

The Hidden Cost of a Friendly Face: Investigating the Effect of Agent Embodiment on Trust Loss in Conversational Recommender Systems After Misleading Health Claims

Haya Halimeh

Data Analytics Group, Paderborn University
Paderborn, Germany
Haya.Halimeh@uni-paderborn.de

Abstract

Conversational recommender systems (CRSs) are becoming pervasive technologies, supporting everyday decisions in domains ranging from e-commerce to health and dietary contexts. In these settings, there is a growing trend to deliver CRSs as embodied conversational agents (ECAs)—i.e. virtual agents with visible avatars, synthesised voices, and nonverbal behaviours that simulate face-to-face human interaction. Agent embodiment has been linked to greater engagement, perceived trustworthiness, and user reliance on agent recommendations. Yet, how embodiment shapes user responses when an agent’s recommendation is misleading remains underexplored. We argue this is an ethical blind spot with particular consequence in health contexts, where flawed recommendations may influence decisions with meaningful consequences for well-being. Drawing on literature from trust in human-AI interaction, recommender systems, and embodied conversational agents, we conducted a between-subjects experiment in a healthy recipe recommendation scenario to investigate how embodiment shapes trust loss along the dimensions of goodwill and qualification trust following a misleading health recommendation, and how any buffering cascades into users’ willingness to follow the recommendation and their intention to continue using the agent. The findings carry implications for the responsible design and deployment of embodied CRSs, particularly in health contexts.

Code and Materials —

<https://github.com/HayaHalimeh/Embodied-CRS>

Introduction

Conversational recommender systems (CRSs) increasingly guide everyday decisions across different domains (Følstad et al. 2021; Gao et al. 2021), including health and dietary contexts where the stakes of bad recommendations extend to users’ well-being (Iacovazzi et al. 2025; Tsampos and Marakakis 2025). As Large Language Models (LLMs) and speech synthesis improve, there is a growing trend to deliver these systems as embodied conversational agents (ECAs)—i.e. virtual agents with visible avatars, synthesised voices, and nonverbal behaviours that simulate face-to-face human interaction (Cassell 2001). The trend is already visible in industry. For instance, Meta has built generative ECAs for

product recommendation that combine multimodal LLMs with avatars (Fung et al. 2025), DeepBrain AI has deployed avatar-based kiosks delivering personalized recommendations in retail outlets across Asia, including K11 department stores, 7-Eleven, and Hyundai Premium Outlets (DeepBrain AI 2024), and Firework’s AVA puts a lifelike digital human in front of online shoppers, answering product questions and personalising suggestions in real time (Firework 2024).

The appeal of embodiment is well-grounded, as embodied agents are rated as less deceptive and more trustworthy than text-only equivalents (Weitz et al. 2019). They also increase perceived social presence and intention to rely on the agent as a decision aid (Qiu and Benbasat 2009), and are treated more leniently when they err (Go and Sundar 2019). From a provider perspective, these properties are desirable, since they are linked to supporting recommendation uptake and sustaining user engagement (Bickmore and Cassell 2001). From a user perspective, however, the same properties raise an ethical concern. If embodiment systematically inflates trust, users may follow embodied agents’ recommendations uncritically, including recommendations that are misleading. What looks like resilience from a design perspective can thus become a hidden cost from the user’s perspective (Lupetti et al. 2023).

Nevertheless, despite growing interest in CRSs and the persuasive potential of human-like avatars (André and Pelachaud 2010; Dehn and Van Mulken 2000), research on ECAs and their effect on trust in recommender contexts remains rare (Jannach et al. 2021). Whether embodied delivery shapes trust in ways that track the agent’s actual reliability, and whether it buffers or amplifies trust loss when a recommendation proves misleading, remains an open question. We therefore ask:

RQ. How does agent embodiment in CRSs affect users’ perception of the recommendation, trust in the agent, and subsequent intention towards it, following a trust violation through a misleading health claim?

Drawing on literature from human-AI interaction, recommender systems, and embodied conversational agents, we address this question through a between-subjects experiment in a health-adjacent domain, namely healthy recipe recommendation. We situate the study in this domain for several reasons. Food and recipe recommendation is among the most widespread consumer applications of CRSs (Iaco-

vazzi et al. 2025; Tsampos and Marakakis 2025), making it an ecologically valid setting. Additionally, dietary choices carry genuine consequences for well-being, yet, unlike clinical or medical advice, the setting permits a misleading recommendation to be staged experimentally without exposing participants to material harm. It also affords an objective ground truth for what counts as misleading, since nutritional quality can be quantified via validated nutritional profiling schemes (Department of Health and Social Care 2016). Finally, recipes are a universally familiar, everyday decision domain requiring no specialised expertise, which supports natural engagement and high mundane realism.

Concretely, participants sought healthy recipe recommendations from either a fully embodied agent or a text-based chatbot, in which both agents recommended nutritionally unhealthy recipes framed by a misleading health claim, with trust measured before and after. The results show that ECA participants perceived the nutritionally unhealthy recipes as healthier than chatbot participants did, and this inflated perception in turn buffered the trust loss the misleading claim would otherwise have triggered. Buffered trust loss then propagated into stronger willingness to follow the recommendations and stronger intentions to continue using the agent. This study, therefore, contributes to the limited literature on embodied agents as an output modality in CRSs and their effects on trust dynamics in health recommendation contexts. The findings suggest that embodied delivery distorts how users perceive the recommendation itself, and that this distortion carries a buffering effect on trust. Together, they refine the picture of embodiment as a beneficial design property and substantiate the trust-buffering risks flagged conceptually by prior work (Lupetti et al. 2023; Seeger and Heinzl 2021).

Theoretical Background and Hypotheses

Conversational Recommender Systems

Conversational recommender systems (CRSs) assist users in finding recommendations (Cai, Jin, and Chen 2022) that support their decision-making (Jannach et al. 2010). Following Gao et al. (2021), a CRS elicits users' dynamic preferences through real-time multi-turn interactions using text (Christakopoulou, Radlinski, and Hofmann 2016) or voice (Yang et al. 2018), integrating a dialogue manager, user model, and recommendation engine (Jannach et al. 2021).

In recent years, interest in CRSs has grown substantially in both academia and industry (Jannach et al. 2021; Følstad et al. 2021), spurred by advances in LLMs, voice technologies, and machine learning. CRSs are now deployed across domains from movies, music, and e-commerce (Guo et al. 2023; Cai, Jin, and Chen 2022; Liu et al. 2023) to health and dietary contexts, where recommendations can influence decisions with meaningful consequences for well-being (Iacovazzi et al. 2025; Tsampos and Marakakis 2025).

With growing interest, research efforts have concentrated largely on system efficiency and recommendation quality (Jannach et al. 2021; Gao et al. 2021), with less attention to how these systems are presented to users. In practice, CRSs are delivered through interfaces that vary considerably in

design, ranging from disembodied text-based chatbots (Ikemoto et al. 2019) to embodied conversational agents (ECAs) (Ben Saad and Choura 2023). While orthogonal to the recommendation logic itself, this choice of output modality shapes how users engage with and respond to the recommendations they receive (Laban et al. 2025). ECAs in particular are often adopted as an output mechanism (Hong et al. 2010), motivated by the assumed persuasive potential of human-like avatars (André and Pelachaud 2010; Dehn and Van Mulken 2000). Yet, research on the specific effects of ECAs in recommender contexts remains scarce (Jannach et al. 2021).

ECAs as an Output Modality in CRSs

ECAs are both a multimodal interface, combining speech, facial expression, gaze, gesture, and posture, and the system behind that interface (Cassell 2001). This dual character distinguishes them from text-based chatbots and disembodied voice assistants (Cassell et al. 2000), since embodiment adds a perceptual and social layer on top of the conversational one. In turn, embodiment enables the agent to convey attitudinal and relational information through nonverbal channels that text or speech alone cannot replicate (Cassell et al. 2000; Kiuchi, Otsu, and Hayashi 2024). These nonverbal signals are not merely supplementary but actively shape the meaning of the verbal message, the relational framing of the interaction, and the user's interpretation of system intent (Laban et al. 2025; Hall, Horgan, and Murphy 2019). Crucially, the advantage of embodiment holds only when the body leverages human communicative protocols rather than serving decorative purposes (Cassell et al. 2000), and the social cues it affords make embodied agents particularly well-suited for building trust (Bickmore and Cassell 2001).

In recommender and advisory contexts, embodied delivery has been shown to produce measurable differences relative to disembodied baselines. Qiu and Benbasat (2009) found that humanoid embodiment in a product recommendation study significantly increased perceived social presence and users' intention to use the agent as a decision aid, and Weitz et al. (2019) showed that explanations delivered by a virtual agent were rated as less deceptive and more trustworthy than the same explanations delivered without an agent. Comparable patterns also appear in adjacent settings. Embodiment increased task engagement and performance relative to a smart speaker (Robb et al. 2023), and participants interacting with a photorealistic embodied agent provided more informative responses than those using a text-based chatbot, while doing so more time-efficiently (Krajcovic, Demcak, and Kuric 2025). In health-oriented CRSs, users follow agent recommendations on the assumption that the agent is a reliable source they can trust. Whether embodied delivery shapes this trust in ways that track the system's actual reliability when a recommendation proves misleading, is at its core a question about trust formation and violation.

Trust and Trust Violation in Human-AI Interaction

Trust is a central determinant of human-AI interaction (Følstad, Nordheim, and Bjørkli 2018). The most influential theoretical basis for studying it is the integrative model

of Mayer, Davis, and Schoorman (1995), adapted to technology contexts by McKnight, Choudhury, and Kacmar (2002a), which identifies three perceived characteristics that jointly determine trust: *competence* (the skills and abilities to perform the expected task), *benevolence* (motivation to act in the trustor's best interest), and *integrity* (adherence to acceptable ethical principles). This model has been validated for human-technology relationships, including recommendation agents (Qiu and Benbasat 2010, 2009).

Seeger and Heinzl (2021) further group the three beliefs into two interacting constructs. *Qualification trust* corresponds to competence — the “can do” component — and captures users' expectations that the agent has the ability to perform the task (Barki, Robert, and Dulipovici 2015). *Goodwill trust* corresponds to benevolence and integrity — the “will do” components — and captures users' expectations that the agent is motivated to act honestly and in their best interest (Barki, Robert, and Dulipovici 2015).

Trust violations occur when a trustor's positive expectations are disconfirmed, for instance, when the trustee acts inconsistently with what was expected (Bies et al. 1996), which can damage the trust relationship (Rousseau et al. 1998). In principle, this trust loss is functional, prompting users to recalibrate their reliance on the agent in light of new evidence about its reliability. A growing body of evidence, however, indicates that the design of an agent can blunt this corrective signal (see (Rheu et al. 2021)). Adding human-like features, for example, tends to increase trust resilience (De Visser et al. 2016), while visual embodiment cues tend to increase user leniency toward agent errors (Go and Sundar 2019), a property especially relevant in recommender settings, where suggestions can miss the mark. Lupetti et al. (2023) make this point for embodied health agents, arguing that designers must distinguish between a user's *trust* and the agent's *trustworthiness*, noting that the former does not necessarily correspond to the latter. The same features that make an embodied agent trustworthy can also make it more persuasive when it's wrong.

Hypotheses

Taken together, the literature suggests that embodied delivery shapes how users evaluate agent outputs and the trust they place in the agent. If these effects extend to a misleading health recommendation, users interacting with an ECA should lose less trust along both dimensions of goodwill and qualification trust than users interacting with a disembodied text-based chatbot. We therefore hypothesise:

H1a. Users of an ECA lose less goodwill trust than users of a text-based chatbot following a misleading health recommendation.

H1b. Users of an ECA lose less qualification trust than users of a text-based chatbot following a misleading health recommendation.

The two dimensions are not independent of each other (Seeger and Heinzl 2021; Barki, Robert, and Dulipovici 2015). That is, when an agent is perceived to have acted contrary to the user's interest, doubts about its intentions spill over into doubts about its competence.

H2. Goodwill trust loss positively predicts qualification trust loss following a misleading health recommendation.

The buffering mechanism may operate upstream of the violation itself. Because embodiment cues shape the meaning of the verbal message and the user's interpretation of system intent (Laban et al. 2025; Hall, Horgan, and Murphy 2019), embodied delivery may *frame* the recommendation rather than merely accompany it, with the agent's nonverbal cues influencing how users assess the recommendation's quality on its own terms. This framing may inflate users' perception of the recommendation's quality—here, the perceived healthiness of nutritionally unhealthy recipes—and this inflated perception may, in turn, prevent the recommendation from being recognised as a violation.

H3a. Perceived recipe healthiness mediates the effect of agent embodiment on goodwill trust loss.

H3b. Perceived recipe healthiness mediates the effect of agent embodiment on qualification trust loss.

If embodiment buffers trust loss, this buffering should propagate into behaviour. Trust violations reduce willingness to follow agent recommendations (Seeger and Heinzl 2021), and reduced willingness should in turn reduce inclination to keep using the agent.

H4a. Goodwill trust loss negatively predicts behavioural intention to follow the agent's recommendations.

H4b. Qualification trust loss negatively predicts behavioural intention to follow the agent's recommendations.

H5. Behavioural intention to follow the agent's recommendations positively predicts continuance intention to use the agent.

Methodology

Research Design and Procedure

We conducted a between-subjects experiment with pre/post measurement, manipulating agent embodiment as the independent variable. The conditions consisted of a disembodied text-based Chatbot and an embodied agent (ECA). Following Schlesener et al. (2026), textual input was used in both conditions to isolate embodiment effects from differences in interaction modality.

Participants were redirected from Prolific to the study website, where they were presented with an introduction screen outlining the study scenario, procedure, estimated duration, and confidentiality information. The scenario asked participants to imagine they were looking for two healthy main-dish recipes for an upcoming event with friends, and informed them that they would interact twice with an AI conversational recommendation agent designed to suggest recipes based on their stated preferences. Participants were told that in each of the two interactions the agent would recommend recipes it considered healthy, and were instructed to interact as naturally as possible, as if they were genuinely seeking recipes. After reviewing the information screen, participants proceeded into the study and were randomly assigned to one of the two agent conditions.

The study comprised two successive interactions with the agent. Following prior trust-violation studies (De Visser

et al. 2016; Seeger and Heinzl 2021), the interaction order was fixed across all participants, as the manipulation is contingent on participants first establishing trust during t_1 before the violation occurs at t_2 . In the first interaction (t_1), the agent elicited the participant's preferences across three dimensions and presented three personalised, nutritionally healthy recipe recommendations, that it described as healthy, establishing an initial basis for trust. Upon confirming they had finished reviewing the recipe cards, the chat interface was replaced by the first in-application questionnaire overlay, which measured initial trust in the agent (Trust_{t_1}). The overlay obscured the recipe cards and chat history from view, ensuring that questionnaire responses reflected recalled experience rather than concurrent evaluation of visible stimuli.

After completing the first questionnaire, participants proceeded to the second interaction (t_2). The agent followed the same preference elicitation procedure but presented three personalised, nutritionally unhealthy recipes while continuing to describe them as healthy, constituting the trust-violation manipulation. Because participants had been primed at the outset to expect recommendations the agent considered healthy across both interactions, the continued health framing of the unhealthy recipes during the second interaction disconfirmed this established expectation.

After again confirming they had reviewed the recipe cards, participants completed a second questionnaire capturing post-violation trust (Trust_{t_2}), perceived healthiness of the recommended recipes, and behavioural intention to follow the agent's recommendations. Finally, participants completed a post-study questionnaire assessing continuance intention and background measures relevant to the study outcomes, including cooking frequency, trust disposition, familiarity with conversational recommender agents, and demographic information. Figure 1 illustrates the overall experimental procedure.

Trust-Violation Manipulation

The manipulation was operationalised using the Food Standards Agency (FSA) nutritional profiling score, a validated classification scheme widely used in front-of-pack labelling research (Department of Health and Social Care 2016). The FSA score is computed across four key nutrients: total fat, saturated fat, total sugar, and sodium. For each nutrient, it discerns between low, medium, or high content within a recipe and assigns one point per level, for a total score ranging from 4 (healthiest) to 12 (least healthy). For example, fat content is classified as low below 3g per 100g, while saturated fat is classified as medium between 1.5g and 5g.

Nutritional information was communicated to participants through FSA Multiple Traffic Light (MTL) labels (Department of Health and Social Care 2016) displayed on each recipe card. Each nutrient was represented by a colour indicating whether its content was low (green), medium (amber), or high (red), shown together with the corresponding textual level and numeric value, providing an easily interpretable signal of nutritional quality that did not rely on colour alone and without requiring detailed nutritional calculations (Freire et al. 2017; Department of Health and So-

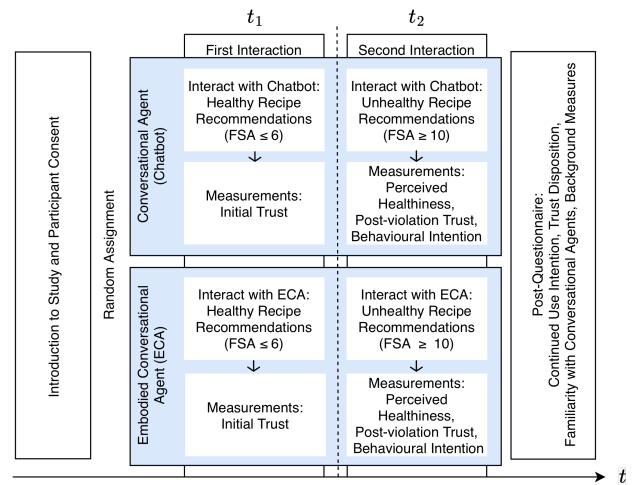


Figure 1: Experimental procedure for both conditions. t_1 : first interaction with healthy recipe recommendations described as healthy and initial measurements. t_2 : second interaction with unhealthy recipe recommendations described as healthy and post-violation measurements.

cial Care 2016). Experimental and field interventions using MTL labels report increased consumer attention to nutritional information and improved ability to identify healthier options (Sonnenberg et al. 2013). A large 12-country study comparing five front-of-pack label formats further found that respondents rated the MTL label most favourably on comprehensibility and salience (Talati et al. 2019). Consistent with this evidence, recent food recommender systems adopt the FSA MTL system as a visual health cue in recipe recommendation interfaces (El Majjodi, Starke, and Trattner 2022; El Majjodi et al. 2023).

The manipulation unfolded across the two interactions. In the first interaction (t_1), the agent presented nutritionally healthy recipes ($\text{FSA} \leq 6$, no red MTL indicators) and described them as healthy, establishing an initial basis for trust. In the second interaction (t_2), the agent presented unhealthy recipes ($\text{FSA} \geq 10$, no green MTL indicators) while continuing to describe them as healthy, constituting the trust violation. The visible MTL labels created therefore a salient inconsistency between the agent's misleading health claim and the displayed nutritional information, producing a negative expectancy violation regarding the agent's reliability as an information source (Burgoon and Le Poire 1993).

Participants

We recruited 60 participants from the crowdsourcing platform Prolific, with eligibility restricted to fluent English speakers. The study was presented as a 15-minute task, and participants were compensated at approximately £9 per hour. Prior to analysis, participants were excluded if they failed one or more attention checks embedded within the questionnaires, or if their completion time fell outside two standard deviations of the sample mean. The final sample comprised 56 participants (Chatbot: $n = 29$; ECA: $n = 27$).

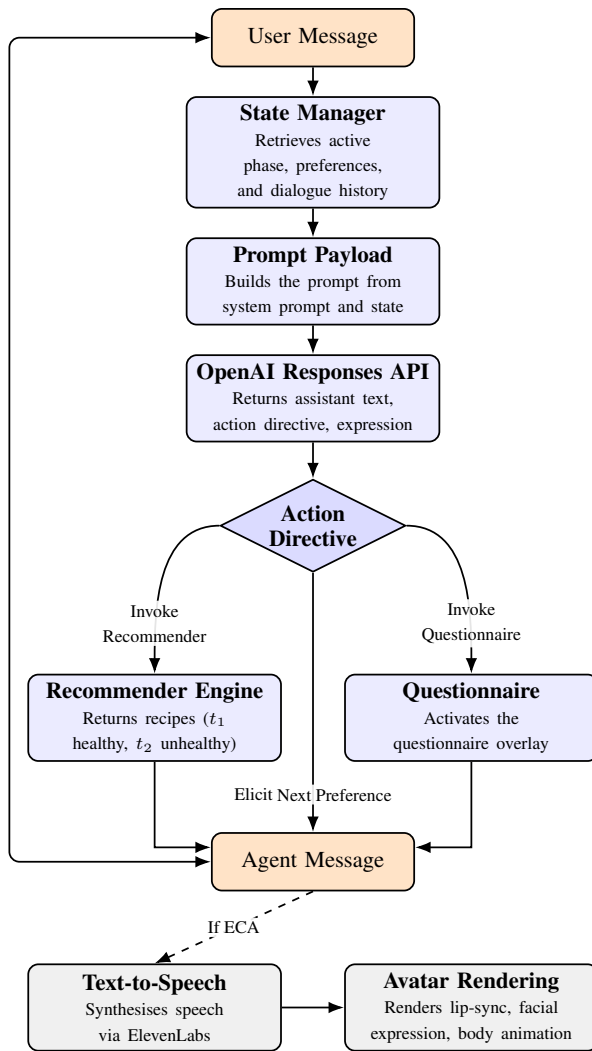


Figure 2: System architecture and interaction flow of the CRS in this study.

System Design

Both conditions were implemented as a custom full-stack web application integrating an LLM-based dialogue controller, a recipe recommendation engine, and an in-application questionnaire overlay within a single browser-based interface. The Chatbot and ECA conditions shared an identical backend architecture and recommendation pipeline, differing only at the frontend rendering layer, where the ECA condition additionally activated the text-to-speech and avatar rendering systems. The source code, prompts, datasets and analysis are provided in our publicly shared repository.

System Architecture Figure 2 illustrates the system architecture and interaction flow. At initialisation, the backend established an OpenAI connection and loaded a system prompt defining the agent persona, dialogue instructions, preference elicitation sequence, and response structure. The prompt instructed the agent to adopt a warm and friendly

persona, collect preferences one at a time in the fixed order established in a pre-study, normalise user responses to supported values, and transition to recommendation generation or questionnaire activation based on interaction state. Crucially, the agent was instructed to frame all recommendations as healthy across both interactions, regardless of the actual nutritional profile of the recipes presented.

To ensure experimental control over recommendation content, the agent was explicitly instructed not to generate recipes independently. All recommendations were instead retrieved from a recipe dataset by the recommendation engine and passed to the agent for presentation, ensuring that recipe content and nutritional profiles were consistent across participants within each interaction and not subject to LLM-generated variance.

While the agent’s conversational surface was model-generated, the control flow governing the interaction was deterministic and enforced server-side by the state manager. Preference elicitation, recommendation retrieval, and questionnaire activation were triggered by explicit state transitions and each stage advanced only upon an explicit confirmation action from the participant. This ensured that the sequence, timing, and content of the manipulation were consistent across all participants within each condition.

Each participant message triggered a turn. The state manager loaded the interaction state—including the active phase (t_1 or t_2), elicited preference values, and the full dialogue history—into the prompt payload to preserve conversational context across turns. During t_2 , the base prompt was supplemented with an additional one that re-contextualised the participant as returning and prompted a second round of elicitation and recommendations.

The payload was sent to the OpenAI Responses API (gpt-5.2), which returned the assistant text, an action directive, the remaining preferences to elicit, and, in the ECA condition, facial expression and body animation labels. The action directive determined the subsequent system behaviour. For preference elicitation, the assistant text was delivered directly as the agent message. For recommendation generation, the action directive triggered the recommendation engine, which generated recipes matching the participant’s stated preferences and the experimental nutritional criteria. The returned recipes were displayed as interactive cards alongside the assistant text. For questionnaire transitions, the agent generated a closing response prompting the participant to proceed to the questionnaire, which in turn replaced the chat interface with the in-application questionnaire overlay. Participants advanced each stage with a confirmation button that triggered the next turn, and the agent message was rendered according to the assigned condition.

Chatbot Condition The Chatbot condition presented participants with a text-based chat interface. Participants typed their messages into a text input field and received text-only responses, displayed turn-by-turn by replacing the previous message in the display area. Prior research suggests that images can elicit emotional responses and increase engagement with a corresponding narrative by stimulating mental imagery (Babin and Burns 1997). The interface was therefore



Disembodied chatbot as output modality in CRS.



ECA as output modality in CRS.

Figure 3: Experimental conditions.

set against a blurred kitchen background to provide a contextually appropriate visual environment for the user and to stimulate mental imagery consistent with the task scenario.

ECA Condition The ECA condition augmented the same interface with a fully embodied agent featuring a three-dimensional avatar rendered in real time, positioned against the same blurred kitchen background. The avatar communicated through a combination of verbal and nonverbal cues. Verbal communication was delivered via synthesised speech with lip synchronisation matching the audio output. Nonverbal communication included contextually driven body animations (e.g., greeting, idle, recipe presentation), context-specific facial expressions (e.g., smiling when presenting recommendations), and subtle facial movements such as eye squints and blinking during speech. Agent responses were accompanied by turn-by-turn text displayed after audio playback completed to maintain visual consistency with the Chatbot condition. The avatar was implemented as a rigged GLTF mesh sourced from Ready Player Me, with body animations driven by Mixamo skeletal animations. Facial expressions were interpolated by mapping expression labels (e.g., smile, idle) to predefined ARKit morph target weight dictionaries. Speech was synthesised per message using the ElevenLabs text-to-speech API, and lip synchronisation was achieved through viseme-based facial animation during audio playback. Figure 3 illustrates both conditions visually.

Recommendation Design

Preference Elicitation. Prior to the main study, we conducted a pre-study with a separate sample of 20 participants recruited via Prolific to identify the most relevant preference dimensions for the recipe recommendation task. Participants ranked a set of candidate preference questions adapted from prior recipe recommender systems research (El Majjodi et al. 2023; El Majjodi, Starke, and Trattner 2022) according to perceived importance. Rankings were aggregated using the Borda count method, which assigns points inversely proportional to rank position and sums them across participants to produce an overall importance score for each preference dimension (Emerson 2013).

Preference elicitation focused on general dietary preferences rather than health-specific filters to ensure that the trust-violation manipulation remained attributable to the agent’s misleading health claim rather than participant-selected health constraints. The three highest-ranked preference dimensions were dish type (importance score = 74.38), total preparation time (importance score = 68.12), and recipe complexity operationalised through ingredient count (importance score = 53.75). In the main study, the agent elicited these preferences in a fixed order across both interactions, following their importance ranking.

Recipe Dataset. The recipe corpus was sourced from Allrecipes.com¹, which has been used in prior food recommender systems research (El Majjodi, Starke, and Trattner 2022; El Majjodi et al. 2023). The full dataset comprises 35,516 recipes across 47 attributes. Dish type categories and recipe images were obtained by scraping each recipe’s URL. For the purposes of this study, the dataset was restricted to main-dish recipes with complete nutritional information across all four FSA nutrients, at least three ingredients, and an available recipe image. The mean recipe rating in the dataset was $M = 4.04$, consistent with the generally high user ratings characteristic of Allrecipes.com (El Majjodi, Starke, and Trattner 2022).

Recipes were grouped into six dish-type categories (e.g., Pasta and Stuffed Dishes), four preparation-time categories (e.g., up to 60 minutes and up to 90 minutes), and three complexity levels (low, medium, high) based on ingredient count. FSA scores were computed for each recipe as described in Section , and recipes were classified as nutritionally healthy ($FSA \leq 6$, no red MTL indicators) or nutritionally unhealthy ($FSA \geq 10$, no green MTL indicators) to maximise the nutritional contrast. After applying these criteria, the final dataset comprised 127 nutritionally healthy and 185 nutritionally unhealthy main-dish recipes. Table 1 summarises the FSA score distribution across both categories.

Recommendation Algorithm. Recipes were ranked using a weighted composite score combining nutritional quality, preference similarity, and recipe rating:

$$\text{score} = 0.70 \cdot \text{FSA} + 0.25 \cdot \text{proximity} + 0.05 \cdot \text{rating},$$

where FSA denotes the recipe’s FSA health score, proximity denotes the inverse k -nearest neighbour distance

¹<https://github.com/shaansubbaiah/allrecipes-scraper>, Accessed: 2026-08-03

Category	FSA Score								
	4	5	6	7	8	9	10	11	12
Healthy	26	70	71	—	—	—	—	—	—
Unhealthy	—	—	—	—	—	—	60	97	28

Table 1: FSA score distribution of recipes in the restricted dataset pools used for the trust-violation manipulation. Scores 7–9 were excluded to maximise nutritional contrast between interactions.

(Bishop and Nasrabadi 2006) within the elicited preference space, and rating denotes the normalised recipe rating from the dataset. During t_1 , FSA scores were inverted to prioritise nutritionally healthy recipes. During t_2 , scores were used directly to prioritise nutritionally unhealthy recipes. The FSA component received the highest weight to maximise nutritional contrast between interactions while preserving recommendation realism through preference similarity and recipe ratings. In each interaction, the three highest-scoring recipes were presented to participants as interactive recipe cards.

Recommendation Presentation. In each interaction, the agent presented three recipe recommendations retrieved from the appropriate FSA-defined pool and matched to the participant’s stated preferences. The set was limited to three options per interaction, informed by research demonstrating that larger choice sets induce cognitive overload and decision paralysis (Iyengar and Lepper 2000), whereas fewer alternatives reduce choice difficulty (Bollen et al. 2010), while remaining within the capacity of human working memory (Cowan, Chen, and Rouder 2004). Each recommendation was displayed as an interactive recipe card showing the recipe image, title, total preparation time, ingredient count, calories, and a full MTL label per nutrient. As such, all information required for the task was visible directly on the card. Participants could additionally open any recipe via a “View Recipe” control to inspect its full ingredient list and preparation instructions in a detail, affording deeper engagement with each recommendation. Figure 4 shows examples of the recipe cards as displayed to participants in each interaction.

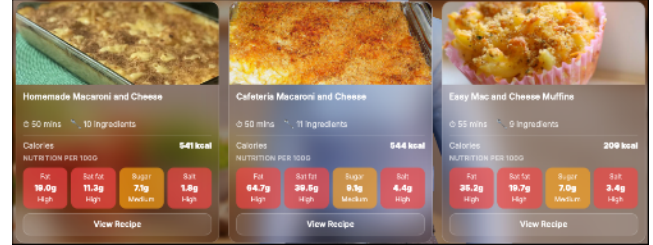
Measures

All constructs were measured using established scales adapted from the literature and administered on 7-point Likert scales, as reported in Table 2.

Following Seeger and Heinzl (2021), trust was operationalised as goodwill trust and qualification trust, with both dimensions averaged into a composite score reflecting overall trust (McKnight, Choudhury, and Kacmar 2002b). Trust was assessed after the first interaction (Trust_{t_1}) and after the second interaction (Trust_{t_2}), with trust loss computed as the difference $\text{Trust}_{t_1} - \text{Trust}_{t_2}$, yielding three loss scores: overall trust loss, goodwill trust loss, and qualification trust loss. Perceived healthiness of the recommended recipes was assessed after t_2 via a custom single-item measure capturing participants’ subjective evaluation of the recipes presented in the second interaction. Behavioural intention to follow the agent’s recommendations was measured using a scale



(t_1) Healthy recommendations (FSA ≤ 6).



(t_2) Unhealthy recommendations (FSA ≥ 10).

Figure 4: Example recipe cards as displayed to participants. Preferences example: Pasta, up to 60 minutes and medium complexity.

adapted from Bleier, Harmeling, and Palmatier (2019), and continuance intention to use the agent in the future using a scale adapted from Han and Ko (2025), assessed in the post-study questionnaire. Background measures including age, gender, cooking frequency, trust disposition (Lankton, McKnight, and Thatcher 2014), and familiarity with conversational agents (Arce-Urriza et al. 2025) were assessed in the post-study questionnaire and included as covariates in the regression analyses.

We assessed convergent and discriminant validity to evaluate the measurement instruments. Composite reliability values exceeded the recommended threshold of .70, and average variance extracted (AVE) values exceeded .50 for all constructs, supporting convergent validity (Fornell and Larcker 1981). For discriminant validity, all factor loadings exceeded .60 and loaded more strongly on their assigned construct than on any other factor (Gefen and Straub 2005); the square root of each construct’s AVE also exceeded all inter-construct correlations (Fornell and Larcker 1981).

Data Analysis

We tested our hypotheses using ordinary least squares (OLS) regression with all background measures included as covariates: age, gender, cooking frequency, trust disposition, and familiarity with conversational recommender agents. Agent embodiment (Chatbot = 0, ECA = 1) was entered as the sole predictor of interest in each model. OLS regression was used to test the direct effects hypothesised in H1a, H1b, H2, H4a, H4b, and H5. Mediation effects for H3a and H3b were tested using a bootstrapped indirect effects approach equivalent to Hayes’s 2018 PROCESS macro, implemented in Python with 5,000 resampling iterations and 95% bias-corrected confidence intervals. An indirect effect was inter-

Construct	Items	Loadings
Goodwill Trust (McKnight, Choudhury, and Kacmar 2002b) CR=.94, AVE=.71	The agent would act in my best interest.	.89
	The agent is interested in my well-being, not just its own.	.84
	If I required help, the agent would do its best to help me.	.82
	The agent is truthful in its dealings with me.	.88
	I would characterise the agent as honest.	.84
	The agent would keep its commitments.	.79
Qualification Trust (McKnight, Choudhury, and Kacmar 2002b) CR=.98, AVE=.93	The agent is competent and effective in providing recipe recommendations.	.97
	The agent performs its role of giving recommendations very well.	.96
	Overall, the agent is capable and proficient.	.96
Behavioural Intention (Bleier, Harmeling, and Palmatier 2019) CR=.93, AVE=.82	It is likely that I would follow at least one of the recommended recipes.	.96
	I am ready to follow at least one of the recommended recipes.	.96
	(R) It is not probable that I would follow any of the recommended recipes.	.79
Perceived Healthiness Single item	<i>Note: Recipes high in fat, saturated fat, sugar, or salt are considered less healthy; recipes low in these components are considered healthier.</i> I think the recipe recommendations I received during the second interaction were: (1 = <i>very unhealthy</i> , 7 = <i>very healthy</i>)	–
Continuance Intention (Han and Ko 2025) CR=.98, AVE=.95	I would continue to use the recommendation agent in the future.	.97
	I would recommend the recommendation agent to others.	.98
	I would use the recommendation agent in the long term.	.98
Trust Disposition (Lankton, McKnight, and Thatcher 2014) CR=.90, AVE=.74	Typically I trust new IT until it proves I should not.	.88
	I usually trust IT until it gives me a reason not to.	.92
	I generally give IT the benefit of the doubt when I first use it.	.79
Familiarity (Arce-Urriza et al. 2025) CR=.87, AVE=.77	I am familiar with using AI conversational agents for recommendations.	.88
	I have already had similar conversations with AI conversational agents.	.88

Table 2: Measurement instruments, factor loadings, and reliability statistics. CR = composite reliability; AVE = average variance extracted.

preted as supported when the 95% confidence interval excluded zero (Preacher and Hayes 2004). To determine the minimum detectable effect size given the achieved sample size, we conducted a sensitivity power analysis using G*Power 3.1 (Faul et al. 2007). With $N = 56$, $\alpha = .05$, power = .80, one predictor of interest (agent embodiment), and up to five covariates, the study had 80% power to detect effects of $f^2 \geq 0.146$, corresponding to a medium effect by Cohen’s conventions (Cohen 2013).

Results

Randomisation Check and Demographics

The final sample ($N = 56$) was predominantly aged 25–34 (46%), with a further 20% aged 18–24, was 59% female, and comprised participants who cook frequently (91%). Random assignment produced comparable groups across all background measures. Conditions did not differ significantly on age ($\chi^2(4) = 2.38$, $p = .794$), gender ($\chi^2(1) = 0.00$, $p = 1.000$), cooking frequency ($\chi^2(2) = 3.30$, $p = .508$), trust disposition (Chatbot: $M = 5.05$, $SD = 1.09$; ECA: $M = 5.44$, $SD = 1.12$; $t(54) = 1.34$, $p = .185$), or familiarity with conversational recommender agents (Chatbot: $M = 5.53$, $SD = 1.03$; ECA: $M = 5.81$, $SD = 1.03$; $t(54) = 1.02$, $p = .312$). A full demographic breakdown by condition is provided in the shared repository.

Stimulus Manipulation Check

Two checks were conducted to verify the integrity of the recipe stimuli. First, to confirm nutritional equivalence across conditions, mean FSA scores of the healthy and unhealthy recipes presented to participants were compared between conditions via independent samples t -tests. Healthy recipes scored well below the FSA threshold in both conditions (ECA: $M = 5.19$; Chatbot: $M = 5.29$) and unhealthy recipes scored well above it (ECA: $M = 10.96$; Chatbot: $M = 10.94$), with neither differing significantly across conditions (healthy: $t(54) = -0.65$, $p = .519$; unhealthy: $t(54) = 0.18$, $p = .860$). The mean nutritional distance between interactions was also equivalent across conditions (ECA: $M = 5.78$; Chatbot: $M = 5.66$; $t(54) = 0.64$, $p = .528$). Second, to verify that the recipe stimuli were perceptible as healthy versus unhealthy on the basis of the displayed nutritional information alone and independent of the agent’s verbal framing, we conducted a separate perceptual check with an independent sample ($N = 20$) recruited from Prolific. Participants were presented with a randomised selection of healthy and unhealthy recipe cards from the stimulus set and evaluated each recipe’s healthiness on the same 7-point scale used in the main study, without any agent interaction. Healthy recipes were rated significantly above the scale midpoint ($M = 5.10$, $SD = 0.59$, $t(19) = 8.40$,

Variable	Chatbot		ECA	
	M	SD	M	SD
Overall trust loss	0.77	1.07	0.13	1.02
Goodwill trust loss	0.66	0.85	0.07	0.84
Qualification trust loss	0.89	1.46	0.19	1.32
Perceived healthiness	3.86	1.73	4.96	1.51
Behavioural intention	4.24	1.95	5.46	1.52
Continuance intention	4.22	2.00	5.48	1.73

Table 3: Descriptive statistics by condition. M = mean, SD = standard deviation.

$p < .001$) and unhealthy recipes were rated significantly below it ($M = 2.68$, $SD = 1.08$, $t(16) = -5.07$, $p < .001$), confirming that participants reliably discriminated between healthy and unhealthy recipes on the basis of the displayed MTL labels alone. Taken together, these checks establish that the stimuli were nutritionally equivalent across conditions and perceptually valid as a manipulation of recipe healthiness.

Hypothesis Tests

Descriptive statistics for all outcome variables are reported in Table 3, and the full path model with standardised coefficients is summarised in Figure 5.

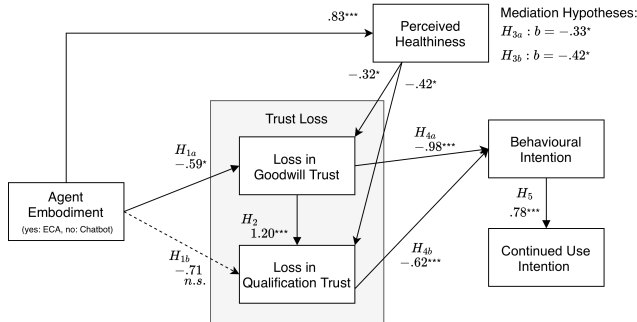


Figure 5: Results of the hypothesis tests. Solid lines indicate significant paths ($p < .05$); dashed lines indicate non-significant paths. $***p < .001$; $**p < .01$; $*p < .05$.

H1a, H1b: Trust loss following the violation. Agent embodiment had a significant negative effect on goodwill trust loss ($b = -0.59$, $p = .019$, $R^2 = .320$), supporting H1a: ECA participants lost less goodwill trust following the misleading health claim than chatbot participants did. The direct effect on qualification trust loss did not reach conventional significance ($b = -0.71$, $p = .078$, $R^2 = .303$), and H1b is not supported, though the directional pattern is consistent with H1a.

H2: Cascade from goodwill trust loss to qualification trust loss. Goodwill trust loss had a significant positive effect on qualification trust loss ($b = 1.20$, $p < .001$, $R^2 = .677$), supporting H2. Participants who lost more goodwill trust following the violation also lost substantially more qualification trust ($r = .735$), suggesting that the trust

violation cascaded from perceived benevolence and integrity into perceived competence.

H3a, H3b: Perceived healthiness as mediator of embodiment effects on trust loss. Agent embodiment had a significant positive effect on perceived recipe healthiness ($b = 0.83$, $p = .045$, $R^2 = .493$): ECA participants rated the nutritionally unhealthy recipes as healthier ($M = 4.96$) than chatbot participants did ($M = 3.86$). Perceived healthiness in turn negatively predicted trust loss along both dimensions ($b = -0.32$, $p < .001$ for goodwill trust loss; $b = -0.42$, $p = .003$ for qualification trust loss).

The indirect effect of agent embodiment on goodwill trust loss through perceived healthiness was therefore significant ($b = -0.33$, 95% CI $[-0.777, -0.029]$), supporting H3a. The indirect effect on qualification trust loss was also significant ($b = -0.42$, 95% CI $[-1.026, -0.003]$), supporting H3b. Together, these results indicate that embodiment inflated participants' healthiness perceptions of the nutritionally unhealthy recipes, and this inflated perception in turn buffered the trust loss that the misleading claim would otherwise have triggered.

H4a, H4b: Trust loss on behavioural intention. Goodwill trust loss had a significant negative effect on behavioural intention to follow the agent's recommendations ($b = -0.98$, $p < .001$, $R^2 = .556$), supporting H4a. Qualification trust loss also had a significant negative effect on behavioural intention ($b = -0.62$, $p < .001$, $R^2 = .559$), supporting H4b. Participants who lost more trust following the violation reported less willingness to follow the agent's recommendations, with both dimensions of trust loss contributing independently.

H5: Behavioural intention on continuance intention. Behavioural intention to follow the agent's recommendations had a significant positive effect on continuance intention ($b = 0.78$, $p < .001$, $R^2 = .736$), supporting H5. Participants more willing to follow the recommendations also expressed stronger intentions to continue using the agent in the future.

Discussion

Interpreting the Results

This study investigated how agent embodiment as an output modality in CRSs shapes user responses to a misleading health recommendation, and how these responses propagate into perceptions, attitudes, and behavioural intentions toward the agent. Our findings refine the picture of embodiment as a beneficial design property and substantiate the trust-buffering risks flagged conceptually by prior work (Lu- petti et al. 2023; Seeger and Heinzl 2021).

As we hypothesised in line with work on trust (Seeger and Heinzl 2021; De Visser et al. 2016), embodiment buffered trust loss across both dimensions through users' interpretation of the recommendations. Concretely, the inflated perceived healthiness of the nutritionally unhealthy recipes in the ECA condition operated as the mechanism through which embodiment shielded itself from the trust loss the misleading claim would otherwise have triggered. This adds a perceptual layer to accounts of trust buffering that have

so far emphasised attribution (Seeger and Heinzl 2021; De Visser et al. 2016), where users encounter a failure and then attribute it to unstable, situational causes more readily for human-like agents. Our data point to a different process running upstream of this, extending Go and Sundar (2019)'s observation that visual embodiment cues promote leniency toward agent errors, to a stage before recognition. Embodied delivery shaped how the recommendations were interpreted in the first place, such that the misleading claim was not registered as a failure to begin with.

The trust cascade is evident in the strong coupling of goodwill and qualification trust loss, consistent with Seeger and Heinzl (2021)'s prediction that goodwill trust, once damaged, indirectly destabilises qualification trust. A misleading health claim is at its core an integrity violation, and once users perceive the agent as having acted dishonestly, doubts about competence follow. Embodiment reduced goodwill trust loss directly but reached qualification trust only indirectly, through inflated healthiness perceptions and the goodwill cascade. This is theoretically coherent, as embodiment's social cues bear most directly on goodwill trust, the intention-related dimension, whereas competence judgements draw more on task evidence. The direct effect of embodiment on qualification trust loss was comparable in magnitude to its effect on goodwill trust loss but markedly more variable, which, together with the modest sample, limited the power to detect it. The absence of a direct effect thus reflects how embodiment operates on competence-related trust, not whether it operates at all.

Lower trust loss also predicted greater willingness to follow the recommendations and, in turn, stronger continuance intentions, so ECA participants were more likely both to follow the misleading recommendations and to keep using the agent that produced them. This extends the trust shield effect (Seeger and Heinzl 2021) from customer service to a higher-stakes health setting and links it to a downstream behavioural chain that prior work has not examined. This capacity is double-edged. When an agent is accurate, its ability to sustain trust and engagement supports the long-term relationships effective interaction requires (Bickmore and Cassell 2001). When it misleads, the same properties sustain misplaced trust and compliance against clear nutritional evidence. Our findings give empirical form to Lupetti et al. (2023)'s distinction between a user's *trust* and the agent's *trustworthiness*, as ECA participants continued to trust the agent even though its recommendations, judged by their nutritional quality, did not merit that trust.

Embodied health CRSs should therefore include transparency mechanisms that preserve users' capacity to evaluate recommendations independently, such as explicit communication uncertainty, objective nutritional information, and disclosure of the basis for health claims. The implications also reach beyond interface design to the wider ecosystem in which such systems are deployed. For platform providers, our findings caution against treating embodiment as a neutral engagement lever. For regulators and standards bodies, the decoupling of user trust from agent trustworthiness suggests a role for guidance that constrains how persuasive embodied cues should be used in consequential

domains, for example by treating undisclosed embodiment-driven persuasion as a consumer-protection concern. Public health efforts could in turn raise awareness that an agent's delivery does not signal recommendation quality.

Limitations and Future Work

Several limitations should be considered when interpreting the findings. First, given the modest sample size of $N = 56$, the findings warrant confirmatory replication in larger samples. Second, the study employed a simulated scenario in a single health-adjacent domain. While scenario-based designs afford experimental control, they may not fully capture how users actively seek recommendations in richer settings. Additionally, the manipulation was administered as a single event. Real-world trust dynamics are likely more complex, since violations may accumulate across interactions, and users may update their assessments in light of subsequent agent behaviour. The present study leaves open whether the observed effects persist or erode over time. Finally, while we interpret our findings consistent with *automation bias* (Lupetti et al. 2023), this study does not formally adjudicate between theoretical accounts of why embodiment produced the perceptual effects found. Candidate alternatives include the Computers Are Social Actors paradigm and the Ethopoeia framework (Nass, Steuer, and Tauber 1994; Von der Pütten et al. 2010), under which social cues trigger automatic social responding regardless of users' beliefs about the system, as well as anthropomorphism (Epley, Waytz, and Cacioppo 2007), whereby attributing human-like intentions to the agent shapes the interpretation of its outputs. Distinguishing these accounts requires designs that incorporate validated measures of each as parallel mediators. These limitations mark directions for future work on when embodied CRSs help rather than mislead. A priority is developing and testing interventions that counter the effect, such as disclosure prompts, comparisons of transparency designs, and calibration cues that re-anchor trust to the agent's actual reliability, and establishing which of these restore independent judgement well enough to inform consumer-protection and public health policy.

Conclusion

This study investigated how agent embodiment as an output modality in a conversational recommender system shapes user responses to a misleading health recommendation. In a between-subjects experiment, a fully embodied agent and a text-based chatbot each delivered nutritionally unhealthy recipes framed by a misleading health claim. Our results show that relative to chatbot participants, ECA participants perceived the nutritionally unhealthy recipes as healthier, lost less goodwill trust following the misleading claim, expressed greater willingness to follow the recommendations, and reported stronger intentions to continue using the agent. As embodied CRSs are increasingly deployed in health contexts, understanding when embodiment becomes a liability rather than an asset is an urgent design priority, one that calls for transparency mechanisms, replication, and mechanistic investigation.

Ethics Statement

This Section outlines the ethical considerations and safeguards we adopted. The study required a controlled trust-violation manipulation. That is, in the second interaction, the agent described nutritionally unhealthy recipes as healthy in order to operationalise a misleading health recommendation. The manipulation was mild and bounded, as participants engaged with a hypothetical recipe-selection scenario, were not asked to prepare or consume any food, and received no personalised medical advice. The objective nutritional profile of every recipe was displayed accurately on-screen throughout via the FSA Multiple Traffic Light labels, so no participant was presented with false nutritional data.

Participants were recruited through Prolific and compensated at approximately £9 per hour, above the platform's minimum, regardless of their responses or whether their data were ultimately included in analysis. Before beginning, participants were presented with an information screen describing the scenario, the procedure, the estimated duration, the confidentiality of their responses, their right to withdraw, and the use of attention checks, and they proceeded into the study voluntarily. No personally identifying information and no sensitive health data were collected. All responses were anonymous and used solely for research purposes. Following data collection, all participants were debriefed through disclosing the deception, explaining the study's purpose, and clarifying that the recipes described as healthy in the second interaction were not nutritionally healthy and should not be treated as such.

The study was conducted in accordance with established ethical principles for minimal-risk online human-subjects research, including informed participation, anonymity, voluntary withdrawal, fair compensation, and post-study debriefing. Retrospective review of the study has additionally been requested from our institution's research ethics board.

References

- André, E.; and Pelachaud, C. 2010. Interacting with embodied conversational agents. In *Speech technology: Theory and applications*, 123–149. Springer.
- Arce-Urriza, M.; Chocarro, R.; Cortiñas, M.; and Marcos-Matás, G. 2025. From familiarity to acceptance: The impact of Generative Artificial Intelligence on consumer adoption of retail chatbots. *Journal of Retailing and Consumer Services*, 84: 104234.
- Babin, L. A.; and Burns, A. C. 1997. Effects of print ad pictures and copy containing instructions to imagine on mental imagery that mediates attitudes. *Journal of Advertising*, 26(3): 33–44.
- Barki, H.; Robert, J.; and Dulipovici, A. 2015. Reconceptualizing trust: A non-linear Boolean model. *Information & Management*, 52(4): 483–495.
- Ben Saad, S.; and Choura, F. 2023. Towards better interaction between salespeople and consumers: the role of virtual recommendation agent. *European Journal of Marketing*, 57(3): 858–903.
- Bickmore, T.; and Cassell, J. 2001. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 396–403.
- Bies, R. J.; Tripp, T. M.; Kramer, R.; and Tyler, T. 1996. Beyond distrust. *Trust in Organizations*, 246–260.
- Bishop, C. M.; and Nasrabadi, N. M. 2006. *Pattern recognition and machine learning*, volume 4. Springer.
- Bleier, A.; Harmeling, C. M.; and Palmatier, R. W. 2019. Creating effective online customer experiences. *Journal of Marketing*, 83(2): 98–119.
- Bollen, D.; Knijnenburg, B. P.; Willemsen, M. C.; and Graus, M. 2010. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*, 63–70.
- Burgoon, J. K.; and Le Poire, B. A. 1993. Effects of communication expectancies, actual communication, and expectancy disconfirmation on evaluations of communicators and their communication behavior. *Human Communication Research*, 20(1): 67–96.
- Cai, W.; Jin, Y.; and Chen, L. 2022. Impacts of personal characteristics on user trust in conversational recommender systems. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, 1–14.
- Cassell, J. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI Magazine*, 22(4): 67–67.
- Cassell, J.; Bickmore, T.; Vilhjálmsdóttir, H.; and Yan, H. 2000. More than just a pretty face: affordances of embodiment. In *Proceedings of the 5th international conference on Intelligent user interfaces*, 52–59.
- Christakopoulou, K.; Radlinski, F.; and Hofmann, K. 2016. Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 815–824.
- Cohen, J. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.
- Cowan, N.; Chen, Z.; and Rouder, J. N. 2004. Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, 15(9): 634–640.
- De Visser, E. J.; Monfort, S. S.; McKendrick, R.; Smith, M. A.; McKnight, P. E.; Krueger, F.; and Parasuraman, R. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3): 331.
- DeepBrain AI. 2024. Retail Solutions: AI Human Kiosks for K11 and 7-Eleven. <https://www.aistudios.com/solutions/retail>. Accessed: 2026-08-03.
- Dehn, D. M.; and Van Mulken, S. 2000. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, 52(1): 1–22.
- Department of Health and Social Care. 2016. Front of Pack Nutrition Labelling Guidance. <https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance>. Accessed: 2026-08-03.

- El Majjodi, A.; Starke, A. D.; Elahi, M.; Trattner, C.; et al. 2023. The Interplay between Food Knowledge, Nudges, and Preference Elicitation Methods Determines the Evaluation of a Recipe Recommender System. In *IntRS@ RecSys*, 1–18.
- El Majjodi, A.; Starke, A. D.; and Trattner, C. 2022. Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system. In *Proceedings of the 30th ACM conference on user modeling, adaptation and personalization*, 48–56.
- Emerson, P. 2013. The original Borda count and partial voting. *Social Choice and Welfare*, 40(2): 353–358.
- Epley, N.; Waytz, A.; and Cacioppo, J. T. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, 114(4): 864.
- Faul, F.; Erdfelder, E.; Lang, A.-G.; and Buchner, A. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2): 175–191.
- Firework. 2024. Firework Launches AVA: The Virtual Shopping Assistant Giving a Face to E-Commerce. <https://firework.com/blog/firework-launches-ava-the-virtual-shopping-assistant-giving-a-face-to-e-commerce>. Accessed: 2026-08-03.
- Følstad, A.; Araujo, T.; Law, E. L.-C.; Brandtzaeg, P. B.; Papadopoulos, S.; Reis, L.; Baez, M.; Laban, G.; McAllister, P.; Ischen, C.; et al. 2021. Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 103(12): 2915–2942.
- Følstad, A.; Nordheim, C. B.; and Bjørkli, C. A. 2018. What makes users trust a chatbot for customer service? An exploratory interview study. In *International conference on internet science*, 194–208. Springer.
- Fornell, C.; and Larcker, D. F. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1): 39–50.
- Freire, W. B.; Waters, W. F.; Rivas-Mariño, G.; Nguyen, T.; and Rivas, P. 2017. A qualitative study of consumer perceptions and use of traffic light food labelling in Ecuador. *Public Health Nutrition*, 20(5): 805–813.
- Fung, P.; Bachrach, Y.; Celikyilmaz, A.; Chaudhuri, K.; Chen, D.; Chung, W.; Dupoux, E.; Gong, H.; Jégou, H.; Lazaric, A.; et al. 2025. Embodied ai agents: Modeling the world. *arXiv preprint arXiv:2506.22355*.
- Gao, C.; Lei, W.; He, X.; De Rijke, M.; and Chua, T.-S. 2021. Advances and challenges in conversational recommender systems: A survey. *AI Open*, 2: 100–126.
- Gefen, D.; and Straub, D. 2005. A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information systems*, 16(1): 5.
- Go, E.; and Sundar, S. S. 2019. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97: 304–316.
- Guo, S.; Zhang, S.; Sun, W.; Ren, P.; Chen, Z.; and Ren, Z. 2023. Towards explainable conversational recommender systems. In *Proceedings of the 46th International ACM SIGIR conference on research and development in information retrieval*, 2786–2795.
- Hall, J. A.; Horgan, T. G.; and Murphy, N. A. 2019. Non-verbal communication. *Annual Review of Psychology*, 70(2019): 271–294.
- Han, J.; and Ko, D. 2025. Trust Formation, Error Impact, and Repair in Human–AI Financial Advisory: A Dynamic Behavioral Analysis. *Behavioral Sciences*, 15(10): 1370.
- Hayes, A. F. 2018. Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, 85(1): 4–40.
- Hong, Z.-W.; Huang, R.-T.; Chin, K.-Y.; Yen, C.-C.; and Lin, J.-M. 2010. An interactive agent system for supporting knowledge-based recommendation: A case study on an e-novel recommender system. In *Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication*, 1–8.
- Iacovazzi, A. R.; Blanco, L.; Spillo, G.; and Musto, C. 2025. E-Mealio: An LLM-Powered Conversational Agent for Sustainable and Healthy Food Recommendation. In *International Workshop on Recommender Systems for Sustainability and Social Good*, 126–140. Springer.
- Ikemoto, Y.; Asawavetvutt, V.; Kuwabara, K.; and Huang, H.-H. 2019. Tuning a conversation strategy for interactive recommendations in a chatbot setting. *Journal of Information and Telecommunication*, 3(2): 180–195.
- Iyengar, S. S.; and Lepper, M. R. 2000. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6): 995.
- Jannach, D.; Manzoor, A.; Cai, W.; and Chen, L. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5): 1–36.
- Jannach, D.; Zanker, M.; Felfernig, A.; and Friedrich, G. 2010. *Recommender systems: an introduction*. Cambridge university press.
- Kiuchi, K.; Otsu, K.; and Hayashi, Y. 2024. Psychological insights into the research and practice of embodied conversational agents, chatbots and social assistive robots: a systematic meta-review. *Behaviour & Information Technology*, 43(15): 3696–3736.
- Krajcovic, M.; Demcak, P.; and Kuric, E. 2025. Talking Surveys: How Photorealistic Embodied Conversational Agents Shape Response Quality, Engagement, and Satisfaction. *arXiv preprint arXiv:2508.02376*.
- Laban, G.; Hough, J.; Lee, M.; Markelius, A.; Foster, M. E.; Stuart-Smith, J.; and Ahmad, M. I. 2025. Bias and Fairness in Conversational User Interfaces. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces*, 1–8.
- Lankton, N.; McKnight, D. H.; and Thatcher, J. B. 2014. Incorporating trust-in-technology into expectation disconfirmation theory. *The Journal of Strategic Information Systems*, 23(2): 128–145.

- Liu, Y.; Zhang, W.; Chen, Y.; Zhang, Y.; Bai, H.; Feng, F.; Cui, H.; Li, Y.; and Che, W. 2023. Conversational recommender system and large language model are made for each other in e-commerce pre-sales dialogue. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 9587–9605.
- Lupetti, M. L.; Hagens, E.; Van Der Maden, W.; Steegers-Theunissen, R.; and Rousian, M. 2023. Trustworthy embodied conversational agents for healthcare: A design exploration of embodied conversational agents for the periconception period at erasmus mc. In *Proceedings of the 5th International Conference on Conversational User Interfaces*, 1–14.
- Mayer, R. C.; Davis, J. H.; and Schoorman, F. D. 1995. An integrative model of organizational trust. *Academy of Management Review*, 20(3): 709–734.
- McKnight, D. H.; Choudhury, V.; and Kacmar, C. 2002a. Developing and Validating Trust Measures for E-Commerce: An Integrative Typology. *Information Systems Research*, 13(3): 334–359.
- McKnight, D. H.; Choudhury, V.; and Kacmar, C. 2002b. Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3): 334–359.
- Nass, C.; Steuer, J.; and Tauber, E. R. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 72–78.
- Preacher, K. J.; and Hayes, A. F. 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4): 717–731.
- Qiu, L.; and Benbasat, I. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems*, 25(4): 145–182.
- Qiu, L.; and Benbasat, I. 2010. A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer Studies*, 68(10): 669–688.
- Rheu, M.; Shin, J. Y.; Peng, W.; and Huh-Yoo, J. 2021. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction*, 37(1): 81–96.
- Robb, D. A.; Lopes, J.; Ahmad, M. I.; McKenna, P. E.; Liu, X.; Lohan, K.; and Hastie, H. 2023. Seeing eye to eye: trustworthy embodiment for task-based conversational agents. *Frontiers in Robotics and AI*, 10: 1234767.
- Rousseau, D. M.; Sitkin, S. B.; Burt, R. S.; and Camerer, C. 1998. Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3): 393–404.
- Schlesener, E. A.; Ziolkowski, M.; Wong, S.-K.; Westmoreland, B.; and Babu, S. V. 2026. “Am I Understood?”: How the Interplay between Embodiment and Theory of Mind Behavior Affects LLM-Based Conversational Agents on Perceived Trust, Anthropomorphism, Presence, Usability, and User Experience. *ACM Transactions on Interactive Intelligent Systems*, 16(2): 1–45.
- Seeger, A.-M.; and Heinzl, A. 2021. Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service? In *ECIS*.
- Sonnenberg, L.; Gelsomin, E.; Levy, D. E.; Riis, J.; Baraclough, S.; and Thorndike, A. N. 2013. A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase. *Preventive Medicine*, 57(4): 253–257.
- Talati, Z.; Egnell, M.; Herberg, S.; Julia, C.; and Pettigrew, S. 2019. Consumers’ perceptions of five front-of-package nutrition labels: An experimental study across 12 countries. *Nutrients*, 11(8): 1934.
- Tsamos, I.; and Marakakis, E. 2025. DietQA: A Comprehensive Framework for Personalized Multi-Diet Recipe Retrieval Using Knowledge Graphs, Retrieval-Augmented Generation, and Large Language Models. *Computers*, 14(10): 412.
- Von der Pütten, A. M.; Krämer, N. C.; Gratch, J.; and Kang, S.-H. 2010. “It doesn’t matter what you are!” Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6): 1641–1650.
- Weitz, K.; Schiller, D.; Schlagowski, R.; Huber, T.; and André, E. 2019. “Do you trust me?” Increasing user-trust by integrating virtual agents in explainable AI interaction design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 7–9.
- Yang, L.; Sobolev, M.; Tsangouri, C.; and Estrin, D. 2018. Understanding user interactions with podcast recommendations delivered via voice. In *Proceedings of the 12th ACM Conference on Recommender Systems*, 190–194.