in verbal, visual, auditory, and invisible cues.

Previous research has predominantly focused on the visual and stylistic aspects of anthropomorphic chatbot design, such as avatars, names, tone, and formality [17, 18]. However, there has been limited exploration of content-specific verbal cues, which are crucial in settings requiring sustained engagement and rapport, like health coaching [19]. These cues, focusing on what is communicated rather than just how it is communicated, can significantly foster a supportive and trusted relationship between users and systems [20].

2.2 Trust

In IS research, trust is often studied in two contexts: interpersonal trust, applicable to both individuals and organizations and trust in IT systems or artifacts [28–30]. Trust formation relies on perceived characteristics such as ability, benevolence, and integrity for individuals, and functionality, helpfulness, and reliability for technology [31–33]. In health coaching, trust hinges on the coach’s perceived competence, goodwill, and reliability, which are crucial for client engagement [8].

We adopt the definition of trust proposed by Mayer et al. ([35], page 712): “Trust is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.” This definition emphasizes the critical role of vulnerability in trust dynamics, where individuals willingly expose themselves to risk in relationships [35, 36]. Designing IT artifacts that foster trust requires a systematic approach grounded in trust theory [37], ensuring that key trust-building factors are embedded the system’s functionality. In LLM-based health coaching, one such approach is anthropomorphic design, which can enhance user perceptions of the system’s trust antecedents [16].

In LLM-based health coaching, the extent to which users perceive human-like qualities in the system, or anthropomorphism, significantly influences trust. Studies indicate that incorporating human-like features into conversational agents makes interactions more natural and relatable, enhancing trust [16]. Specifically, anthropomorphic design elements like human-like behaviors and verbal social cues shape user perceptions of trustworthiness. Supporting research shows that users form richer mental models of anthropomorphic systems than of mechanical ones, suggesting that these effects also apply to digital systems [24]. Moreover, trust in conversational agents is notably improved by anthropomorphic features, including visual and auditory cues [38].

3 Methodology

This research follows the Design Science Research (DSR) methodology as described by Peffers et al. [22], which is structured into six distinct phases depicted in Fig. 1. We followed the problem-centered approach, where the initial step involves identifying the problem and establishing motivation – details of which are presented in the introduction of this paper.

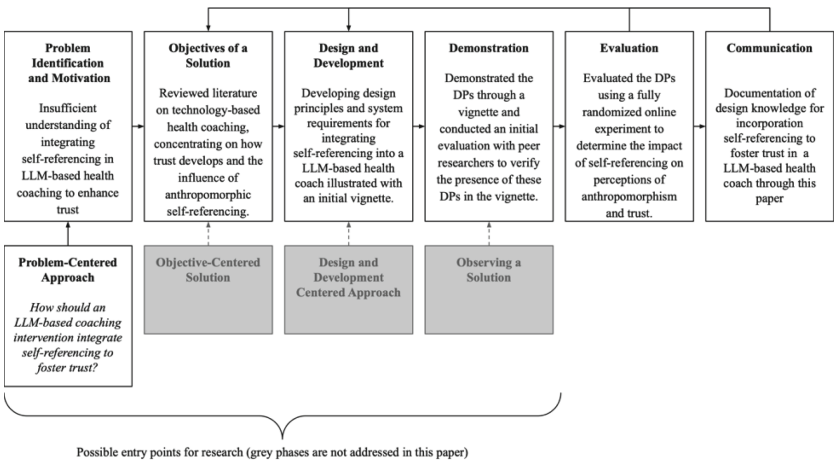


Fig. 1. Design Science Research Process adapted from Peffers et al. [22]

The second phase focuses on specifying the objectives of the solution. The primary goal of our solution is to integrate the content-specific verbal social cue of self-referencing into an LLM-based health coaching intervention to enhance users’ perception of anthropomorphism and trust. More broadly, self-referencing contributes to anthropomorphic design by making the conversational agent appear more human-like. Sharing personal experiences and demonstrating behaviors are inherently human traits that foster personalization, build rapport, and strengthen trust. Feine et al. highlight content-specific social cues, such as self-disclosure and references to past events, as key factors in enhancing user engagement [21]. By incorporating relevant self-references,

the LLM-based coach can create more relatable and engaging interactions, ultimately improving the overall coaching experience.

In the third phase, Design and Development, we adopt a theory-driven approach, deriving design principles from the literature on anthropomorphic chatbot design and health coaching. After establishing initial design principles based on the guidelines of Gregor et al. [39], we developed a vignette-based prototype. This prototype serves to demonstrate and evaluate our design principles in phases five and six, respectively, through a fully randomized online experiment.

The research cycle concludes with the refinement of our design principles, incorporating insights gained from the experimental results and additional qualitative user feedback. This iterative process ensures that our findings are both robust and relevant to the field of digital health coaching.

4 Design and Development

4.1 Deriving Initial Design Principles from Theory

Self-referencing is a content-specific social cue with multiple potential benefits in LLM-based health coaching. Coaching inherently involves clients disclosing personal information about their goals and habits, often on intimate topics, particularly in health-related discussions [8]. However, individuals are generally hesitant to share such information unless reciprocity is present. Research has demonstrated that users are more willing to disclose personal details when their conversational partner also shares intimate information [40]. Moon's study [40], grounded in the CASA paradigm, shows that this reciprocity effect extends to human-computer interactions. Similarly, Lee et al. found that greater self-disclosure by a chatbot led to increased self-disclosure from users, as well as higher enjoyment and intimacy in the interaction [41]. We therefore propose the following design principle:

DPI: For designers and researchers to design LLM-based health behavior change interventions effectively for individuals desiring to change their health behaviors, ensure that the LLM-based coach engages in self-referencing by sharing relevant information about itself in a way that fosters reciprocity, so that user openness is encouraged, thereby strengthening trust and engagement.

Beyond fostering openness, self-referencing aligns with established behavior change techniques (BCTs). The taxonomy by Michie et al. [42] identifies 93 distinct BCTs, including the "comparison of behavior" cluster, which highlights the role of demonstrating behaviors and fostering social comparison. Self-referencing allows the LLM-based coach to position itself as a role model, verbally demonstrating behaviors that clients might adopt. For example, the coach could reference its own strategies for maintaining habits or overcoming challenges, encouraging clients to reflect on their own behaviors and implement changes accordingly. We therefore propose the following design principle:

DP2: For designers and researchers to design LLM-based health behavior change interventions effectively for individuals desiring to change their health behaviors, ensure that the LLM-based coach verbally demonstrates desirable behaviors, so that social comparison is facilitated and users are encouraged to adopt healthier habits.

Furthermore, self-referencing plays a crucial role in fostering trust. When assessing trustworthiness, people evaluate key characteristics such as ability, benevolence, and integrity [34]. By referencing its own coaching-relevant experiences, the LLM-based coach can establish its ability and credibility, demonstrating expertise in guiding users toward their health goals. We therefore propose the following design principle:

Table 1. Exemplary Dialogue Elements Based on Initial Design Principles

Design Principles		Implementation in Vignette
For designers and researchers to design LLM-based health behavior change interventions effectively for individuals desiring to change their health behaviors, ...		Embedded within conversation between LLM-based coach and client aimed at fostering healthy nutrition and fitness.
DP1	ensure that the LLM-based coach engages in self-referencing by sharing information about itself in a way that fosters reciprocity, so that the user’s openness is encouraged thereby enhancing trust and engagement.	“A bit about me — I’ve always been passionate about helping others live their healthiest lives, inspired by my own wellness journey.”
		“I used to do homemade pizza nights with friends to see who could make the healthiest, tastiest toppings.”
		“Wednesday: Rest day — I usually use this to catch up on reading or plan the week. “
DP2	ensure that the LLM-based coach verbally demonstrates desirable behaviors, so that social comparison is reinforced and users are encouraged to adopt healthier habits.	“For lunch, how about switching to wraps with lean meats and veggies instead of the usual sandwiches? It’s an easy, tasty swap I love for eating better on the go.”
		“Here’s a plan for the week that includes some of my personal favorites for inspiration: Monday & Thursday: 15-minute morning walk. — starting the day with a walk really energizes me, and it might do the same for you.”
DP3	ensure that the LLM-based coach utilizes self-referencing to demonstrate its expertise and reliability, so that the user feels confident in the coach’s abilities and trusts the guidance provided, thereby fostering a committed and effective coaching relation-ship.	„I have other clients who are vegan, so I’m quite familiar with a variety of dishes that could work well for you.”
		„For example, a client of mine also had a desk job and found success by scheduling short, 10 minute walks during her lunch break.”

DP3: For designers and researchers to design LLM-based health behavior change interventions effectively for individuals desiring to change their health behaviors, ensure that the LLM-based coach utilizes self-referencing to demonstrate its expertise and reliability, so that the user feels confident in the coach's abilities and trusts the guidance provided, thereby fostering a committed and effective coaching relationship.

4.2 Development of Vignettes and Demonstration

We developed a vignette-based LLM health coaching intervention designed to simulate trust-building interactions in a realistic setting [43] and to demonstrate our initial design principles. The vignette represents an initial chat-based session with a lifestyle coach focused on goal-setting – an essential component of health coaching [6]. To reinforce this, we incorporated techniques from the “goals and planning” cluster of the behavior change technique taxonomy [42] and integrated dietary advice and strategic recommendations. These elements address key trust factors, such as sharing personal goals [8] and following expert advice [44].

Building on design principles derived from theory, we crafted a prototype dialogue demonstrating DP1, DP2, and DP3. Table 1 outlines these principles alongside the corresponding self-reference phrases embedded in the vignettes. The dialogue includes self-references aligned with our initial design knowledge, while a control vignette, omitting these self-references, was developed for baseline comparison. Each interaction began with an introductory message clarifying that the coach was powered by an LLM.

These messages were implemented in a Figma prototype designed to replicate the mobile version of ChatGPT, ensuring a familiar user experience while avoiding additional social cues beyond those typical of LLM-based interactions.

We conducted a preliminary evaluation of the vignettes with five peer researchers to ensure they effectively demonstrated our design principles. Based on their feedback, we made slight adjustments to refine clarity and alignment with the intended design elements.

5 Evaluation

5.1 Experimental Procedure

We conducted a fully randomized online experiment to assess the impact of our design principles on user trust and perceptions of anthropomorphism. Our evaluation followed the DSR evaluation framework proposed by Venable [45], adopting a formative intermediate approach situated between ex-ante and ex-post evaluation. The experiment was well suited for this purpose, as it focused on relational mechanisms – such as trust, anthropomorphism, and reliance – that are central to the intended effects of the proposed design principles.

A total of 149 participants were recruited via Prolific [46] in accordance with guidelines for online studies [47]. After excluding six responses for failing attention checks and three responses due to excessively rapid survey completion (under five minutes), the final sample comprised 140 valid responses, with 74 participants in the control group and 66 in the treatment group.

The sample consisted of slightly more males (49.3%) than females (47.9%), with a small percentage of non-binary participants (2.9%). Participants had a mean age of 38 years (ranging from 18 to 74 years). In terms of education, 43.6% held a bachelor's degree as their highest qualification, while 29.3% had completed a high school diploma. Moreover, 63.6% reported no prior experience with health or lifestyle coaching, and 80% aimed to improve their fitness and nutrition.

The experiment unfolded in three steps. First, participants received a brief overview of the study's objectives and context. Second, they were introduced to a vignette scenario in which they assumed the role of an individual receiving LLM-based coaching for healthy nutrition and physical activity – initiating goal-setting via an online chat. Finally, participants were presented with the prototype vignette depicting the initial coaching session and were asked to evaluate its content through an online survey.

Perceived anthropomorphism was measured using seven indicators adapted from Seymour et al. [48], reliance was assessed with three indicators adopted from Chua et al. [49] and Gursoy et al. [50], and trust was evaluated using three indicators adapted from Gefen [51]. Additionally, open-ended questions were included to capture exploratory insights during this initial iteration.

Afterward, we used structural equation modeling (SEM) with the variance-based partial least squares (PLS) [52, 53] approach to analyze our data. PLS-SEM is a multivariate analysis technique widely used to investigate causal-predictive relationships in information systems. It does this by employing weighted composites of variables as proxies for latent variables, which allows for the estimation of their relationship. [54]

5.2 Results

We started by evaluating our measurement model. We checked the univariate normality of the measurements and found that none of our measures exceeded the threshold of 2 or -2 for skewness or kurtosis values of 2 or -2 [55]. Subsequently, we assessed the construct's reliability and convergent validity. We employed Cronbach's Alpha, composite reliability (ρ_c and ρ_a), and average variance extracted (AVE) as indicators. Our findings indicated that all indicators met the thresholds of 0.7 or above for Cronbach's alpha, ρ_c , and ρ_a , and 0.5 or above for AVE [55]. Furthermore, we evaluated the discriminant validity by measuring the heterotrait-monotrait (HTMT) ratios of correlations. The discriminant validity measures to what extent the latent constructs are truly distinct from each other [56]. None of our latent constructs exhibited a value above the suggested threshold of 0.9 [55].

The results of the structural model are presented in detail using path coefficients (β), significance levels (p -values), and effect sizes (f^2). A 5,000-sample bootstrapping algorithm was used to determine the statistical significance of the path coefficients. The results are summarized in Fig. 2.

Our results show a significant relationship with a p -value less than 0.01 between anthropomorphism and trust (p -value = 0.000). Additionally, we found a significant relationship between self-referencing and anthropomorphism with a p -value of 0.003. Furthermore, we identified a marginally significant relationship between self-referencing and trust with a p -value of 0.086. Moreover, we revealed a significant relationship between trust and reliance.

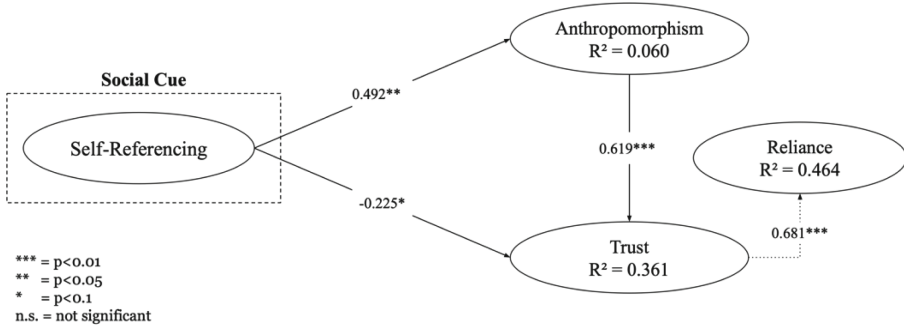


Fig. 2. Structural Model

In terms of effect size, values above 0.020, 0.150, and 0.350 indicate a low, moderate, and high effect at the structural level, respectively [57]. The analysis confirmed a significant and positive relationship between anthropomorphism and trust, as indicated by a strong path coefficient ($\beta = 0.619$, $p \leq 0.01$) and a high effect size ($f^2 = 0.563$). Similarly, the relationship between self-referencing and anthropomorphism was significant and positive, with a moderate path coefficient ($\beta = 0.492$, $p \leq 0.05$) and a low effect size ($f^2 = 0.064$). In contrast, the relationship between self-referencing and trust was significant but negative, with a path coefficient ($\beta = -0.225$, $p \leq 0.1$) and a marginal effect size ($f^2 = 0.019$). Moreover, the analysis identified a significant relationship between trust and reliance with a strong path coefficient ($\beta = 0.681$, $p = 0.000$) and a substantial explained effect size ($f^2 = 0.866$). The R^2 values for trust ($R^2 = 0.361$) and reliance ($R^2 = 0.464$) are considered substantial based on Chin’s [52] guidelines.

6 Discussion and Revision of Design Principles

Our results reveal a dual effect of self-referencing in digital coaching applications. While it enhances anthropomorphism and can foster trust, it may also undermine trust if not carefully calibrated. Participants responded negatively to experiential self-references – such as anecdotes about pizza nights or personal routines – often perceiving them as unsettling or inauthentic. One participant noted, “It also wrote as though it was alive and had friends and ate things, which is weird.” Another remarked, “I did NOT like the LLM coach pretending to be human. An LLM doesn’t have pizza nights, take morning walks, and spend Wednesday breaks reading.”

These reactions highlight a key insight: overly human-like self-references can compromise the system’s perceived integrity – a cornerstone of trust [34]. In contrast, professional self-referencing, such as references to experiences with other clients, was perceived as helpful and credible. Since self-referencing increases anthropomorphism – positively affecting trust –, we recommend incorporating it at a professional level while integrating emotions and empathy. These elements were not perceived as uncanny and contributed to a more human-like yet credible interaction.

In light of these findings, we revised our original design principles to reflect the nuanced role of self-referencing in building trust. Specifically, we recommend avoiding

fabricated personal anecdotes and instead suggest referencing generalized client experiences and providing empathetic responses that mirror user emotions without implying human-like experiences.

DP1, which originally focused on fostering openness through self-referencing, is now refined to emphasize professional reciprocity rather than fabricated personal stories. Empathy remains important but should be conveyed in ways that do not imply human identity.

DP2 emphasized self-referencing as a role-modeling strategy, aligning with behavior change techniques such as social comparison. However, user feedback suggests this should not include personal anecdotes or imply human-like experiences. Instead, LLM-based coaches should reference other clients’ experiences to encourage reflection while maintaining an AI-appropriate persona.

Based on our evaluation, we find no need to revise DP3, which focuses on establishing the coach’s ability through references to professional experiences. However, we propose adding DP4, which underscores the importance of avoiding experiential self-references in LLM-based coaching. Table 2 provides an overview of the revised design principles.

Table 2. Revised Design Principles

Revised Design Principles	
For designers and researchers to design LLM-based health behavior change interventions effectively for individuals desiring to change their health behaviors, ...	
DP1	...ensure that the LLM-based coach engages in self-referencing by sharing generalized insights and experiences from other clients in a professional manner and providing empathetic responses that mirror users’ emotions without implying personal experiences, so that users feel a sense of commonality and reciprocity, encouraging openness and fostering trust and engagement.
DP2	...ensure that the LLM-based coach verbally demonstrates desirable behaviors by referencing other clients’ strategies, so that social comparison is reinforced and users are encouraged to adopt healthier habits.
DP3	...ensure that the LLM-based coach utilizes self-referencing to demonstrate its expertise and reliability, so that the user feels confident in the coach’s abilities and trusts the guidance provided, thereby fostering a committed and effective coaching relationship.
DP4	...ensure that the LLM-based coach avoids experiential self-references, such as sharing human-like personal experiences, so that users accept the system and do not experience feelings of eeriness.

7 Conclusion

In our study, we synthesized three theory-driven design principles for integrating self-referencing in LLM-based health coaching to foster trust. We demonstrated these principles through a vignette-based prototype, which we evaluated in a single-factorial

between-subjects experiment. Using PLS-SEM analysis and additional qualitative data, we identified a dual effect of self-referencing on trust and reliance in digital coaching applications: while self-referencing can enhance trust by increasing perceived anthropomorphism, it can also directly undermine trust when not carefully calibrated. Based on these findings, we refined our design principles to optimize the use of self-referencing in LLM-based health coaching for trust-building.

While our study provides valuable insights, it is important to acknowledge the limitations inherent in the iterative nature of the Design Science Research methodology. This study represents the first iteration, where evaluation results informed the revision of our design principles, which should be further tested in subsequent iterations. Future research should extend beyond vignette-based prototypes to implement and evaluate the revised design principles in real-world settings. Additionally, further studies should assess the intervention's effectiveness in collaboration with behavioral healthcare experts, such as therapists, certified dietitians, and researchers, to validate its practical applicability. Finally, future research must also address potential biases inherent in the natural training data of LLMs to ensure ethical and unbiased coaching interactions.

References

1. Statista: Digital Health - Worldwide, <https://www.statista.com/outlook/hmo/digital-health/worldwide>, last accessed 16 Apr 2024
2. Park, M.J., Kim, D.J., Lee, U., Na, E.J., Jeon, H.J.: A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. *Front. Psych.* **10** (2019). <https://doi.org/10.3389/fpsy.2019.00505>
3. OECD/European Union: Health at a glance: Europe 2022: state of health in the EU cycle. OECD (2022). <https://doi.org/10.1787/507433b0-en>
4. Huffman, M.: Health coaching: a new and exciting technique to enhance patient self-management and improve outcomes. *Home Healthc. Now.* **25**, 271 (2007). <https://doi.org/10.1097/01.NHH.0000267287.84952.8f>
5. Mitchell, E.G., Elhadad, N., Mamykina, L.: Examining AI methods for micro-coaching dialogs. In: CHI Conference on Human Factors in Computing Systems, pp. 1–24. ACM, New Orleans, LA (2022). <https://doi.org/10.1145/3491102.3501886>
6. Mitchell, E.G., Maimone, R., Cassells, A., Tobin, J.N., Davidson, P., Smaldone, A.M., Mamykina, L.: Automated vs.: Human health coaching: exploring participant and practitioner experiences. *Proc. ACM Hum.-Comput. Interact.* **5**, 99 (2021). <https://doi.org/10.1145/3449173>
7. Brandt, C.J., Søgaaard, G.I., Clemensen, J., Søndergaard, J., Nielsen, J.B.: Determinants of successful eHealth coaching for consumer lifestyle changes: qualitative interview study among health care professionals. *J. Med. Internet Res.* **20**, e237 (2018). <https://doi.org/10.2196/jmir.9791>
8. Schiemann, S.J., Mühlberger, C., Schoorman, F.D., Jonas, E.: Trust me, I am a caring coach: the benefits of establishing trustworthiness during coaching by communicating benevolence. *J. Trust Res.* **9**, 164–184 (2019). <https://doi.org/10.1080/21515581.2019.1650751>
9. Alanezi, F.: Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. *Nutr. Health*, 02601060241244563 (2024). <https://doi.org/10.1177/02601060241244563>
10. Arslan, S.: Exploring the potential of chat GPT in personalized obesity treatment. *Ann. Biomed. Eng.* **51**, 1887–1888 (2023). <https://doi.org/10.1007/s10439-023-03227-9>

11. Kirk, D., van Eijnatten, E., Camps, G.: Comparison of answers between ChatGPT and human dieticians to common nutrition questions. *J. Nutrit. Metabol.* **2023**, e5548684 (2023). <https://doi.org/10.1155/2023/5548684>
12. Willms, A., Liu, S.: Exploring the feasibility of using ChatGPT to create just-in-time adaptive physical activity mHealth intervention content: case study. *JMIR Med. Educ.* **10**, e51426 (2024). <https://doi.org/10.2196/51426>
13. Demsky, D., Yang, D., Yeager, D.S., Bryan, C.J., Clapper, M., Chandhok, S., Eichstaedt, J.C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., JonesMitchell, N., Ong, D.C., Dweck, C.S., Gross, J.J., Pennebaker, J.W.: Using large language models in psychology. *Nat Rev Psychol.*, 1–14 (2023). <https://doi.org/10.1038/s44159-023-00241-5>
14. Terblanche, N.H.D., Heyns, M.: The impact of coachee personality traits, propensity to trust and perceived trustworthiness of a coach, on a coachee's trust behaviour in a coaching relationship. *SA J. Ind. Psychol.* **46**, 1–11 (2020). <https://doi.org/10.4102/sajip.v46i0.1707>
15. Wu, Z., Helaoui, R., Reforgiato Recupero, D., Riboni, D.: Towards effective automatic evaluation of generated reflections for motivational interviewing. In: Companion Publication of the 25th International Conference on Multimodal Interaction, pp. 368–373. Association for Computing Machinery, New York, NY (2023). <https://doi.org/10.1145/3610661.3616127>
16. de Visser, E., Monfort, S., Mckendrick, R., Smith, M., Mcknight, P., Krueger, F., Parasuraman, R.: Almost human: anthropomorphism increases trust resilience in cognitive agents. *J. Exp. Psychol. Appl.* **22** (2016). <https://doi.org/10.1037/xap0000092>
17. Janson, A.: How to leverage anthropomorphism for chatbot service interfaces: the interplay of communication style and personification. *Comput. Hum. Behav.* **149**, 107954 (2023). <https://doi.org/10.1016/j.chb.2023.107954>
18. Fadhil, A., Schiavo, G., Wang, Y., Yilma, B.A.: The Effect of emojis when interacting with conversational interface assisted health coaching system. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 378–383. Association for Computing Machinery, New York (2018). <https://doi.org/10.1145/3240925.3240965>
19. Jörke, M., Sapkota, S., Warkenthien, L., Vainio, N., Schmiedmayer, P., Brunskill, E., Landay, J.: Supporting Physical Activity Behavior Change with LLM-Based Conversational Agents, <http://arxiv.org/abs/2405.06061> (2024). <https://doi.org/10.48550/arXiv.2405.06061>
20. Ryan, K., Dockray, S., Linehan, C.: Understanding how eHealth coaches tailor support for weight loss: towards the design of Person-centered coaching systems. In: CHI Conference on Human Factors in Computing Systems, pp. 1–16. ACM, New Orleans, LA (2022). <https://doi.org/10.1145/3491102.3501864>
21. Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A taxonomy of social cues for conversational agents. *Int. J. Human-Comput. Stud.* **132**, 138–161 (2019). <https://doi.org/10.1016/j.ijhcs.2019.07.009>
22. Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A design science research methodology for information systems research. *J. Manag. Inf. Syst.* **24**, 45–77 (2007)
23. Epley, N., Waytz, A., Cacioppo, J.T.: On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* **114**, 864–886 (2007). <https://doi.org/10.1037/0033-295X.114.4.864>
24. Fink, J.: Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: Ge, S.S., Khatib, O., Cabibihan, J.-J., Simmons, R., Williams, M.-A. (eds.) *Social Robotics*, pp. 199–208. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34103-8_20
25. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. *J. Soc. Issues.* **56**, 81–103 (2000). <https://doi.org/10.1111/0022-4537.00153>

26. Lee, S.-L., Lau, I., Kiesler, S., Chiu, C.Y.: Human Mental Models of Humanoid Robots. (2005). <https://doi.org/10.1109/ROBOT.2005.1570532>
27. Stein, J.-P., MacDorman, K.F.: After confronting one uncanny valley, another awaits. *Nat. Rev. Electr. Eng.* **1**, 276–277 (2024). <https://doi.org/10.1038/s44287-024-00041-w>
28. Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., Leimeister, J.M.: Understanding the formation of trust in IT artifacts. In: *International Conference on Information Systems (ICIS)* 2012, Orlando, FL (2012)
29. Söllner, M., Hoffmann, A., Leimeister, J.M.: Why different trust relationships matter for information systems users. *Eur. J. Inf. Syst.* **25**, 274–287 (2016). <https://doi.org/10.1057/ejis.2015.17>
30. Thiebes, S., Lins, S., Sunyaev, A.: Trustworthy artificial intelligence. *Electron Markets*. **31**, 447–464 (2021). <https://doi.org/10.1007/s12525-020-00441-4>
31. McKnight, D.H., Carter, M., Thatcher, J.B., Clay, P.F.: Trust in a specific technology: an investigation of its components and measures. *ACM Trans. Manag. Inf. Syst.* **2**, 1–25 (2011). <https://doi.org/10.1145/1985347.1985353>
32. Söllner, M., Pavlou, P.A., Leimeister, J.M.: Understanding Trust in IT Artifacts – A New Conceptual Approach, <https://papers.ssrn.com/abstract=2475382> (2013). <https://doi.org/10.2139/ssrn.2475382>
33. Thatcher, J.B., McKnight, D.H., Baker, E.W., Arsal, R.E., Roberts, N.H.: The role of trust in Postadoption IT exploration: an empirical examination of knowledge management systems. *IEEE Trans. Eng. Manag.* **58**, 56–70 (2011). <https://doi.org/10.1109/TEM.2009.2028320>
34. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. *Acad. Manag. Rev.* **20**, 709–734 (1995). <https://doi.org/10.2307/258792>
35. Rousseau, D., Sitkin, S., Burt, R., Camerer, C.: Not so different after all: a cross-discipline view of trust. *Acad. Manag. Rev.* **23** (1998). <https://doi.org/10.5465/AMR.1998.926617>
36. Lee, J.D., See, K.A.: Trust in Automation: designing for appropriate reliance. *Hum. Factors*. (2004)
37. Hoffmann, H., Söllner, M.: Incorporating behavioral trust theory into system development for ubiquitous applications. *Pers. Ubiquit. Comput.* **18**, 117–128 (2014). <https://doi.org/10.1007/s00779-012-0631-1>
38. Rheu, M., Shin, J.Y., Peng, W., Huh-Yoo, J.: Systematic review: trust-building factors and implications for conversational agent design. *Int. J. Human-Comp. Interact.* **37**, 1–16 (2020). <https://doi.org/10.1080/10447318.2020.1807710>
39. Gregor, S., Chandra Kruse, L., Seidel, S.: The anatomy of a design principle. *J. Assoc. Inf. Syst.* **21**, 1622–1652 (2020). <https://doi.org/10.17705/1jais.00649>
40. Moon, Y.: Intimate exchanges: using computers to elicit self-disclosure from consumers. *J. Consum. Res.* **26**, 323–339 (2000). <https://doi.org/10.1086/209566>
41. Lee, Y.-C., Yamashita, N., Huang, Y., Fu, W.: “I Hear You, I Feel You”: encouraging deep self-disclosure through a Chatbot. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. Association for Computing Machinery, New York (2020). <https://doi.org/10.1145/3313831.3376175>
42. Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M., Cane, J., Wood, C.: The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann. Behav. Med. Publ. Soc. Behav. Med.* **46** (2013). <https://doi.org/10.1007/s12160-013-9486-6>
43. Aguinis, H., Bradley, K.J.: Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organ. Res. Methods.* **17**, 351–371 (2014). <https://doi.org/10.1177/1094428114547952>

44. Wu, D., Lowry, P.B., Zhang, D., Tao, Y.: Patient Trust in Physicians Matters—Understanding the role of a Mobile patient education system and patient-physician communication in improving patient adherence behavior: field study. *J. Med. Internet Res.* **24**, e42941 (2022). <https://doi.org/10.2196/42941>
45. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: a framework for evaluation in design science research. *Eur. J. Inf. Syst.* **25**, 77–89 (2016). <https://doi.org/10.1057/ejis.2014.36>
46. Palan, S., Schitter, C.: Prolific.ac—a subject pool for online experiments. *J. Behav. Experi. Finan.* **17**, 22–27 (2018). <https://doi.org/10.1016/j.jbef.2017.12.004>
47. Lowry, P.B., D’Arcy, J., Hammer, B., Moody, G.D.: “Cargo cult” science in traditional organization and information systems survey research: a case for using nontraditional methods of data collection, including mechanical Turk and online panels. *J. Strateg. Inf. Syst.* **25**, 232–240 (2016). <https://doi.org/10.1016/j.jsis.2016.06.002>
48. Seymour, M., Yuan, L., Dennis, A.R., Riemer, K.: Have we crossed the uncanny valley? understanding affinity, trustworthiness, and preference for realistic digital humans in immersive environments. *JAIS.* **22**, 591–617 (2021). <https://doi.org/10.17705/1jais.00674>
49. Chua, A.Y.K., Pal, A., Banerjee, S.: AI-enabled investment advice: will users buy it? *Comput. Hum. Behav.* **138**, 107481 (2023). <https://doi.org/10.1016/j.chb.2022.107481>
50. Gursoy, D., Chi, O.H., Lu, L., Nunkoo, R.: Consumers acceptance of artificially intelligent (AI) device use in service delivery. *Int. J. Inf. Manag.* **49**, 157–169 (2019). <https://doi.org/10.1016/j.ijinfomgt.2019.03.008>
51. Gefen, D.: Customer loyalty in E-commerce. *J. Assoc. Inf. Syst.* **3** (2002). <https://doi.org/10.17705/1jais.00022>
52. Chin, W.W.: The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (ed.) *Modern Methods for Business Research*. LEA, London (1998)
53. Wold, H.O.A.: Soft modeling: the basic design and some extensions. In: Jöreskog, K.G., Wold, H.O.A. (eds.) *Systems Under Indirect Observation*, pp. 1–54. North-Holland, Amsterdam (1982)
54. Richter, N.F., Schubring, S., Hauff, S., Ringle, C.M., Sarstedt, M.: When predictors of outcomes are necessary: guidelines for the combined use of PLS-SEM and NCA. *IMDS.* **120**, 2243–2267 (2020). <https://doi.org/10.1108/IMDS-11-2019-0638>
55. Hair, J., Hult, G.T.M., Ringle, C., Sarstedt, M.: *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2022)
56. Ab Hamid, M.R., Sami, W., Mohmad Sidek, M.H.: Discriminant validity assessment: use of Fornell & Larcker criterion versus HTMT criterion. *J. Phys. Conf. Ser.* **890**, 012163 (2017). <https://doi.org/10.1088/1742-6596/890/1/012163>
57. Henseler, J., Ringle, C.M., Sinkovics, R.R.: The use of partial least squares path modeling in international marketing. In: Sinkovics, R., Ghauri, P. (eds.) *New Challenges to International Marketing*, pp. 277–319. Emerald Group Publishing Limited (2009). [https://doi.org/10.1108/S1474-7979\(2009\)0000020014](https://doi.org/10.1108/S1474-7979(2009)0000020014)