Please quote as: Meywirth, S., Janson, A. & Söllner, M. (2026). Do human trust models transfer to AI? a necessary condition analysis of interpersonal trust antecedents in large language model-based coaching. Behaviour & Information Technology (BIT), online first, 1-19. doi: 10.1080/0144929X.2026.2676746

Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: www.tandfonline.com/journals/tbit20

Do human trust models transfer to AI? a necessary condition analysis of interpersonal trust antecedents in large language model-based coaching

Sophia Meywirth, Andreas Janson & Matthias Söllner

To cite this article: Sophia Meywirth, Andreas Janson & Matthias Söllner (21 May 2026): Do human trust models transfer to AI? a necessary condition analysis of interpersonal trust antecedents in large language model-based coaching, Behaviour & Information Technology, DOI: 10.1080/0144929X.2026.2676746

To link to this article: https://doi.org/10.1080/0144929X.2026.2676746

Published online: 21 May 2026.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2026.2676746

Do human trust models transfer to AI? a necessary condition analysis of interpersonal trust antecedents in large language model-based coaching

Sophia Meywirth [a], Andreas Janson [b] and Matthias Söllner [a]

[a]Information Systems and Systems Engineering, University of Kassel, Research Center for IS Design (ITeG), Kassel, Germany; [b]Institute of Information Management, University of St.Gallen, St. Gallen, Switzerland

ABSTRACT

As large language models (LLMs) increasingly serve as a technology basis for conversational health behaviour interventions, understanding how users evaluate and respond to these systems becomes critical for their acceptance. When LLM-based coaches are anthropomorphically designed – e.g. through human names, avatars, and messaging interfaces – they may trigger interpersonal trust expectations traditionally reserved for human relationships. This study investigates whether and to what extent the interpersonal trust antecedents of ability, benevolence, and integrity influence trust-related user responses in LLM-based health coaching for fitness and nutrition. Using a vignette-based survey design and combining partial least-squares structural equation modelling (PLS-SEM) with necessary condition analysis (NCA), we identify which antecedents are essential for perceived usefulness and intention to adopt. Results show that integrity is a necessary condition for both outcomes, with a medium to large effect, and that ability significantly influences perceived usefulness but is not necessary. Benevolence, however, was found to be neither a necessary nor a significant predictor in this context. These findings offer theoretical insights into the selective transfer of interpersonal trustworthiness beliefs to AI and provide practical guidance for the design of trustworthy anthropomorphic coaching systems.

ARTICLE HISTORY
Received 10 November 2025
Accepted 12 May 2026

KEYWORDS
Large language models; behavior change; digital health; trust; anthropomorphic design

1. Introduction

Recent advances in large language models (LLMs) such as ChatGPT have enabled conversational agents to communicate in ways that closely resemble human interaction (Demszky et al. 2023; Dwivedi et al. 2023; Vardhan et al. 2023). These conversational systems are increasingly applied in domains traditionally grounded in interpersonal relationships (Triantafyllopoulos et al. 2024; Zhang and Zhao 2024) – such as health coaching – where natural language interaction, empathy, and perceived social presence play a vital role (e.g. Brunswicker et al. 2025). As AI systems adopt human-like communication patterns and social cues, they invite users to engage with them not merely as tools, but as social actors (Nass and Moon 2000).

This anthropomorphic shift raises a critical question: Do users evaluate these agents using human trust expectations, and if so, which interpersonal dimensions of perceived trustworthiness matter most? In human relationships, perceived trustworthiness is commonly conceptualised through the dimensions of ability, benevolence, and integrity – known as the interpersonal trust model (Mayer, Davis, and Schoorman 1995). These beliefs are generally important for reliance on AI (Klingbeil, Grützner, and Schreck 2024) and become particularly critical in coaching contexts that depend on disclosure (Schiemann et al. 2019), sustained engagement, and behaviour change (Wu et al. 2022). Traditionally, trust in technology has been explained through constructs such as functionality, reliability, and helpfulness (McKnight and Chervany 2001), or through dimensions like performance and process (Lee and See 2004). However, if users apply interpersonal trustworthiness beliefs to anthropomorphic AI systems, this suggests that designing trustworthy agents goes beyond optimising technical performance: it would require shaping the agent's perceived personality traits to foster trust-related outcomes. As conversational agents become increasingly human-like, users may transfer these interpersonal expectations to AI systems (Lankton, McKnight, and Tripp 2015), blurring the

CONTACT Sophia Meywirth sophia.meywirth@uni-kassel.de Ringgold Standard Institution, Information Systems and Systems Engineering, Universitat Kassel, Kassel, Germany

longstanding distinction between trust in people and trust in technology (McKnight et al. 2011).

Maintaining healthy behaviours such as regular physical activity and balanced nutrition remains a persistent challenge (WHO 2025), despite well-established benefits for preventing chronic disease and promoting mental well-being (Warburton, Nicol, and Bredin 2006). While coaching interventions have proven effective in supporting behaviour change, access to such support is often limited (Greaves et al. 2011). Face-to-face coaching typically requires trained professionals and sustained engagement, which can be associated with high costs, limited scalability, and long waiting times (Mitchell et al. 2021). More broadly, structural barriers such as resource constraints and shortages of qualified professionals continue to restrict access to personalised support in health-related contexts (Ni and Jia 2025). These limitations highlight the need for scalable, accessible alternatives that can provide ongoing guidance and support for behaviour change.

While prior research has explored digital coaching systems, much of it has focused on scripted or rule-based chatbots that struggle to replicate the dynamic, relational aspects of human coaching (Mitchell et al. 2021). The emergence of LLM-based coaches marks a shift toward systems capable of open-ended, context-sensitive, and emotionally expressive dialogue (Kohnke and Moorhouse 2025). These agents can be embedded in familiar formats (e.g. messaging apps), assigned names and avatars, and introduced explicitly as AI-powered – creating an experience that feels both technological and interpersonal.

As these systems blur the boundary between technological tool and interpersonal partner, a central question is how interpersonal trustworthiness beliefs – particularly perceived ability, benevolence and integrity (Mayer, Davis, and Schoorman 1995) – shape users' responses to them. In coaching contexts, trust becomes meaningful when users perceive the system as valuable and are willing to engage with it over time, as sustained interaction, disclosure, and adherence are central to coaching effectiveness (Schiemann et al. 2019; Thom et al. 2016; Wu et al. 2022). Prior research in information systems has shown that trustworthiness beliefs influence such evaluative and behavioural responses (Gefen, Karahanna, and Straub 2003; Gefen 2002; Wang and Benbasat 2005). Accordingly, we examine how perceived ability, benevolence, and integrity influence perceived usefulness and intention to adopt an LLM-based coaching system. These outcomes capture how interpersonal trustworthiness beliefs translate into meaningful user engagement with human-like AI systems.

In consequence, this paper addresses an underexplored area in human-AI interaction research: understanding whether the interpersonal trust antecedents not only remain relevant but also represent necessary conditions for the emergence of trust-related responses in human-like, LLM-based coaching systems. While prior studies have examined factors that facilitate trust, such as ability, benevolence, and integrity, most research has relied on sufficiency logic (e.g. Lalot and Bertram 2025), focusing on factors that can enhance trust when present. However, little is known about which antecedents are non-negotiable. As trust has long been recognised as critical for user adoption (Wang and Benbasat 2005), it is essential to move beyond identifying influential predictors toward uncovering foundational preconditions. To address this, we complement traditional sufficiency analyses with a necessity perspective, enabling the identification of indispensable interpersonal trust antecedents in LLM-based coaching.

In consequence, we investigate the following research question (RQ):

> *RQ:* To what extent are interpersonal trust antecedents – ability, benevolence, and integrity – sufficient and necessary for trust-related user responses in anthropomorphic LLM-based health coaching systems?

To address our research question, we conducted a vignette-based survey in which participants assessed health coaching interactions and reported their perceptions of the AI coach's ability, benevolence, and integrity. The coaching agent was presented in a human-like format incorporating multiple established verbal and visual social cues (Feine et al. 2019) such as WhatsApp interface, human name and photo, and conversational language, while being explicitly labelled as an AI. We apply a dual-method approach – partial least squares structural equation modelling (PLS-SEM) and Necessary Condition Analysis (NCA; Dul 2016) – to assess both the strength and necessity of each trust belief in driving perceived usefulness and intention to adopt. To our knowledge, this is the first study to apply NCA to investigate trust antecedents in AI-mediated coaching.

By integrating interpersonal trust theory with research on human-AI interaction, this study contributes to the ongoing debate on how users relate to increasingly anthropomorphic systems (Kim and Im 2023). Theoretically, we examine whether the interpersonal trust model, long used to study human trust, applies to human-like AI coaches. Practically, our findings inform the design of trustworthy AI systems by identifying which interpersonal traits must be communicated to foster user engagement.

2. Related work and theoretical foundations

2.1. Trust in human and technological agents

To understand how trust-related responses towards anthropomorphic AI systems emerge, it is essential to first examine how trust is conceptualised in both human and technological contexts. Trust plays a central role in facilitating cooperation and effective transactions, particularly in situations marked by uncertainty and complexity (Lee and See 2004). It allows individuals to engage with others despite risks, by relying on expected behaviour (Rousseau et al. 1998).

Across disciplines such as psychology, sociology, and information systems, trust is seen as a multifaceted construct (Gulati et al. 2024). In this paper, we adopt the widely used definition by Mayer, Davis, and Schoorman (1995, 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 has become one of the most widely adopted conceptualizations of trust in the field of Information Systems (Rousseau et al. 1998). A central component of this definition is the notion of vulnerability - trust only becomes relevant when individuals willingly place themselves at risk based on positive expectations about another party's actions (Lee and See 2004).

When referring to trust in technological artifacts, trust is commonly distinguished into two categories: interpersonal trust, which refers to trust in individuals or organisations, and technology trust, which refers to trust in IT systems or digital artifacts (Söllner et al. 2012, 2016; Thiebes, Lins, and Sunyaev 2021). This distinction clarifies who or what the 'trustee' is in a given interaction and determines which trustworthiness beliefs apply (McKnight and Chervany 2001). Trust in technology is often linked to attributes such as functionality, reliability, or helpfulness (McKnight et al. 2011; Söllner, Pavlou, and Leimeister 2013; Thatcher et al. 2011).

In contrast, interpersonal trust is commonly conceptualised through the three trust antecedents proposed by Mayer, Davis, and Schoorman (1995): ability, benevolence, and integrity. Ability refers to the competence and skills of the trustee; benevolence captures the perception that the trustee genuinely cares about the trustor's well-being, independent of self-interest; and integrity reflects adherence to principles that the trustor finds acceptable. These antecedents have been widely applied in studies of organisational and interpersonal relationships and form the basis for our investigation into how trust forms in AI systems that simulate human interaction.

Many of these models treat AI as autonomous tools and focus on system-related attributes such as automation, reliability, and transparency (Thiebes, Lins, and Sunyaev 2021). For instance, Lee and See (2004) proposed a model of trust in automation based on performance, process, and purpose. However, such technology-centered trust models may no longer suffice in highly social and human-like settings. When AI systems are anthropomorphized – through natural language interaction, self-reference, and expressive cues – users may apply social and interpersonal expectations to evaluate trust. In these contexts, trust is not solely based on technical performance but also shaped by perceptions of social presence (Janson 2023). For an integrative overview of both interpersonal and technology-related trust antecedents – including those applied to AI systems – see Thiebes, Lins, and Sunyaev (2021).

Lankton, McKnight, and Tripp (2015) argue that the relevance of specific trust constructs depends on the perceived humanness of the system and found that interpersonal trust models – particularly the ability, benevolence, and integrity – were more predictive in high-humanness systems. Similarly, Lalot and Bertram (2025) demonstrated that users meaningfully apply the interpersonal trust model to human-like AI agents.

2.2. Human-likeness in AI and conversational agents

As AI systems increasingly mimic human traits, the line between interpersonal and technological trust begins to blur. This shift is driven by the phenomenon of anthropomorphism, which fundamentally alters how users relate to these systems. Anthropomorphism refers to the attribution of human traits to non-human entities, including AI systems (Epley, Waytz, and Cacioppo 2007; Fink 2012). As technological agents become more human-like in how they speak, respond, and present themselves, users often evaluate them not merely as tools, but as social partners. A seminal framework that captures this phenomenon is the 'Computers are Social Actors' (CASA) paradigm (Nass and Moon 2000), which shows that people apply interpersonal norms to computers – even while being aware that these systems are not human. For instance, users tend to show politeness, assign gender roles, or attribute personality traits to conversational interfaces, especially when those systems exhibit social cues or engage in natural dialogue (Nass and Moon 2000).

This tendency to treat computers as social actors has been explored through various theoretical lenses (Fink 2012). Two dominant perspectives explain why users respond to non-human agents as if they were human: the first suggests that people automatically react to social or life-like cues, applying heuristics and stereotypes without deliberate thought (Lee et al. 2005). The second, more cognitive perspective, argues that users build mental models of how a system behaves; when a system acts in ways that resemble human behaviour, users may align their expectations with those they have for other people (Lee et al. 2005). This distinction – often described as 'mindless' (automatic) versus 'mindful' (cognitive) anthropomorphism – has important implications for trust (Araujo 2018; Kim and Sundar 2012). Mindless anthropomorphism can lead to intuitive judgments of trustworthiness based on surface-level cues, while mindful anthropomorphism involves more deliberate comparisons to human-like expectations. Understanding these mechanisms is especially relevant today, as recent advancements in artificial intelligence have given rise to systems that exhibit increasingly sophisticated human-like behaviours.

Since the early 2000s, advances in artificial intelligence have dramatically changed how users conversationally interact with computer systems. Whereas earlier conversational agents were limited to rule-based responses and lacked the capacity to simulate human-like behaviour (Schöbel, Söllner, and Leimeister 2016) today's LLMs can generate fluent, responsive dialogue, use self-referential language, and express seemingly emotional or moral stances (Stein and MacDorman 2024). This development is especially evident in conversational agents, which now communicate in ways that closely resemble human interaction (Dwivedi et al. 2023)

To better understand how conversational agents evoke human-like responses, Feine et al. (2019) developed a widely used taxonomy of social cues for conversational agents. They distinguish between verbal, visual, auditory, and invisible cues that conversational agents can use to simulate social presence. In text-based systems, verbal cues are especially important, as all user interaction occurs through language. These verbal cues can be further categorised into style-related cues – how something is said (e.g. tone, formality) – and content-related cues – what is said (e.g. expressions of empathy, moral judgment, or encouragement).

While much prior research on human-like chatbot design has focused on visual design choices (e.g. avatars, names) or stylistic language features (e.g. formal vs. casual tone) (Fadhil et al. 2018; Janson 2023), the role of content-specific verbal has received comparatively less attention. Yet in text-based interactions, particularly those requiring emotional resonance and sustained engagement, users may be especially sensitive to the agent's verbal content.

2.3. Trust in anthropomorphic coaching systems

The importance of verbal social cues becomes especially pronounced in health coaching, where human-like conversation plays a central role in building trust and rapport (Ryan, Dockray, and Linehan 2022). Particularly in lifestyle behaviour change interventions, users expect more than competence alone; they seek relational signals that foster emotional connection and sustained engagement (Jörke et al. 2024). Prior work has highlighted that in coaching, benevolence may be particularly important because coaching relationships revolve around supporting clients' autonomy, self-valued goals, and personal growth (Schiemann et al. 2019). In this context, verbal expressions of empathy, fairness, and understanding can strengthen users' perceptions of the coach's benevolence and integrity. Meaningful dialogue beyond purely task-oriented exchanges is therefore critical for creating a supportive and trusted relationship between the user and the system (Ryan, Dockray, and Linehan 2022). Nadarzynski et al. (2019) found that while users appreciated the nonjudgmental nature of AI coaches, they also emphasised that trust and meaningful interaction were still required to build rapport, even with a chatbot. A qualitative study by Thom et al. (2016, 512) similarly found that trust was central to the coach's ability to support patients. As one coach in their study noted, 'A patient would probably tell me the things I want to hear and not the things that are really happening, and then I wouldn't be able to help the patient that well.' if trust was missing. This shows how trust enables not only adherence but also honest, meaningful interaction – without which coaching loses its effectiveness. In digital settings, this need for rapport persists, as we outline in our review of related work in the context of conversational agents for the context of coaching and (mental) health support (see Table 1).

The reviewed studies collectively underscore the central role of anthropomorphic design, communication style (e.g. Scholich et al. 2025), and trust-related factors in shaping user perceptions and interactions with conversational agents across health, wellness, and medical contexts (e.g. Jin and Eastin 2025). Common themes include the importance of perceived intelligence and personalisation (Beinema et al. 2021; Kettunen et al. 2024), the moderating role of demographic or

Table 1. Related work and study positioning.

Source	Study	Main Results and Contribution
Beinema et al. (2021)	Tailoring coaching strategies to users' motivation in a multi-agent health coaching application	ECAs adapted strategies to user motivation profiles. Dual-motivated users preferred positive strategies. Strategy appreciation increased with personalisation; agent likeability had limited impact.
Brunswicker et al. (2025)	Influence of systemising and empathising language structures on trust and cognitive effort in chatbots	Empathetic language increased trust and helpfulness unless users were angry – then trust dropped. Systemising logic reduced cognitive effort. Identified boundaries of language-based anthropomorphism.
Jin and Eastin (2025)	Effects of chatbot design cues and threat perception on trust in health information	Found that doctor-like and bot-like design cues enhanced perceived expertise and trust in health information compared to layperson-like cues. These effects were significant only under high perceived health threat. Design cues influenced privacy concerns in interaction with threat perception, highlighting the conditional impact of anthropomorphic and professional cues.
Kettunen et al. (2024)	Critical usage experiences with digital coaching in sport and wellness technologies across age groups	Identified critical positive and negative experiences with digital coaches among young adults and young elderly. Positive incidents increased motivation via feedback and perceived competence; negative incidents (e.g. data inaccuracy, complexity) reduced trust and use. Elderly valued supportive feedback more than young adults.
Mayer et al. (2024)	User preferences and trust in digital vs. analog and AI-based medical consultation formats	Face-to-face formats preferred for sensitive/emotional topics. Higher perceived efficiency increased preference for digital formats. Trust in digital was lower for serious topics. Preferences shaped by topic, not demographics.
Scholich et al. (2025)	Comparison of LLM-based chatbot and therapist responses in therapeutic communication	Mixed-methods study comparing chatbots to licensed therapists. Chatbots performed well at validation and psychoeducation but showed lack of inquiry, feedback seeking, and crisis response. Chatbots overused generic and directive advice, raising safety and ethical concerns.
Triantafyllopoulos et al. (2024)	Evaluation of ChatGPT interactions by Greek physicians across specialties and experience levels	Moderate satisfaction with ChatGPT responses. Younger/less experienced doctors had more positive views. Internal medicine/psychiatry specialists rated it more favourably. Highlighted need for local language adaptation.
Zhang and Zhao (2024)	Influence of ChatGPT perceptions and doctor satisfaction on adoption of GPT-powered medical consultation systems	Adoption predicted by perceived usefulness/ease of use. Positive ChatGPT perceptions increased adoption; high satisfaction with doctors increased resistance. Resistance mediated the doctor satisfaction–adoption link.
This Study	*Transfer of interpersonal trust antecedents (ability, benevolence, integrity) to LLM-based coaching systems*	*Integrity is a necessary and influential condition for key trust-related user responses (perceived usefulness and intention to adopt) in anthropomorphic LLM coaches. Ability affects perceived usefulness but is not essential, while benevolence is neither necessary nor significant for both outcomes. Highlighting the selective transferability of interpersonal trustworthiness beliefs to AI and implications for coach design.*

contextual variables like age, domain familiarity, or doctor satisfaction (Mayer et al. 2024; Triantafyllopoulos et al. 2024; Zhang and Zhao 2024), and the impact of agent language and emotional framing on trust and engagement (Brunswicker et al. 2025). Across these studies, trust emerges as a critical user response, frequently linked to factors such as performance feedback, system usability, social presence, and emotional alignment. However, most prior work treats trust as a composite outcome or explores it through traditional user experience variables.

This study extends prior research by introducing a necessity-based perspective to the analysis of trust in LLM-based coaching systems – an approach largely absent from existing work. While previous studies have primarily focused on identifying factors that are sufficient to increase trust, this study draws on necessity logic to explore which interpersonal trust antecedents must be present for trust-related user responses to emerge. By combining variance-based and necessity-oriented methods, it allows for a deeper understanding of the impact of interpersonal trustworthiness beliefs on user responses, potentially uncovering conditions that are critical and non-compensable. This methodological shift complements existing findings by revealing not just what influences user adoption, but what is fundamentally required to enable it.

3. Hypotheses development and research model

As conversational agents in coaching contexts increasingly resemble human partners, traditional models of trust in technology may no longer fully explain how users respond to them. In such socially rich settings, users are likely to apply interpersonal expectations, making interpersonal trustworthiness beliefs like ability, benevolence, and integrity more appropriate.

While prior studies (e.g. Lalot and Bertram 2025; Lankton, McKnight, and Tripp 2015) have demonstrated that the interpersonal trust model better predicts trust in highly human-like systems, they focus primarily on how strongly these beliefs influence trust (i.e. their sufficiency). We build on this work by asking an additional and critical question: Which trustworthiness

beliefs are also functionally required – that is, necessary – for trust-related user responses to emerge?

This dual focus on sufficiency and necessity offers a more nuanced understanding of how trustworthiness beliefs shape user engagement in anthropomorphic systems, especially in sensitive domains like health coaching. We therefore apply both perspectives in our research model (Figure 1) to examine the influence of perceived ability, benevolence, and integrity on two trust-relevant outcomes: perceived usefulness and intention to adopt. The following sections outline our hypotheses.

3.1. Sufficiency of interpersonal trustworthiness beliefs for trust outcomes

Building on the theoretical grounding that LLM-based coaches increasingly resemble human partners, we argue that it is not only important to apply interpersonal trust models – but also to understand which specific antecedents most strongly influence trust-relevant responses to such systems. As Mayer, Davis, and Schoorman (1995) emphasise, ability, benevolence, and integrity are distinct yet interrelated dimensions of trustworthiness. Each contributes a unique perceptual perspective from which users judge the trustee. While a system may be strong in one area, weaknesses in another may significantly impair overall trust. 'Thus, it is possible for a perceived lack of any of the three factors to undermine trust. […] Each of the three factors can vary along a continuum' (Mayer, Davis, and Schoorman 1995, 721). This dual insight – that trustworthiness beliefs vary in degree yet that deficiencies in any one dimension may undermine trust – suggests the relevance of examining both sufficiency and necessity in anthropomorphic AI contexts.

Sufficiency logic allows us to assess the extent to which each of these antecedents contributes to trust-related responses in LLM-based health coaching. Specifically, it helps determine which interpersonal trustworthiness beliefs are particularly influential in driving trust-related outcomes such as perceived usefulness and intention to adopt.

To motivate this approach, we draw on Mayer et al.'s (1995) foundational framework and link it to practical reasoning in coaching contexts. As Terblanche and Heyns (2020, 10) observe in the context of behavioural coaching, 'the coach can play an active role in building the perceived [trustworthiness] by demonstrating ability, benevolence and integrity,' underscoring the practical relevance of all three dimensions in shaping trust-related user responses.

Perceived ability reflects the extent to which users believe the coach possesses the competence and expertise required to provide effective guidance (Mayer, Davis, and Schoorman 1995). In health coaching contexts, users expect coaches to demonstrate relevant knowledge and provide actionable, medically informed advice (McQueen et al. 2020). When a coaching system is perceived as competent, users are more likely to evaluate its recommendations as valuable for achieving their health goals. Higher perceptions of ability also increase confidence in the system's effectiveness, thereby strengthening users' willingness to engage with and adopt the coaching system. Thus, we hypothesise:

> H1a: The perceived ability of a human-like LLM-based health coach positively influences the user's perceived usefulness.

> H2a: The perceived ability of a human-like LLM-based health coach positively influences the user's intention to adopt the LLM-based coaching system.

Benevolence reflects the extent to which users believe that the coach genuinely cares about their well-being and acts in their interest (Mayer, Davis, and Schoorman 1995). In coaching contexts, communicating benevolence –

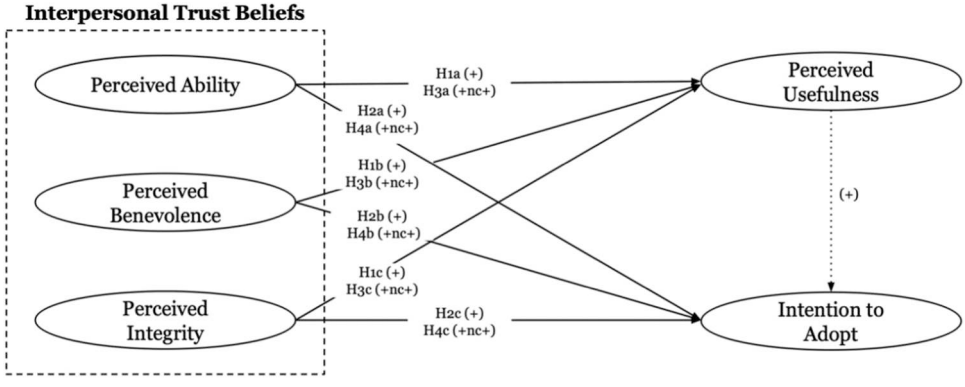

Figure 1. Research model.

through empathy, appreciation, and supportive interaction – plays a central role in establishing trustworthiness (Schiemann et al. 2019). Research on coaching with compassion further suggests that experiences of empathic concern foster openness, learning, and sustained engagement (Boyatzis, Smith, and Beveridge 2013), which are critical mechanisms for effective health coaching and behaviour change (Thom et al. 2016). When users perceive a coaching system as supportive and genuinely invested in their development, they are more likely to experience the interaction as personally meaningful and valuable. Higher perceptions of benevolence therefore strengthen users' evaluations of the system's usefulness and increase their willingness to engage with and adopt the coaching system. Thus, we hypothesise:

H1b: The perceived benevolence of a human-like LLM-based health coach positively influences the user's perceived usefulness.

H2b: The perceived benevolence of a human-like LLM-based health coach positively influences the user's intention to adopt the LLM-based coaching system.

Integrity reflects the extent to which users perceive that the coach adheres to a set of principles that they find acceptable (Mayer, Davis, and Schoorman 1995). This includes perceptions of consistency between words and actions, credible communication, and the belief that the trustee acts according to stable and just principles. In health coaching contexts, users must evaluate whether the system's recommendations align with their own values and goals and whether its guidance appears consistent and principled. When a health coach is perceived as acting in accordance with acceptable and value-congruent principles, users are more likely to view its recommendations as credible and worthwhile. Higher perceptions of integrity therefore strengthen users' evaluations of the system's usefulness and increase their willingness to engage with and adopt the coaching system. Thus we, hypothesise:

H1c: The perceived integrity of a human-like LLM-based health coach positively influences the user's perceived usefulness.

H2c: The perceived integrity of a human-like LLM-based health coach positively influences the user's intention to adopt the LLM-based coaching system.

3.2. Necessity of interpersonal trustworthiness beliefs for trust outcomes

While sufficiency logic tells us how strongly different antecedents contribute to trust-related outcomes, it does not tell us whether any of them are essential. Mayer, Davis, and Schoorman (1995) note that a perceived deficiency in ability, benevolence, or integrity may undermine trust, suggesting that the absence of any one dimension can constrain the development of trust. Building on this insight, we adopt a necessity perspective that asks whether certain trustworthiness beliefs must reach a minimal level for trust-related user responses to occur.

In health coaching contexts, users expect coaches to possess relevant expertise and domain knowledge. Qualitative evidence suggests that clients view health professionals as the experts on medical issues and value coaching guidance that is informed, competent, and educational in nature (McQueen et al. 2020). These expectations imply that a minimal perception of competence is foundational for coaching interactions to be taken seriously. If users do not perceive the system as sufficiently knowledgeable, they may question the credibility and appropriateness of its recommendations, regardless of how supportive or well-intentioned the interaction appears. Without at least a basic level of perceived ability, users may be unwilling to rely on the system's guidance, preventing perceived usefulness from reaching meaningful levels and limiting intention to adopt. Perceived ability may therefore constitute a necessary condition for trust-related outcomes in LLM-based health coaching. Thus, we hypothesise:

H3a: The perceived ability of a human-like LLM-based health coach is necessary for the user's perceived usefulness.

H4a: The perceived ability of a human-like LLM-based health coach is necessary for the user's intention to adopt the LLM-based coaching system.

Health coaching frequently involves the disclosure of personal goals, habits, and challenges. If users do not perceive the system as supportive and genuinely concerned with their well-being, they may hesitate to engage openly with the coaching process. As Schiemann et al. (2019) note, empathic and autonomy-supportive behaviour in coaching is closely aligned with the perception of benevolence. In the absence of a minimal perception of benevolence, users may therefore withhold engagement, thereby limiting the development of perceived usefulness and intention to adopt. Benevolence may thus constitute a necessary relational foundation for meaningful interaction with LLM-based coaches. Thus, we hypothesise:

H3b: The perceived benevolence of a human-like LLM-based health coach is necessary for the user's perceived usefulness.

H4b: The perceived benevolence of a human-like LLM-based health coach is necessary for the user's intention to adopt the LLM-based coaching system.

Integrity concerns the perception that the coach adheres to acceptable and consistent principles (Mayer, Davis, and Schoorman 1995). In health-related contexts, users must believe that recommendations align with their own values, as coaching relationships depend on confidence in the coach's principles, consistency, and ethical orientation – particularly when advice affects health or lifestyle decisions. If a minimal perception of integrity is absent, users may question the legitimacy of the system's guidance, preventing perceived usefulness from reaching meaningful levels and undermining adoption intentions. Integrity may therefore represent a foundational condition for trust-related user responses. Thus, we hypothesise:

> Hypothesis H3c: The perceived integrity of a human-like LLM-based health coach is necessary for the user's perceived usefulness.

> Hypothesis H4c: The perceived integrity of a human-like LLM-based health coach is necessary for the user's intention to adopt the LLM-based coaching system.

To empirically test this proposition, we use Necessary Condition Analysis (NCA; Dul 2016, 2021), a method that identifies bottleneck conditions – that is, factors whose absence prevents an outcome regardless of other strengths. In doing so, we address an important research gap. While prior studies (e.g.Lalot and Bertram 2025; Lankton, McKnight, and Tripp 2015) have explored the influence of interpersonal trustworthiness beliefs on perceived trust and trust-related user responses, they have not examined whether these beliefs are functionally required to reach key thresholds. Our study contributes a novel perspective by identifying which trustworthiness beliefs are indispensable for trust-related user responses of LLM-based health coaching systems.

4. Research method

4.1. Overview of research method approach

To empirically examine the proposed research model, we employed a vignette-based online survey with 218 participants. Participants were presented with health coaching interaction scenarios and asked to evaluate an AI-based coach with respect to perceived ability, benevolence, and integrity. The coach was depicted using human-like cues, including a messaging-style interface, a human name, and a profile picture, while being clearly identified as an AI system. We analyzed the data using a dual-method approach combining partial least squares structural equation modelling (PLS-SEM) and Necessary Condition Analysis (NCA; Dul 2016) to assess both sufficiency and necessity relationships between trustworthiness beliefs, perceived usefulness, and intention to adopt.

The approach to use structural equation modelling (SEM) with the variance-based partial least squares (PLS) (Chin 1998; Wold 1982) suits our research goal because we aim to identify key 'driver' constructs and need to use latent variable scores in our subsequent necessary condition analysis (J. F. Hair, Ringle, and Sarstedt 2011). 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. (Richter et al. 2020) PLS-SEM findings follow a sufficiency logic ('X increased Y') and therefore is suited to identify should-have factors of information systems (Richter et al. 2020). NCA, on the other hand, can identify the necessary conditions that must be met in order to achieve a specific level of the desired outcome ('X is necessary for Y') (Dul 2019). Both of these factors are crucial for the design and development of successful information systems, as the identification of the should-have factors is relevant to increasing an outcome, but this is only possible after the must-have factors are fulfilled. (Richter et al. 2020)

NCA is a relatively novel approach to data analysis that identifies necessary conditions in data sets (Dul 2016, 2019). In contrast to traditional data analysis techniques, NCA does not analyze the average relationships between dependent and independent variables. Instead, it aims to reveal areas in scatter plots with no cases. When conducting NCA, a ceiling line is determined and placed on top of the data to separate the space with and without observations. The larger the identified empty space, the greater the constraint that X places on Y. The ceiling line also indicates the minimum level of X required to achieve a certain level of Y. The bottleneck table presents the ceiling line in a table format, showing the outcome in the first column and the conditions that must be fulfilled in the following columns. The two key metrics in NCA are ceiling accuracy (c-accuracy) and necessity effect size (d). Ceiling accuracy is defined as the ratio of observations that lie on or beneath the ceiling line to the total observations multiplied by 100. While there is no definitive threshold for acceptable c-accuracy, it is often compared against a benchmark, such as 95%, for evaluative purposes. The necessity effect size (d) measures the proportion of the empty ceiling zone to the occupied space. The value of d ranges from 0 to 1. Values between 0 and less than 0.1 indicate a small effect, 0.1 and less than 0.3 indicate a medium

effect, 0.3 and less than 0.5 signify a large effect, and 0.5 or greater denotes a very large effect.

Regarding sample size considerations, we followed established guidelines for PLS-SEM by assessing sample size adequacy based on the maximum number of structural paths directed at any endogenous construct in the model (Hair et al. 2017). In the proposed structural model, intention to adopt represents the most complex endogenous construct, with four incoming paths originating from perceived ability, perceived benevolence, perceived integrity, and perceived usefulness. Referring to the minimum sample size recommendations provided by Hair et al. (2017), and assuming a conservative minimum R^2 of 0.10 and a significance level of 0.05, a minimum sample size of 113 observations is required.

With respect to the Necessary Condition Analysis (NCA), recent methodological work emphasises that NCA does not rely on a single fixed minimum sample size, as statistical power depends on assumptions that are typically unknown ex ante; accordingly, sample size is reported transparently, and necessity results are interpreted with appropriate caution (Dul 2024).

4.2. Participants

We recruited 218 participants from Prolific (Palan and Schitter 2018) and adhered to the guidelines for studies on online platforms (Lowry et al. 2016). We removed 14 of the collected responses because the participants failed attention checks. Additionally, we removed nine responses because the participants rushed through the survey (duration of less than five minutes). Through this process, we ended up with 196 valid responses.

More males (58.2%) than females (40.3%) and non-binaries (1.5%) participated in our study. The participants were, on average, 38 years old (minimum age 18 years, maximum age 78 years). Most participants had a completed bachelor's degree (40.8%), followed by participants with a High School Diploma (35.7%) as their highest level of education. 75% of the participants said that they do not have any prior experience in health or lifestyle coaching, and 67.3% said that they currently have a desire to change their behaviour when it comes to fitness and nutrition.

4.3. Survey procedures

In line with prior studies and methodological guidance, we developed a vignette of a human-like LLM-based health coaching intervention that provides a realistic setting (Aguinis and Bradley 2014; Ötting and Maier 2018) to elicit relevant trustworthiness beliefs and user evaluations. This vignette displayed the first session of the chat-based interaction with the health coach. The session was focused on goal-setting, as this is a pivotal aspect of health coaching (Mitchell et al. 2021). Therefore, techniques from the 'goals and planning' cluster of the behaviour change technique taxonomy (Michie et al. 2013) were implemented. Additionally, passages were incorporated in which the coach provided advice and suggested strategies regarding health-related behaviour, namely dietary advice. This approach included two interaction features that are commonly associated with trsut in coaching contexts: sharing one's goals and wishes (Schiemann et al. 2019) and adhering to advice (Wu et al. 2022).

Each interaction began with a message that informed the users that the interaction was based on a LLM:

> Hello! I'm Chris, your virtual lifestyle coach for nutrition and physical activity. Just so you know, I'm powered by a large language model designed to provide information and support in your health and fitness journey. While I can offer tips and guidance based on the knowledge I've been trained on, I'm not a human. If you have any specific health concerns, please consult a professional. Now, let's get started! How can I assist you today?

The messages were incorporated into a prototype through Figma to complete the vignette. The interface was designed to mirror the layout of the messaging service WhatsApp. This included the use of a profile picture of a male human coach. In line with prior research on conversational agents, the interface incorporates multiple established verbal and visual social cues – such as a human name, profile picture, and conversational language – to evoke a human-like interaction setting (Feine et al. 2019).

The integration of the messages into this specific interface serves two purposes. Firstly, it ensures that the design closely replicates the familiar user experience of human-to-human chat interactions. Secondly, it facilitates a realistic implementation of an LLM. Tools such as Chatbase[1] facilitate the incorporation of LLMs, including ChatGPT, into popular messaging platforms such as WhatsApp. Finally, we account for the probabilistic nature of large language models as they typically assign a probability to every possible subsequent token and then select among the most likely token candidates. Thus, we sampled four slightly different interactions, all with the same goal. Figure 2 shows the survey procedure.

The survey was conducted in three steps. Firstly, we briefly described the objectives and context of the study. Secondly, we introduced the participants to the vignette scenario, explaining that they needed to put themselves in the position of receiving LLM-based coaching for healthy nutrition and physical activity.

Figure 2. Survey procedure.

We informed them that they were beginning their training with their LLM-based coach, Chris, and setting goals for their behaviour change through an online chat. Lastly, we presented the participants with the prototype vignette, in which the first coaching session was performed, and asked them to assess its content in an online survey.

To mitigate potential common method bias, we implemented several procedural safeguards ex ante in the design of the study. Following methodological recommendations (Podsakoff et al. 2003), we separated measurement blocks by thematic context, used clearly worded and behaviourally anchored items, and incorporated attention checks to ensure participant engagement with the survey. These design-based remedies are considered more effective than relying solely on statistical post hoc controls. In line with recent methodological guidance (Podsakoff et al. 2024), which emphasises the importance of preventive design strategies and transparent reporting, we did not conduct an additional post hoc statistical test for common method bias.

4.4. Survey measures

The questionnaire included several subsections regarding formative trustworthiness belief constructs and reflective outcomes constructs. Additionally, the questionnaire included questions regarding sociodemographic data, such as personal innovativeness in the domain of information technology (Agarwal and Prasad 1998).

Building upon previous research (Kamis, Koufaris, and Stern 2008; Söllner, Hoffmann, and Leimeister 2016; Wang and Benbasat 2005), we included three attributes indicating formative trustworthiness beliefs measures perceived ability (5 items), perceived benevolence (4 items) and perceived integrity (4 items) and two attributes indicating reflective outcome measures, namely perceived usefulness (3 items) and intention to adopt (5 items). The items were formulated as statements and assessed using a 7-point Likert scale. Table 2 provides an overview of the indicators and statements used to assess the attributes of trust.

5. Results

5.1. Measurement models

We began by assessing the measurement model. An examination of univariate normality showed that all indicators were within acceptable limits, with skewness and kurtosis values remaining between −2 and 2 (Hair et al. 2022).

Next, we evaluated the constructs' reliability and convergent validity using Cronbach's alpha, composite reliability (ρ_c and ρ_a), and average variance extracted (AVE) as metrics. All values met or exceeded the recommended thresholds – 0.7 for Cronbach's alpha, ρ_c, and ρ_a, and 0.5 for AVE (Hair et al. 2022). A summary of these results is presented in Table 3.

Furthermore, we evaluated the discriminant validity by measuring the heterotrait-monotrait (HTMT) ratios

Table 2. Indicators and attributes of trust in LLM-based coaching.

Attribute	Indicator	Statement	Source
Perceived Ability	ability1	The virtual coach does a good job.	Söllner, Hoffmann, and Leimeister (2016); Wang and Benbasat (2005)
	ability2	The virtual coach is like a real expert in fitness and nutrition	
	ability3	The virtual coach has the expertise to understand my needs and preferences about fitness and nutrition.	
	ability4	The virtual coach has the ability to understand my needs and preferences about fitness and nutrition.	
	ability5	The virtual coach considers my needs and all important attributes of fitness and nutrition.	
Perceived Benevolence	benevolence1	It is important for the virtual coach that he supports me in achieving my goals.	Söllner, Hoffmann, and Leimeister (2016)
	benevolence2	The virtual coach puts my interests first.	Wang and Benbasat (2005)
	benevolence3	The virtual coach keeps my interests in mind.	
	benevolence4	The virtual coach wants to understand my needs and preferences.	
Perceived Integrity	integrity1	I can count on the statements of the virtual coach.	Söllner, Hoffmann, and Leimeister (2016)
	integrity2	The virtual coach provides unbiased recommendations.	Wang and Benbasat (2005)
	integrity3	The virtual coach is honest.	
	integrity4	I consider the virtual coach to possess integrity.	
Perceived Usefulness	PU1	Using the virtual coaching intervention could improve my fitness and nutritional habits.	Söllner, Hoffmann, and Leimeister (2016)
	PU2	Using the virtual coaching intervention could improve my effectiveness in fitness and nutrition.	
	PU3	The virtual coaching intervention is a useful tool to support me in reaching my goals related to fitness and nutrition.	Adapted from Kamis, Koufaris, and Stern (2008)
Intention to Adopt	INT_A1	Assuming I had access to the virtual coaching intervention, I am willing to use this virtual coaching intervention as an aid to help me set goals related to my physical activity and diet.	Wang and Benbasat (2005)
	INT_A2	Assuming I had access to the virtual coaching intervention, I am willing to share my goals and intentions with regard to physical activity and nutrition with this virtual coach.	
	INT_A3	Assuming I had access to the virtual coaching intervention, I am willing to let this virtual coaching intervention assist me in deciding how to reach my goals related to physical activity and nutrition.	
	INT_A4	Assuming I had access to the virtual coaching intervention, I am willing to use this virtual coaching intervention as a tool to suggest a strategy for achieving my physical activity and nutrition goals.	
	INT_A5	Assuming I had access to the virtual coaching intervention, I am willing to follow the physical activity and nutrition recommendations provided by the virtual coach.	

of correlations (Table 4). The discriminant validity measures to what extent the latent constructs are truly distinct from each other (Ab Hamid, Sami, and Mohmad Sidek 2017). None of our latent constructs exhibited a value above the suggested threshold of 0.9.

To establish discriminant validity, we calculated the heterotrait-monotrait (HTMT) ratios of correlations (Table 4), which indicate the degree to which latent constructs are distinct from one another (Ab Hamid, Sami, and Mohmad Sidek 2017). All HTMT values were below the recommended cut-off of 0.9, confirming adequate discriminant validity (Hair et al. 2022) (Hair et al. 2022).

5.2. Structural model evaluation

We report the structural model results by presenting the path coefficients (ß), corresponding p-values, and effect sizes (f^2). To assess the significance of the paths, we applied a bootstrapping procedure with 5,000 resamples. An overview of the outcomes is provided in Figure 3.

The analysis demonstrated the validity of hypotheses H1a, H1c, and H2c, indicating significant relationships between perceived ability and perceived usefulness ($p \leq 0.001$), perceived integrity and perceived usefulness ($p \leq 0.001$), and perceived integrity and intention to adopt ($p \leq 0.01$). Additionally, a post-hoc analysis

Table 3. Construct reliability and validity.

Latent Construct	Cronbach's Alpha	Composite Reliability (rho_c)	Composite Reliability (rho_a)	Average variance extracted (AVE)
Perceived Ability	0.963	0.964	0.972	0.872
Percieved Benevolence	0.907	0.927	0.936	0.787
Perceived Integrity	0.887	0.898	0.922	0.747
Perceived Usefulness	0.967	0.967	0.978	0.938
Intention to Adopt	0.978	0.979	0.983	0.921

Table 4. HTMT for the assessment of discriminant validity.

Latent Construct	Intention to Adopt	Perceived Ability	Perceived Benevolence	Perceived Integrity
Intention to Adopt				
Perceived Ability	0.794			
Perceived Benevolence	0.650	0.741		
Perceived Integrity	0.830	0.856	0.771	
Perceived Usefulness	0.864	0.820	0.608	0.809

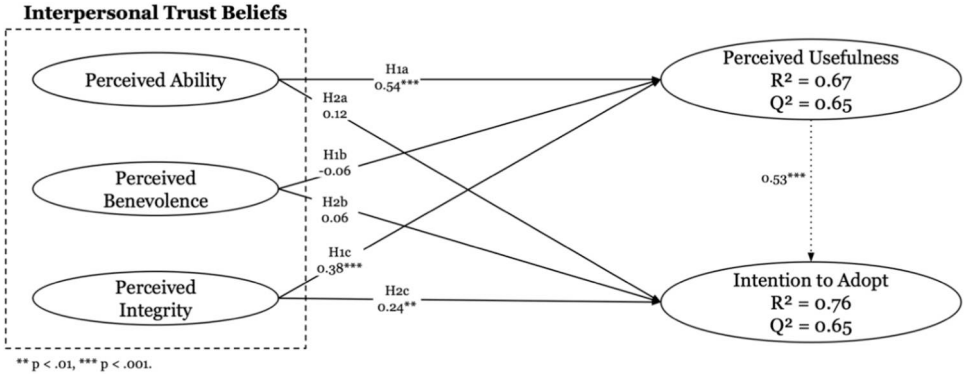

Figure 3. Structural model.

beyond a-priori formulated hypotheses revealed a significant relationship between perceived usefulness and intention to adopt ($p \leq 0.001$). Therefore, hypotheses H1b, H2a, and H2b are rejected.

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, as defined by Henseler, Ringle, and Sinkovics (2009). Perceived ability significantly affects perceived usefulness with a medium effect size ($f^2 \geq 0.15$), as indicated by a strong path coefficient ($\beta = 0.536$) and a highly significant p-value ($p \leq 0.001$, $f^2 = 0.281$). Similarly, perceived usefulness exerts a significant influence on adoption intention (ß = 0.533, $p \leq 0.001$) and is associated with a large effect size ($f^2 = 0.388$). In contrast, perceived integrity has a positive but comparatively modest effect on perceived usefulness ($\beta = 0.376$, $f^2 = 0.03$) and intention to adopt ($\beta = 0.283$, $f^2 = 0.07$).

The R^2 values for perceived usefulness ($R^2 = 0.67$) and intention to adopt ($R^2 = 0.76$) can be considered substantial (Chin 1998). Finally, the predictive power was assessed using the PLSPredict approach (Shmueli et al. 2016) in accordance with the guidelines set forth by Shmueli et al. (2019). All the endogenous constructs exhibited $Q^2_{predict}$ values above 0 for both perceived usefulness ($Q^2_{predict} = 0.65$) and intention to adopt ($Q^2_{predict} = 0.65$). This indicates that the prediction error of the PLS-SEM results is smaller than that of using the mean values, thereby demonstrating that the PLS-SEM models offer superior predictive performance and strong predictive relevance (Shmueli et al. 2016). In conclusion, we confirm hypotheses H1a, H1c, and H2c.

5.3. Necessary condition analysis

To further investigate the relationship between interpersonal trustworthiness beliefs and user responses, we enhanced our PLS-SEM approach with NCA. Following the methodology outlined by Richter et al. (2020), we utilised the latent variable scores generated by PLS-SEM as the foundation for our NCA. The latent variable scores, specifically those for perceived ability, benevolence, integrity, usefulness, and intention to adopt, were imported into R software in accordance with the procedures detailed in Dul's (2021) NCA quick start guide.

We applied Ceiling Regression-Free Disposal Hull (CR-FDH) (Dul 2021) to analyze the scatter plots for empty spaces, effectively distinguishing between the areas with observations and those without. This analysis enabled us to determine the extent to which each interpersonal trust belief limits the levels of perceived usefulness and intention to adopt. The resulting ceiling line also delineates the minimum required level of a particular trust belief necessary to achieve specific outcomes in perceived usefulness or intention to adopt.

The primary metrics for conducting Necessary Condition Analysis (NCA) are reflected in the necessity effect size (denoted as d) derived from the Ceiling Regression-Free Disposal Hull (CR-FDH) and the

ceiling accuracy (C-accuracy). The findings are detailed in Table 5.

C-accuracy quantifies the proportion of data points that fall on or below the ceiling line, relative to the total number of observations, expressed as a percentage. Although there is no universally accepted threshold for C-accuracy, it is commonly evaluated against a standard benchmark – often 95% – for assessment purposes. In our study, none of the values surpassed this benchmark.

The necessity effect size, or *d*, calculates the ratio of the empty space above the ceiling line to the occupied space below it, with values ranging from 0 to 1. A value of *d* between 0 and less than 0.1 is indicative of a small effect, while a value between 0.1 and less than 0.3 is indicative of a medium effect. A value between 0.3 and less than 0.5 is indicative of a large effect, while a value of 0.5 or greater is indicative of a very large effect. Previous research suggests that a necessity effect size of at least 0.1, indicative of a medium effect, is generally necessary to support hypotheses concerning necessary conditions. The analysis revealed a medium necessity effect size for the relationship between perceived integrity and perceived usefulness ($d = 0.238$, C-accuracy = 96.9%), and a large effect for the relationship between perceived integrity and intention to adopt ($d = 0.30$, C-accuracy = 97.4%). This indicates that perceived integrity is a necessary condition for enhancing both perceived usefulness and intention to adopt. However, the hypotheses proposed in sections 3a, 3b, 4a, and 4b must be rejected.

To further refine our analysis, we examined the bottleneck table, which quantifies the minimum level of one variable (X) required to achieve a specified level of another variable (Y), presented as percentages.

Table 6 shows the bottleneck table for perceived integrity and perceived usefulness. It shows that to achieve the maximum level of perceived usefulness (100%), at least a medium to high level (minimum 59.8%) of perceived integrity is required. This means that unless the specified level of perceived integrity is achieved, this variable will hinder the establishment of a high level of perceived usefulness, regardless of the value of other influencing factors.

Table 5. Effect sizes of necessary condition analysis.

	Perceived Ability	
Construct	CR-FDH (d)	c-accuracy
Perceived Usefulness	0.01	100%
Intention to Adopt	0.03	99.0%
	Perceived Benevolence	
Construct	CR-FDH (d)	c-accuracy
Perceived Usefulness	0.06	100%
Intention to Adopt	0.08	98.0%
	Perceived Integrity	
Construct	CR-FDH (d)	c-accuracy
Perceived Usefulness	0.24	96.9%
Intention to Adopt	0.30	97.4%

Table 6. Bottleneck table showing the levels of integrity necessary to achieve a specific level of preceived usefulness.

Perceived Usefulness (in %)	Perceived Integrity (in %)
0	NN
10	NN
20	NN
30	7.1
40	14.7
50	22.2
60	29.7
70	37.3
80	44.8
90	52.3
100	59.8

Table 7. Bottleneck table showing the levels of integrity necessary to achieve a specific level of intention to adopt.

Intention to Adopt (in %)	Perceived Integrity (in %)
0	NN
10	5.2
20	11.1
30	17.1
40	23.0
50	29.0
60	34.9
70	40.9
80	46.8
90	52.8
100	58.7

Similarly, the bottleneck table between perceived integrity and adoption intent (Table 7) shows comparable thresholds. A medium to high level of perceived integrity (at least 58.7%) is required to achieve the maximum level of adoption intention (100%).

These findings underscore the critical role of perceived integrity in influencing both perceived usefulness and intention to adopt. Without achieving a medium to high level of perceived integrity, the levels of perceived usefulness and intention to adopt will not cross the threshold of achieving very high levels. This highlights the indispensable role of high perceived integrity in influencing key outcomes.

6. Discussion

The objective of this study was to investigate how interpersonal trustworthiness beliefs shape early trust-related user responses in human-like LLM-based health coaches. Using both probabilistic sufficiency logic and necessity logic, we examined whether users' perceptions of ability, benevolence, and integrity shape two early trust-related outcomes: perceived usefulness and intention to adopt. Our results reveal a nuanced picture of

how interpersonal trust models transfer to AI-based coaching systems.

6.1. Theoretical implications

This study contributes to research on trust in anthropomorphic AI systems by introducing a necessity-based perspective to the analysis of interpersonal trustworthiness beliefs. While prior research (e.g. Lalot and Bertram 2025; Lankton, McKnight, and Tripp 2015), has primarily examined how dimensions such as ability, benevolence, and integrity influence trust-related outcomes using sufficiency logic, our findings highlight that these dimensions may play fundamentally different roles, including functioning as non-compensable conditions.

Among the three interpersonal trustworthiness beliefs, integrity emerged as the only necessary condition for the emergence of both perceived usefulness and intention to adopt, with medium to large necessity effects. While integrity showed a statistically significant but comparatively modest sufficiency effect, its necessity indicates that it functions as a threshold condition: without a minimal perception of integrity, perceived usefulness and intention to adopt do not reach meaningful levels, regardless of how strong other beliefs may be. This finding extends prior research (e.g. Lalot and Bertram 2025; Lankton, McKnight, and Tripp 2015) by showing that integrity operates as a bottleneck condition rather than merely a stronger predictor.

Perceived ability, in contrast, was not necessary but had a clear and significant sufficiency effect. It positively influenced perceived usefulness, reinforcing the notion that users value competence in digital coaches. This finding aligns with Mayer et al.'s (1995) suggestion that trustworthiness is composed of separable but interdependent components: each trust belief can vary independently, and while some may be required, others may enhance trust without being essential. Similar observations were made by Triantafyllopoulos et al. (2024), who showed that trust among physicians toward LLM-based systems was shaped more by system clarity and competence than by relational cues, further illustrating ability's sufficiency but not necessity in trust dynamics. This complements findings by Jin and Eastin (2025), who showed that design cues enhancing perceived expertise significantly shaped trust, particularly when users experienced heightened health-related threat, highlighting the context-dependent importance of ability-related perceptions in shaping user responses.

Contrary to expectation, benevolence neither significantly influenced intention to adopt and perceived usefulness nor was it necessary. This contrasts with findings from human coaching literature, where benevolence is often seen as the most critical trust driver – particularly because it supports user autonomy and emotional safety (Schiemann et al. 2019).

In digital contexts, however, this role may differ. Our findings are consistent with prior research suggesting that when ability and integrity are established, perceptions of benevolence play a less central role in fostering trust-related user responses in human-like AI agents (Lalot and Bertram 2025). One possible interpretation is that users prioritise an AI coach's reliability and impartial, honest behaviour over relational warmth or care, especially when interacting with a non-human agent. This aligns with findings Scholich et al. (2025), who show that while chatbots can express empathy and validation, they may fall short in fostering deeper engagement, suggesting that benevolence is not evaluated in the same way as in human interactions.

At the same time, alternative explanations should be considered. For instance, the limited role of benevolence may be related to how it is perceived and attributed in AI-based systems. While users may interact with the LLM agent directly, they may assign emotional and ethical responsibilities to the developers or platform providers. In this sense, benevolence might be assessed at the system level rather than the agent level (Knote et al. 2021). This perspective is also in line with Brunswicker et al. (2025), who highlight that users' responses to empathic cues can enhance trust under certain conditions. However users may respond more critically to such cues when emotional states are complex or expectations misaligned, for instance expectations set by the chatbot platform itself, e.g. ChatGPT. While these explanations offer plausible interpretations, additional factors not captured in the present study may also account for the limited role of benevolence, warranting further investigation.

6.2. Practical implications

The findings of this study have significant practical implications, particularly regarding the key factor of perceived integrity in fostering positive user responses in LLM-based health coaching. It is therefore recommended that developers and designers prioritise the creation of LLM-based coaching apps that are perceived as having high integrity by clients. This necessitates the development of a coach character that is truthful, authentic, and provides impartial recommendations (Schiemann et al. 2019). It is critical to maintain a medium to high level of perceived integrity (at least 60%) to achieve a very high level of perceived usefulness and intention to adopt.

Once a high level of integrity has been established in the coaching application, it becomes crucial to prioritise the perceived ability of the coach, as this significantly enhances the perceived usefulness of the intervention. Although it may appear self-evident that the competence of a coaching intervention should be a primary consideration, it is important to underscore that the perceived ability extends beyond the mere helpfulness of the application. It is essential that users view the coach as an expert in the field. This perception is also conveyed through the language used by the coach (Schiemann et al. 2019). Consequently, the use of uncertain or ambiguous language should be avoided to maintain and reinforce the coach's image as a knowledgeable and reliable figure. Regarding the proposition that benevolence does not influence the formation of trust, it is necessary to clarify that we are not proposing the idea of creating coaching interventions where coaches do not have a benevolent attitude toward their users. It is of the utmost importance that users' safety and autonomy are always ensured in any health coaching intervention. Our research findings merely suggest that the lack of perceived benevolence does not hinder the formation of trust between the coach and the user.

Furthermore, our research demonstrates the significance of employing necessary condition analysis. Had we solely relied on traditional probabilistic sufficiency logic, we would have failed to recognise the impact of integrity, which had only a modest effect size. This oversight would have resulted in us overlooking a crucial obstacle for the emergence of perceived usefulness and intention to adopt in LLM-based coaching.

7. Limitations and further research

It is important to acknowledge the limitations of our research, which present opportunities for further investigation. The study was limited to an LLM-based coaching intervention for fitness and nutrition. Therefore, it is not possible to assume that the same trustworthiness beliefs are influencing evaluative and behavioural responses in other contexts. Consequently, further research should explore interpersonal trustworthiness beliefs in other health coaching scenarios. It is furthermore essential to acknowledge that the findings of our study were derived from the analysis of vignettes representing only the initial session of coaching. The impact of trustworthiness beliefs may be subject to change over time and with increased frequency of interaction. Nevertheless, the initial session is critical in establishing an initial impression and laying the foundation for ongoing user engagement. Consequently, it is of critical importance to ensure integrity from the outset if the success and effectiveness of LLM-based health coaching programmes are to be achieved. Further research should evaluate the influence of interpersonal trustworthiness beliefs in long-term LLM-based coaching scenarios. Furthermore, our research was limited to assessing interpersonal trustworthiness beliefs. While we believe that this is the most appropriate approach for generating initial insights, future research should also explore the influence of technological trustworthiness beliefs as well as those related to organisations.

In addition, the non-significant role of benevolence observed in this study should be interpreted with caution. While we offer several possible explanations, such as attribution effects within layered AI systems, other factors such as the operationalisation of benevolence or limited variance in short-term vignette settings may also have influenced this finding. Future research should further investigate the conditions under which benevolence becomes more or less salient in AI-based interactions.

The use of a vignette study introduces additional limitations to the research process. While vignette studies are an appropriate methodology for our research endeavour (Aguinis and Bradley 2014), they cannot fully substitute for the individual interaction with a prototype. Therefore, further research should assess the findings in experimental settings using a prototype. Additionally, it is important to acknowledge that the coach in question was male. Future research should consider a wider range of coaches to address potential gender biases.

Finally, NCA is a relatively novel technique, and not all issues pertaining to statistical and causal inference have been resolved. Further research on the statistical properties of estimated ceiling lines and confidence interval estimation is necessary to gain a deeper understanding of the necessity of conditions. (van der Valk et al. 2016).

8. Conclusion

The objective of our study was to examine how interpersonal trustworthiness beliefs shape early user responses to LLM-based health coaching systems. Probabilistic sufficiency and necessity logic were employed to assess trustworthiness beliefs shape user responses.

Participants were presented with a vignette of human-like LLM-based health coaching and then surveyed on formative trustworthiness belief constructs (ability, benevolence, and integrity) as well as reflective outcome constructs (perceived usefulness and intention to adopt). Our analysis using PLS-SEM demonstrated that ability is a significant factor in in LLM-based

coaching and our NCA analysis revealed that integrity is a necessary condition for eliciting positive user responses in terms of intention to adopt and perceived usefulness. Our findings contribute to both research and practice by providing a deeper understanding of the must-have and should-have factors for user engagement in LLM-based health coaching. This knowledge can be utilised to design effective health coaching interventions.

Note

1. https://www.chatbase.co/docs/integrations/whatsapp

Author contributions

CRediT: **Sophia Meywirth:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft; **Andreas Janson:** Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing; **Matthias Söllner:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) to improve the clarity, coherence, and readability of selected paragraphs. After using this tool, the authors reviewed and edited all content carefully and take full responsibility for the content of the published article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethics approval

This study was reviewed and determined to be exempt from further ethical approval by the Institutional Review Board of the University of St. Gallen.

ORCID

Sophia Meywirth https://orcid.org/0009-0005-6643-6864
Andreas Janson http://orcid.org/0000-0003-3149-0340
Matthias Söllner https://orcid.org/0000-0002-1347-8252

References

Ab Hamid, M. R., W. Sami, and M. H. Mohmad Sidek. 2017. "Discriminant Validity Assessment: Use of Fornell & Larcker Criterion versus HTMT Criterion." *Journal of Physics: Conference Series* 890:012163. https://doi.org/10.1088/1742-6596/890/1/012163.

Agarwal, R., and J. Prasad. 1998. "A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology." *Information Systems Research* 9 (2): 204–215.

Aguinis, H., and K. J. Bradley. 2014. "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17 (4): 351–371. https://doi.org/10.1177/1094428114547952.

Araujo, T. 2018. "Living up to the Chatbot Hype: The Influence of Anthropomorphic Design Cues and Communicative Agency Framing on Conversational Agent and Company Perceptions." *Computers in Human Behavior* 85:183–189. https://doi.org/10.1016/j.chb.2018.03.051.

Beinema, T., H. op den Akker, L. van Velsen, and H. Hermens. 2021. "Tailoring Coaching Strategies to Users' Motivation in a Multi-agent Health Coaching Application." *Computers in Human Behavior* 121:106787. https://doi.org/10.1016/j.chb.2021.106787.

Boyatzis, R. E., M. L. Smith, and Alim J. Beveridge. 2013. "Coaching with Compassion: Inspiring Health, Well-Being, and Development in Organizations." *The Journal of Applied Behavioral Science* 49 (2): 153–178. https://doi.org/10.1177/0021886312462236.

Brunswicker, S., Y. Zhang, C. Rashidian, and D. W. Linna. 2025. "Trust through Words: The Systemize-Empathize-Effect of Language in Task-Oriented Conversational Agents." *Computers in Human Behavior* 165:108516. https://doi.org/10.1016/j.chb.2024.108516.

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling." In *Modern Methods for Business Research*, edited by G. A. Marcoulides. New York: LEA.

Demszky, D., D. Yang, D. S. Yeager, C. J. Bryan, M. Clapper, S. Chandhok, J. C. Eichstaedt, et al. 2023. "Using Large Language Models in Psychology." *Nature Reviews Psychology* : 1–14. https://doi.org/10.1038/s44159-023-00241-5.

Dul, J. 2016. "Necessary Condition Analysis (NCA)." *Organizational Research Methods* 19 (1): 10–52. https://doi.org/10.1177/1094428115584005.

Dul, J. 2019. *Conducting Necessary Condition Analysis for Business and Management Students*. 1–160.

Dul, J. 2021. *Advances in Necessary Condition Analysis*. https://bookdown.org/ncabook/advanced_nca2/summary.html.

Dul, J. 2024. "How to Sample in Necessary Condition Analysis." *European J. of International Management* 1 (1): 10063098. https://doi.org/10.1504/EJIM.2024.10063098.

Dwivedi, Y. K., N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, et al. 2023. "So What If ChatGPT Wrote It?": Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy." *International Journal of Information Technology and Management* 71 (102642), https://doi.org/10.1016/j.ijinfomgt.2023.102642.

Epley, N., A. Waytz, and J. T. Cacioppo. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism."

Psychological Review 114 (4): 864–886. https://doi.org/10.1037/0033-295X.114.4.864.

Fadhil, A., G. Schiavo, Y. Wang, and B. A. Yilma. 2018. "The Effect of Emojis When Interacting with Conversational Interface Assisted Health Coaching System." Proceedings of the 12th EAI international conference on pervasive computing technologies for healthcare, PervasiveHealth '18, 378–383. https://doi.org/10.1145/3240925.3240965.

Feine, J., U. Gnewuch, S. Morana, and A. Maedche. 2019. "A Taxonomy of Social Cues for Conversational Agents." *International Journal of Human-Computer Studies* 132:138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009.

Fink, J. 2012. "Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction." In *Social Robotics*, Vol. 7621, edited by S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, and M.-A. Williams, 199–208. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-34103-8_20.

Gefen, D. 2002. "Customer Loyalty in E-Commerce." *Journal of the Association for Information Systems* 3 (1), https://doi.org/10.17705/1jais.00022.

Gefen, D., E. Karahanna, and D. W. Straub. 2003. "Trust and TAM in Online Shopping: An Integrated Model." *MIS Quarterly* 27 (1): 51. https://doi.org/10.2307/30036519.

Greaves, C. J., K. E. Sheppard, C. Abraham, W. Hardeman, M. Roden, P. H. Evans, P. Schwarz, and The IMAGE Study Group. 2011. "Systematic Review of Reviews of Intervention Components Associated with Increased Effectiveness in Dietary and Physical Activity Interventions." *BMC Public Health* 11 (1): 119. https://doi.org/10.1186/1471-2458-11-119.

Gulati, S., J. McDonagh, S. Sousa, and D. Lamas. 2024. "Trust Models and Theories in Human–Computer Interaction: A Systematic Literature Review." *Computers in Human Behavior Reports* 16:100495. https://doi.org/10.1016/j.chbr.2024.100495.

Hair, J. F., G. T. M. Hult, C. M. Ringle, and M. Sarstedt. 2017. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (Second Edition)*. Los Angeles: SAGE.

Hair, J., G. T. M. Hult, C. Ringle, and M. Sarstedt. 2022. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM).

Hair, J. F., C. M. Ringle, and M. Sarstedt. 2011. "PLS-SEM: Indeed a Silver Bullet." *Journal of Marketing Theory and Practice* 19 (2): 139–152. https://doi.org/10.2753/MTP1069-6679190202.

Henseler, J., C. M. Ringle, and R. R. Sinkovics. 2009. "The use of Partial Least Squares Path Modeling in International Marketing." In *New Challenges to International Marketing*, edited by R. R. Sinkovics and P. N. Ghauri, 277–319. Emerald Group Publishing Limited. https://doi.org/10.1108/S1474-7979(2009)0000020014.

Janson, A. 2023. "How to Leverage Anthropomorphism for Chatbot Service Interfaces: The Interplay of Communication Style and Personification." *Computers in Human Behavior* 149:107954. https://doi.org/10.1016/j.chb.2023.107954.

Jin, E., and M. Eastin. 2025. "Towards More Trusted Virtual Physicians: The Combinative Effects of Healthcare Chatbot Design Cues and Threat Perception on Health Information Trust." *Behaviour and Information Technology* 44 (4): 829–842.

Jörke, M., S. Sapkota, L. Warkenthien, N. Vainio, P. Schmiedmayer, E. Brunskill, and J. Landay. 2024. *Supporting Physical Activity Behavior Change with LLM-Based Conversational Agents* (arXiv:2405.06061). arXiv. https://doi.org/10.48550/arXiv.2405.06061.

Kamis, A., M. Koufaris, and T. Stern. 2008. "Using an Attribute-Based Decision Support System for User-Customized Products Online: An Experimental Investigation." *MIS Quarterly* 32 (1): 159–177.

Kettunen, E., T. Kari, W. Critchley, and L. Frank. 2024. "Critical Experiences with Sport and Wellness Technology Digital Coach – Differences between Young Adults and Young Elderly." *Behaviour & Information Technology* 43 (12): 3010–3026. https://doi.org/10.1080/0144929X.2023.2267692.

Kim, J., and I. Im. 2023. "Anthropomorphic Response: Understanding Interactions between Humans and Artificial Intelligence Agents." *Computers in Human Behavior* 139:107512. https://doi.org/10.1016/j.chb.2022.107512.

Kim, Y., and S. S. Sundar. 2012. "Anthropomorphism of Computers: Is It Mindful or Mindless?" *Computers in Human Behavior* 28 (1): 241–250. https://doi.org/10.1016/j.chb.2011.09.006.

Klingbeil, A., C. Grützner, and P. Schreck. 2024. "Trust and Reliance on AI — an Experimental Study on the Extent and Costs of Overreliance on AI." *Computers in Human Behavior* 160:108352. https://doi.org/10.1016/j.chb.2024.108352.

Knote, R., A. Janson, M. Söllner, and J. M. Leimeister. 2021. "Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants." *Journal of the Association for Information Systems* 22:418–458. https://doi.org/10.17705/1jais.00667.

Kohnke, L., and B. L. Moorhouse. 2025. "Enhancing the Emotional Aspects of Language Education through Generative Artificial Intelligence (GenAI): A Qualitative Investigation." *Computers in Human Behavior* 167:108600. https://doi.org/10.1016/j.chb.2025.108600.

Lalot, F., and A.-M. Bertram. 2025. "When the bot Walks the Talk: Investigating the Foundations of Trust in an Artificial Intelligence (AI) Chatbot." *Journal of Experimental Psychology: General* 154 (2): 533–551. https://doi.org/10.1037/xge0001696.

Lankton, N., D. H. McKnight, and J. Tripp. 2015. "Technology, Humanness, and Trust: Rethinking Trust in Technology." *Journal of the Association for Information Systems* 16 (10): 880–918. https://doi.org/10.17705/1jais.00411.

Lee, S.-L., I. Lau, S. Kiesler, and C. Y. Chiu. 2005. *Human Mental Models of Humanoid Robots* (p. 2772). https://doi.org/10.1109/ROBOT.2005.1570532.

Lee, J. D., and K. A. See. 2004. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors* 46 (1): 50–80. https://doi.org/10.1518/hfes.46.1.50_30392.

Lowry, P. B., J. D'Arcy, B. Hammer, and G. D. Moody. 2016. "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." *The Journal of Strategic Information Systems* 25 (3): 232–240. https://doi.org/10.1016/j.jsis.2016.06.002.

Mayer, R. C., J. H. Davis, and F. D. Schoorman. 1995. "An Integrative Model of Organizational Trust." *The Academy of Management Review* 20 (3): 709–734. https://doi.org/10.2307/258792.

Mayer, C. J., J. Mahal, D. Geisel, E. J. Geiger, E. Staatz, M. Zappel, S. P. Lerch, J. C. Ehrenthal, S. Walter, and B. Ditzen. 2024. "User Preferences and Trust in Hypothetical Analog, Digitalized and AI-Based Medical Consultation Scenarios: An Online Discrete Choice Survey." *Computers in Human Behavior* 161:108419. https://doi.org/10.1016/j.chb.2024.108419.

McKnight, D. H., M. Carter, J. B. Thatcher, and P. F. Clay. 2011. "Trust in a Specific Technology: An Investigation of Its Components and Measures." *ACM Transactions on Management Information Systems* 2 (2): 1–25. https://doi.org/10.1145/1985347.1985353.

McKnight, D. H., and N. L. Chervany. 2001. "What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology." *International Journal of Electronic Commerce* 6 (2): 35–59.

McQueen, A., M. L. Imming, T. Thompson, R. Garg, T. Poor, and M. W. Kreuter. 2020. "Client Perspectives on Health Coaching: Insight for Improved Program Design." *American Journal of Health Behavior* 44 (5): 591–602. https://doi.org/10.5993/AJHB.44.5.4.

Michie, S., M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. Eccles, J. Cane, and C. Wood. 2013. "The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions." *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine* : 46. https://doi.org/10.1007/s12160-013-9486-6.

Mitchell, E. G., R. Maimone, A. Cassells, J. N. Tobin, P. Davidson, A. M. Smaldone, and L. Mamykina. 2021. "Automated vs. Human Health Coaching: Exploring Participant and Practitioner Experiences." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1): 99:1–99:37. https://doi.org/10.1145/3449173.

Nadarzynski, T., O. Miles, A. Cowie, and D. Ridge. 2019. "Acceptability of Artificial Intelligence (AI)-led Chatbot Services in Healthcare: A Mixed-Methods Study." *Digital Health* 5:2055207619871808. https://doi.org/10.1177/2055207619871808.

Nass, C., and Y. Moon. 2000. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56 (1): 81–103. https://doi.org/10.1111/0022-4537.00153.

Ni, Y., and F. Jia. 2025. "A Scoping Review of AI-Driven Digital Interventions in Mental Health Care: Mapping Applications across Screening, Support, Monitoring, Prevention, and Clinical Education." *Healthcare* 13 (10): 1205. https://doi.org/10.3390/healthcare13101205.

Ötting, S. K., and G. W. Maier. 2018. "The Importance of Procedural Justice in Human–Machine Interactions: Intelligent Systems as new Decision Agents in Organizations." *Computers in Human Behavior* 89:27–39. https://doi.org/10.1016/j.chb.2018.07.022.

Palan, S., and C. Schitter. 2018. "Prolific.ac—A Subject Pool for Online Experiments." *Journal of Behavioral and Experimental Finance* 17:22–27. https://doi.org/10.1016/j.jbef.2017.12.004.

Podsakoff, P. M., S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88 (5): 879–903. https://doi.org/10.1037/0021-9010.88.5.879.

Podsakoff, P. M., N. P. Podsakoff, L. J. Williams, C. Huang, and J. Yang. 2024. "Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix." *Annual Review of Organizational Psychology and Organizational Behavior* 11:17–61. https://doi.org/10.1146/annurev-orgpsych-110721-040030.

Richter, N. F., S. Schubring, S. Hauff, C. M. Ringle, and M. Sarstedt. 2020. "When Predictors of Outcomes Are Necessary: Guidelines for the Combined use of PLS-SEM and NCA." *Industrial Management & Data Systems* 120 (12): 2243–2267. https://doi.org/10.1108/IMDS-11-2019-0638.

Rousseau, D., S. Sitkin, R. Burt, and C. Camerer. 1998. "Not So Different after All: A Cross-Discipline View of Trust." *Academy of Management Review* 23), https://doi.org/10.5465/AMR.1998.926617.

Ryan, K., S. Dockray, and C. Linehan. 2022. "Understanding How EHealth Coaches Tailor Support for Weight Loss: Towards the Design of Person-Centered Coaching Systems." *CHI Conference on Human Factors in Computing Systems* : 1–16. https://doi.org/10.1145/3491102.3501864.

Schiemann, S. J., C. Mühlberger, F. D. Schoorman, and E. Jonas. 2019. "Trust Me, I Am a Caring Coach: The Benefits of Establishing Trustworthiness during Coaching by Communicating Benevolence." *Journal of Trust Research* 9 (2): 164–184. https://doi.org/10.1080/21515581.2019.1650751.

Scholich, T., M. Barr, S. W. Stirman, and S. Raj. 2025. "A Comparison of Responses from Human Therapists and Large Language Model–Based Chatbots to Assess Therapeutic Communication: Mixed Methods Study." *JMIR Mental Health* 12 (1): e69709. https://doi.org/10.2196/69709.

Schöbel, S., M. Söllner, and J. M. Leimeister. 2016. "The Agony of Choice – Analyzing User Preferences regarding Gamification Elements in Learning Management Systems." Icis 2016 proceedings https://aisel.aisnet.org/icis2016/HumanBehavior/Presentations/10.

Shmueli, G., S. Ray, J. M. Velasquez Estrada, and S. B. Chatla. 2016. "The Elephant in the Room: Predictive Performance of PLS Models." *Journal of Business Research* 69 (10): 4552–4564. https://doi.org/10.1016/j.jbusres.2016.03.049.

Shmueli, G., M. Sarstedt, J. F. Hair, J.-H. Cheah, H. Ting, S. Vaithilingam, and C. M. Ringle. 2019. "Predictive Model Assessment in PLS-SEM: Guidelines for Using PLSpredict." *European Journal of Marketing* 53 (11): 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189.

Söllner, M., A. Hoffmann, H. Hoffmann, A. Wacker, and J. M. Leimeister. 2012. "Understanding the Formation of Trust in IT Artifacts." International conference on information systems (ICIS) 2012, Orlando, FL, USA. http://aisel.aisnet.org/icis2012/proceedings/HumanBehavior/11

Söllner, M., A. Hoffmann, and J. M. Leimeister. 2016. "Why Different Trust Relationships Matter for Information

Systems Users." *European Journal of Information Systems* 25 (3): 274–287. https://doi.org/10.1057/ejis.2015.17.

Söllner, M., P. A. Pavlou, and J. M. Leimeister. 2013. *Understanding Trust in IT Artifacts – A New Conceptual Approach* (SSRN Scholarly Paper No. 2475382). https://doi.org/10.2139/ssrn.2475382.

Stein, J.-P., and K. F. MacDorman. 2024. "After Confronting One Uncanny Valley, Another Awaits." *Nature Reviews Electrical Engineering* 1 (5): 276–277. https://doi.org/10.1038/s44287-024-00041-w.

Terblanche, N. H. D., and M. Heyns. 2020. "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 Journal of Industrial Psychology* 46 (1): 1–11. https://doi.org/10.4102/sajip.v46i0.1707.

Thatcher, J. B., D. H. McKnight, E. W. Baker, R. E. Arsal, and N. H. Roberts. 2011. "The Role of Trust in Postadoption IT Exploration: An Empirical Examination of Knowledge Management Systems." *IEEE Transactions on Engineering Management* 58 (1): 56–70. https://doi.org/10.1109/TEM.2009.2028320.

Thiebes, S., S. Lins, and A. Sunyaev. 2021. "Trustworthy Artificial Intelligence." *Electronic Markets* 31 (2): 447–464. https://doi.org/10.1007/s12525-020-00441-4.

Thom, D. H., J. Wolf, H. Gardner, D. DeVore, M. Lin, A. Ma, A. Ibarra-Castro, and G. Saba. 2016. "A Qualitative Study of How Health Coaches Support Patients in Making Health-Related Decisions and Behavioral Changes." *Annals of Family Medicine* 14 (6): 509–516. https://doi.org/10.1370/afm.1988.

Triantafyllopoulos, L., G. Feretzakis, L. Tzelves, A. Sakagianni, V. S. Verykios, and D. Kalles. 2024. "Evaluating the Interactions of Medical Doctors with Chatbots Based on Large Language Models." *Computers in Human Behavior* 161 (C), https://doi.org/10.1016/j.chb.2024.108404.

van der Valk, W., R. Sumo, J. Dul, and R. Schroeder. 2016. "When Are Contracts and Trust Necessary for Innovation in Buyer-Supplier Relationships? A Necessary Condition Analysis." *Journal of Purchasing and Supply Management*, https://doi.org/10.1016/j.pursup.2016.06.005.

Vardhan, M., N. Hegde, D. Nathani, E. Rosenzweig, A. Karthikesalingam, and M. Seneviratne. 2023. "Infusing Behavior Science into Large Language Models for Activity Coaching [Preprint]." *Health Systems and Quality Improvement*, https://doi.org/10.1101/2023.03.31.23287995.

Wang, W., and I. Benbasat. 2005. "Trust in and Adoption of Online Recommendation Agents." *Journal of the Association for Information Systems* 6 (3): 72–101. https://doi.org/10.17705/1jais.00065.

Warburton, D. E. R., C. W. Nicol, and S. S. D. Bredin. 2006. "Health Benefits of Physical Activity: The Evidence." *Canadian Medical Association Journal* 174 (6): 801–809. https://doi.org/10.1503/cmaj.051351.

WHO. 2025, August 12. *Obesity and Overweight*. https://www.Who.int/news-room/fact-sheets/detail/obesity-and-overweight.

Wold, H. O. A. 1982. "Soft Modeling: The Basic Design and Some Extensions." In *Systems under Indirect Observation*, edited by K. G. Jöreskog and H. O. A. Wold, 1–54. North-Holland, Amsterdam.

Wu, D., P. B. Lowry, D. Zhang, and Y. Tao. 2022. "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." *Journal of Medical Internet Research* 24 (12): e42941. https://doi.org/10.2196/42941.

Zhang, D., and X. Zhao. 2024. "Understanding Adoption Intention of Virtual Medical Consultation Systems: Perceptions of ChatGPT and Satisfaction with Doctors." *Computers in Human Behavior* 159:108359. https://doi.org/10.1016/j.chb.2024.108359.