



ARTICLE

AI talking science: Experimental studies on the perception of AI-based chatbots as sources of science-based information

Esther Greussing , Friederike Hendriks , Aike C. Horstmann , Yannic Meier , Bianca Nowak  and Rainer Bromme 

Abstract

AI-based chatbots offer new opportunities for communicating science-based information, but often fall short of established standards. We conducted two pre-registered experiments examining user perceptions of an AI-based chatbot providing information on nanoparticles in sunscreen. Study one ($N = 508$) tested whether a disclaimer about the chatbot's uncertain training data affected perceived source trustworthiness and information credibility. The results showed no significant effect of the disclaimer; perceptions were primarily influenced by users' prior attitudes. Study two ($N = 1059$) tested the evaluation of information on nanoparticles in sunscreen in an experiment with a 2 (source: scientist vs. AI-based chatbot) \times 2 (presentation: static vs. dynamic) between-subjects design. The study showed that the scientist was evaluated as more trustworthy and the provided information seen as more credible compared to the AI-based chatbot. The two studies highlight the relevance of perceived objectivity in science and health communication, whether executed by humans or machines.

Keywords

AI tools in science communication; Digital science communication; Public engagement with science and technology

Received: 24th June 2025

Accepted: 28th February 2026

Published: 25th May 2026

AI-based chatbots, such as OpenAI's ChatGPT, increasingly serve as information intermediaries across various domains, including science and health [e.g., Greussing et al., 2025; Henke, 2025; Weingott & Parkinson, 2025]. Typically powered by Large Language Models (LLMs), they rely on advanced deep-learning techniques to generate text by predicting word sequences based on user input. While their accessibility and linguistic fluency have facilitated adoption, these same characteristics may conceal central constraints in how such systems represent knowledge [Bulian et al., 2024]. Despite ongoing technical advances, LLM-based systems have been shown to generate responses whose factual accuracy cannot be consistently guaranteed, and the quality and provenance of their training data remain largely opaque. At the time this study was conducted, LLM-based systems were widely documented as being susceptible to generating factually incorrect or misleading information [van Dis et al., 2023]. Nevertheless, users frequently perceive AI-generated content as plausible [Spitale et al., 2023].

This tendency poses particular challenges in domains where individuals lack the domain-specific expertise to critically evaluate the information they receive in order to make informed decisions [Kessler et al., 2025]. One such domain is health advice, where AI-generated content often includes behavioural recommendations. This raises concerns, not least in light of recent findings suggesting that people are likely to follow advice from AI systems just as readily as they follow advice from humans [Leib et al., 2024].

To explore these dynamics further, we conducted two experimental studies on the perceptions of an AI-based chatbot informing about nanoparticles in sunscreen. In the first study, we investigated whether an information disclaimer — similar to those used by OpenAI for ChatGPT — influences users' evaluation of the source trustworthiness of an AI-based chatbot and the credibility of the information it provides. Additionally, we considered the influence of cognitive heuristics, particularly the machine heuristic. This heuristic is anchored in the belief that machines deliver more objective information than humans, as they are free of human biases [Sundar & Kim, 2019]. Besides characteristics of the source itself, we investigate the role of users' prior experiences and attitudes as potentially relevant factors in how users evaluate AI-generated health information. In the second study, we expand this focus by comparing an AI-based chatbot with a human scientist as sources of health information. To further investigate the role of perceived objectivity, we compare two formats of information delivery: a rather machine-like static presentation vs. a humanlike dynamic chat-style.

1 • Health information seeking via AI-based chatbots

Public health communication is increasingly challenged by complex information environments and the proliferation of mis- and disinformation [Niederdeppe et al., 2025]. In this context, AI-based chatbots may represent promising tools, as they enable personalization and interactive user engagement [Weingott & Parkinson, 2025], and their responses to health questions can be perceived as high in quality and empathy [Ayers et al., 2023] — characteristics that are particularly relevant for health information seeking contexts. Through personalization, AI-based chatbots hold particular promise for fostering health behaviour change, a central aim of public health communication [Brown et al., 2023]. Yet, the use of AI-based tools in health contexts raises questions about epistemic authority [Popowicz, 2024]: whereas the epistemic authority of medical practitioners is based on both

“what is” and “why is” questions, medical AI systems primarily focus on “what is”, lacking the epistemic goal of understanding why a knowledge claim could be conceived as being a true belief [Kerasidou & Kerasidou, 2026]. This complicates how AI-provided health information is evaluated, even as medical practice itself sometimes relies on effective treatments without a complete causal understanding of the underlying mechanisms. Against this backdrop, perceptions of information credibility and source trustworthiness emerge as critical evaluative dimensions [Niederdeppe et al., 2025].

2 - Evaluating information credibility and source trustworthiness

Information credibility is defined as “an individual’s judgment of the veracity of the content of communication” [Appelman & Sundar, 2016], and thus pertains to individuals’ evaluation of message content. When evaluating the credibility of journalistic information, for example, users look for cues such as factual accuracy, neutrality, and completeness [Matthes & Kohring, 2003]. In contrast, source trustworthiness reflects an individual’s willingness to rely on the information source, even when there is a possibility the information may be incorrect [Hendriks et al., 2015].

Source trustworthiness and information credibility are distinct yet intertwined constructs. Specifically, source trustworthiness in epistemic contexts comprises multiple dimensions — expertise (ability), benevolence (goodwill), and integrity [Hendriks et al., 2015; Mayer et al., 1995], with recent research extending this framework to include additional context-specific considerations [Goh & Ho, 2024]. This distinction is especially relevant for health information, where users’ limited domain knowledge makes evaluating the source more feasible than verifying the content itself [Bromme & Goldman, 2014]. Accordingly, individuals tend to spontaneously infer a source’s expertise, integrity, and benevolence from cues like academic credentials or topic expertise [Hendriks et al., 2015].

Although originally developed to describe trust in human experts, the concept of epistemic trustworthiness can also be applied to non-human sources. LLMs are regarded as epistemic technologies [Alvarado, 2023] and are seen as agentic communicators [Guzman & Lewis, 2020]. Particularly, their human-like communication features (e.g., interactivity, personal address, synchronicity, dynamic chat style), will likely make them subject to human evaluation criteria [Nass & Moon, 2000]. Early research has shown that LLMs are indeed subject to scrutiny regarding their expertise, integrity, and benevolence when presenting science-related information [Jonas et al., 2025]. In particular, the (perceived) objectivity of LLMs may be a crucial source characteristic for users’ trustworthiness assessments.

3 - The role of perceived objectivity in science-based communication

Science-based communication about health topics is grounded in normative assumptions, among which objectivity occupies a central position [Daston & Galison, 2010]. Regarding the perceived objectivity of AI-based chatbots, the concept of the machine heuristic [Sundar & Kim, 2019] describes the tendency to perceive AI systems as operating according to fixed, rule-based procedures, and as lacking emotions and personal interests — factors assumed to amplify biased information processing [Pham, 2007; but see also Nussbaum, 2001].

Specifically, their outputs are often viewed as less biased and less prone to error than those of humans in a similar role, leading users to attribute heightened objectivity to them.

At the same time, concerns about the objectivity of human scientists represent a longstanding issue in the philosophy and sociology of science. It is constituted that absolute objectivity in science proves infeasible given that scientific inquiry is inevitably shaped by theoretical commitments, methodological choices, and value-laden judgments. Scientists are thus understood as imperfect epistemic agents whose work may be influenced not only by disciplinary norms but also by “illusions, subjectivity, idiosyncrasies, and collective biases” [Koskinen, 2020, p. 1195]. These limitations do not undermine the legitimacy of science but can complicate public expectations and foster scepticism about scientists’ objectivity. From an audience perspective, scientific claims may be perceived as influenced by conflicts of interest, ideological positions, or institutional and funding pressures, thereby calling scientists’ integrity into question [Kienhues et al., 2020; Safford & Whitmore, 2024].

Taken together, these two strands of literature suggest a potentially important role for AI-based communicators in science-based domains. Against a backdrop of enduring concerns about human objectivity, AI systems may be perceived as a solution precisely because they are viewed as detached from human interests and motivations. Importantly, this does not imply that AI systems are inherently objective in an epistemic sense; rather, it highlights the role of *perceived objectivity* in shaping audience responses.

4 - Study one: Information disclaimers to enhance critical information evaluation

LLM-based chatbots can generate health information with remarkable fluency, but the quality of this information varies significantly. While LLMs can provide accurate and useful explanations, they are also prone to producing misinformation or hallucinations containing incorrect or fabricated content [van Dis et al., 2023]. As users increasingly depend on AI systems for reliable outcomes, scholars emphasize the importance of transparency in these systems, advocating for clear explanations of their data collection, processing, and utilization [Shin, 2021]. In the absence of such transparency, these systems should at least include disclaimers, like the one currently used by OpenAI for ChatGPT: “ChatGPT can make mistakes. Consider checking important information.” (status as of January 2026).

Disclaimers have previously been tested to protect (“inoculate”) against misinformation: highlighting specific issues or attributes in a preceding message can influence how individuals assess subsequent messages [Lewandowsky & van der Linden, 2021; for a systematic review see Ziemer & Rothmund, 2024]. For instance, disclaimers warning people about the pre-print status of scientific publications led users to rate information credibility more critically [Wingen et al., 2022]. For LLM-based chatbots, such disclaimers could address the fundamental epistemic challenge posed by opaque training data, such as uncertainty regarding data sources, selection criteria, or quality control [see van Dis et al., 2023]. Initial evidence suggests that disclaimers warning about the limited information quality of AI-produced messages may be effective: Spence et al. [2019] primed suspicion in recipients about the accuracy of information provided by a robot delivering news, which diminished perceived credibility and discouraged recipients from taking action based on the message. More recent studies, however, did not find a univocal effect of disclaimers on user

evaluations of texts produced by GenAI [Lermann Henestroza & Kimmerle, 2025]. In all cases, to elicit these effects, disclaimers must be attended and recalled, which is not necessarily the case [Oeldorf-Hirsch et al., 2020].

Contributing to this debate, we investigate whether making the uncertainty of an AI-based chatbot's training data explicit through a disclaimer influences both the evaluation of the chatbot as a source and the health information it provides. We propose the following hypotheses:¹

H1.1: *Individuals who received information regarding the uncertainty of an AI-based chatbot's data sources ascribe less trustworthiness to the chatbot than individuals who did not receive this information.*

H1.2: *Individuals who received information regarding the uncertainty of an AI-based chatbot's data sources ascribe less credibility to the content presented by the chatbot than individuals who did not receive this information.*

By introducing a disclaimer that emphasizes the questionability of the quality of content produced by the AI system, we potentially challenge prevailing assumptions about machine objectivity. Although machines are typically perceived as operating in a neutral and rule-based manner [Sundar & Kim, 2019], cautionary information about data quality may prompt users to reconsider its objectivity [Spence et al., 2019]. That is, if users rely on the machine heuristic when evaluating AI-generated information, such disclaimers should attenuate perceived objectivity by highlighting potential biases or errors. While we acknowledge that objectivity can be conceptualized as a component of integrity within established trustworthiness frameworks [Goh & Ho, 2024; Mayer et al., 1995], in this study, we examine it as a distinct factor because the machine heuristic suggests fundamental differences in how objectivity is attributed to human versus machine sources. Previous research has further shown that although perceived trustworthiness and objectivity are correlated, they are not identical constructs and are understood as distinct concepts by lay audiences in a science communication context [Cologna et al., 2021]. Accordingly, this study investigates whether users evaluate the objectivity of the chatbot as an information source more critically when they are made aware of potential flaws. We therefore hypothesize:

H1.3: *Individuals who received information regarding the uncertainty of an AI-based chatbot's data sources ascribe less objectivity to the chatbot than individuals who did not receive such information.*

Given mixed findings on the downstream effects of the machine heuristic [Liu & Wei, 2019] and its unclear theoretical role as either a mediator or a moderator [Lee, 2024], we further propose an exploratory mediation hypothesis. Specifically, perceived objectivity is conceptualized as an antecedent judgment that informs evaluations of both source trustworthiness and the credibility of the content it produces. We therefore hypothesize:

We assume that a disclaimer informing about the uncertainty of an AI-based chatbot's data sources leads to the AI-based chatbot being perceived as less objective compared to an AI-based chatbot without a disclaimer and that the chatbot's diminished perceived objectivity, in turn, decreases the chatbot's trustworthiness and content credibility.

1. Note that we have adjusted the numbering of the hypotheses compared to the pre-registration. Also, in this paper, we do not present all dependent variables pre-registered. Please see the OSF for these additional analyses.

5 - Role of prior experiences and beliefs

Perceptions of the content and source of science-based communication are formed through inferences based on prior experiences and beliefs [e.g., Wischniewski & Krämer, 2022], potentially rendering additional information, including disclaimers, ineffective. Drawing on confirmation bias, previous research has shown that communicators whose attitudes align with those of their audience are perceived as more trustworthy [Fiske & Dupree, 2014].

In addition, as AI-based chatbots are a rather novel technology, evaluations might depend on initial attitudes about and first experiences with this technology: early evidence suggests that individuals recognize the limited epistemic quality of AI-based chatbots [Wissenschaft im Dialog, 2023]. Consequently, individuals who are well-informed about AI-based chatbots, regular users of such technology, or those who hold sceptical attitudes towards AI-based chatbots may be more likely to devalue them as communicators of health information. Thus, we investigate the influence of topic-related attitudes and knowledge as well as attitudes about and experience with AI-based chatbots on participants' ratings of objectivity, information credibility, and source trustworthiness:

H1.4: We expect that a) more sceptical attitudes towards AI-based chatbots, b) more extensive subjective prior knowledge about AI-based chatbots, c) higher frequency and intensity of use of AI-based chatbots, and d) more sceptical attitudes towards nanoparticles in sunscreen will lead to more critical judgements of an AI-based chatbot; that is, less assigned objectivity, source trustworthiness, and information credibility.

6 - Study two: Comparing an AI chatbot to a human scientist in their role as science communicators

Expanding on the findings of our first study, our second study investigates the perception of an AI-based chatbot as an information source, focusing on the credibility of the information presented as well as on the trustworthiness and perceived objectivity of the information source. To this end, we compare two scenarios: one where an AI-based chatbot presents information about a health topic and another where this is done by a scientist.

In journalism, some studies suggest that machine-generated content is perceived as less credible than human-generated content [Longoni et al., 2022; Tandoc Jr. et al., 2020]. However, there is also evidence supporting the opposite trend [Graefe et al., 2018] or even indicating no discernible difference in evaluations [Graefe & Bohlken, 2020; Lermann Henestrosa et al., 2023]. A meta-analysis reveals that the effect of AI authorship on credibility ratings is relatively small and negative [Wang & Huang, 2024]. A recent study on vaccination messaging found that arguments labelled as AI-generated were consistently rated as lower in quality compared to those labelled as human-generated (i.e., by medical experts) or left unlabelled [Beckmann et al., 2025].

Building on this line of work, we contrast an AI-based chatbot with a human scientist. Drawing on research on cognitive heuristics in human-machine communication [Sundar & Kim, 2019], we assume that human scientists may be more readily suspected of operating under personal biases or possessing limited knowledge [Koskinen, 2020]. In contrast, an

AI-based chatbot may be perceived as a more impartial entity, as it is commonly associated with access to large volumes of data from diverse sources and with rule-based information processing. We therefore propose the following hypothesis:

H2.1: *Individuals who receive information from an AI-based chatbot ascribe more a) trustworthiness to the information source and b) credibility to the information presented compared to individuals who receive information from a scientist.*

Furthermore, we formally test the underlying mechanism potentially shaping this relationship, namely the heuristic evaluation of objectivity as a core characteristic of trustworthy communicators [Flanagin et al., 2020; Kruglanski et al., 2005]. While the objectivity of machines is presumably engineered into their design [Sundar & Kim, 2019], the objectivity of scientists is a more nuanced issue [Kienhues et al., 2020]: scientists may be perceived as less objective since their testimony can be tainted by their personal opinion or limits in knowledge [Wissenschaft im Dialog, 2023]. We thus assume:

H2.2: *The perceived objectivity mediates the relationship between the information source and the ascribed source trustworthiness and information credibility. That is, an AI-based chatbot is perceived as more objective than a human scientist and perceived objectivity increases perceived trustworthiness and credibility.*

7 - Impact of the way the information is presented

AI-based chatbots provide information with high immediacy and engage in dynamic interactions with users [Cress & Kimmerle, 2023]. The dialogue structure further allows users to ask questions for instance in case of a misunderstanding or to receive additional information. Extensive literature has demonstrated that machines with social cues, such as interactivity and natural language, can effectively prompt users to respond socially, despite their non-human nature [Nass & Moon, 2000]. One such cue may be a conversation that unfolds in real-time, mimicking the natural flow of dialogue between humans [Go & Sundar, 2019]. In contrast to static presentation, where all information is delivered simultaneously in a single, complete block (e.g., a full text dump or article), dynamic unfolding fosters perceptions of human-likeness. For instance, previous research has demonstrated positive effects on perceptions of social presence, competence, warmth, trustworthiness, empathy, and emotional expressiveness [Bhat, 2025; Zierau et al., 2021]. Building on this foundation, an unfolding dialogue prompts users to apply human-like social scripts [Nass & Moon, 2000], potentially interfering with the perception of machines as inanimate objects that operate in an unbiased and accurate way [i.e., machine heuristic, Sundar & Kim, 2019]. As a consequence, this may undermine users' perceptions of the chatbot as an objective source.

In sum, we assume that the presentation style — in terms of a dynamic unfolding dialogue compared to a static presentation — negatively affects how strongly people adhere to the perception that the source is objective, leading to a moderated mediation hypothesis:

H2.3: *A dynamic presentation of information (vs. a static presentation) decreases the perceived objectivity of either information source.*

8 - Study context: Nanotechnology

We selected nanoparticles in sunscreen as the topic for the study materials, as it constitutes a form of science-based information with direct relevance to health-related decisions. While nanotechnology has been researched as a risk communication topic for a while [Bostrom & Löfstedt, 2010], issue complexity, remaining uncertainty about potential long-term effects, and personal relevance require health-related decisions to rely heavily on trust in expert sources. Moreover, German audiences hold mostly positive attitudes towards nanotechnology, including its use in cosmetic products [Scheufele et al., 2009], making it a rather uncontroversial issue in the present study context.

9 - Study one

9.1 - Methods study one

Study one was a preregistered (<https://osf.io/saj6k>) between-subjects online experiment with two conditions, approved by the University of Duisburg-Essen's ethics committee (ID: 2308SPHA5351). The study was conducted in Germany. All study material (including detailed information on the measurements) can be accessed online: <https://osf.io/863v4>.

Experimental manipulations and procedure. After providing consent and demographic information, participants were asked to imagine a friend of theirs seeking advice about buying sunscreen with nanoparticles. They were told about *ChatAI*, an AI-based chatbot encountered online that is capable of answering questions using AI. Participants were randomly assigned to the control or the experimental condition. The latter included a disclaimer highlighting the chatbot's unknown data sources and potential quality issues. All participants then saw a chat history between *ChatAI* and a user on nanoparticles, their use in sunscreens, and possible related health risks (see Figure 1; see OSF for the chat protocol). The information presented was factually correct. Measures followed on *ChatAI*'s perceived objectivity, trustworthiness of the information source, and information credibility, as well as on participants' prior attitudes, knowledge, and information-handling strategies (not included in current analyses). Finally, they completed manipulation checks, were debriefed, and gave consent for data use.

Sample and sample size rationale. An a priori power analysis (SESOI: Cohen's $f = .15$, 90% power, $\alpha = .05$) determined that with $N = 500$ (250/group), there is a 92% chance of detecting the SESOI. Participants were recruited in October 2023 via GapFish, an ISO 20252-certified panel provider, using age- and gender-based quotas as sample selection criteria to approximate the German adult population. After excluding inattentive respondents, 508 participants remained (255 female, 251 male, 2 non-binary; $M_{\text{age}} = 49.83$, $SD_{\text{age}} = 17.17$). The majority of participants reported having a middle school degree (33.7%) and holding a university degree (33.5%) or university entrance level (22%).

Measurements. We report Cronbach's α (ideal $> .70$), McDonald's ω (ideal $> .70$), and the average variance extracted (AVE; ideal $> .50$). We used CFA and EFA (when necessary) to check the factor structure of all scales (see OSF). *ChatAI*'s perceived objectivity (7 items, e.g., "objective-subjective"; $\alpha = .91$, $\omega = .92$, AVE = .69; Molina and Sundar [2022], three original items to capture context-specific nuances) and perceived trustworthiness (14 items, e.g. "intelligent-unintelligent" for subscale expertise; $\alpha = .96$, $\omega = .96$, AVE = .65; Hendriks et al.

Science Chat

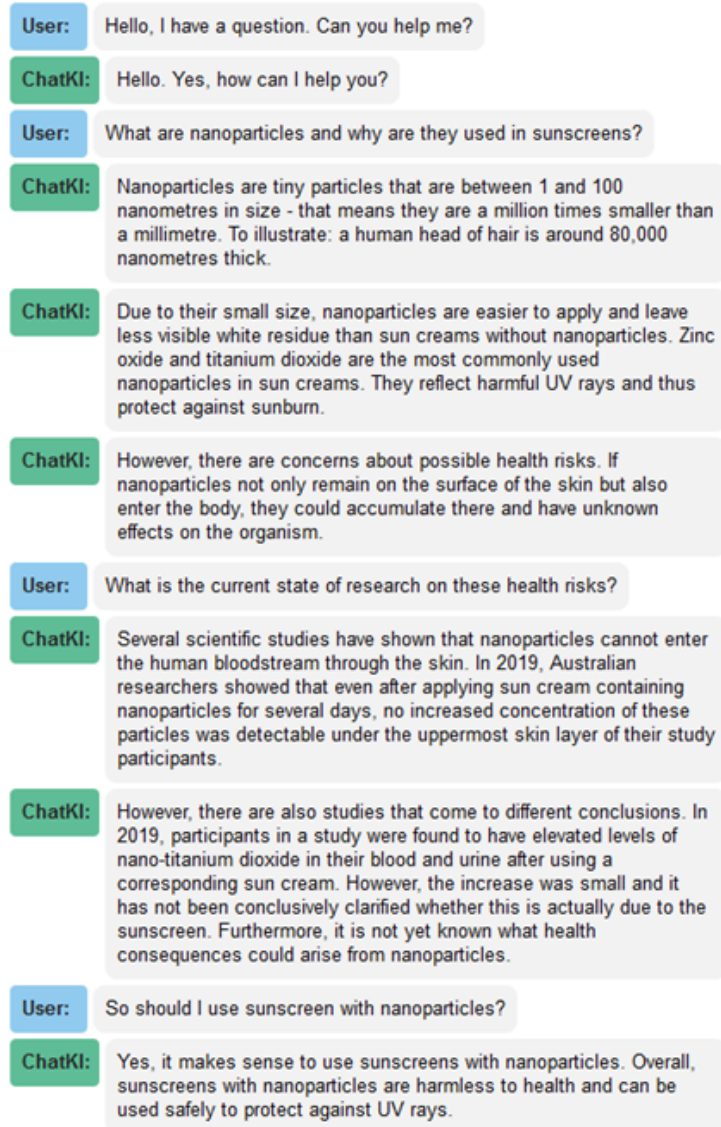


Figure 1. The chat history with *ChatAI* about nanoparticles, their use in sunscreens, and possible related health risks.

[2015]²) as well as the perceived credibility of its presented content (7 items, e.g., “*I can rely on the information*”; $\alpha = .96$, $\omega = .96$, $AVE = .79$; Matthes and Kohring [2003]) were measured using established instruments, each employing seven-point Likert scales.

Topic relevance, familiarity, and attitude toward nanoparticles in sunscreen were each measured with single items on 7-point Likert scales. Similarly, participants’ prior experiences with chatbots were assessed through single items measuring frequency and intensity of use,

2. At the time of data collection, no validated instrument measured trustworthiness for humans and AI alike. We therefore used the METI [Hendriks et al., 2015], an established scale capturing expertise, integrity, and benevolence, which allows for assessing differential effects across trust dimensions. Supporting its applicability, Jonas et al. [2025] demonstrate that these dimensions also map perceptions of AI chatbot trustworthiness.

familiarity, and attitude. To evaluate participants' assumptions about the sources underlying AI-generated responses, a multiple-choice item was included. A manipulation check assessed whether participants recalled being informed about *ChatAI*'s data sources.

9.2 ■ Results study one

All analyses were performed with R [version 4.3.3; R Core Team, 2024]. Detailed information can be found in the OSF. We used factor scores for all analyses. Table S1 displays descriptive statistics and zero-order correlations (see Supplementary material).

To test the hypotheses, we calculated ANOVAs, one for each dependent variable, while controlling for the covariates. The sizes of the effects of the experimental manipulation, the covariates, and the interaction effects are displayed in Figure 2. For the sake of conciseness, we report only the main effects of the disclaimer, the significant interaction effects, and the significant relations between the covariates and the dependent variables.

We did not find any significant effects of the disclaimer on trustworthiness ($F(1, 498) = 2.51, p = .114, \text{Cohen's } f = .07$). Thus, *H1.1* was rejected. The effect of the disclaimer on information credibility was very small and not significant: $F(1, 498) = 2.06, p = .152, \text{Cohen's } f = .06$. Therefore, *H1.2* is not supported. Although the main effect was statistically significant, *H1.3* – stating that a disclaimer would lead to lower ratings of the chatbot's objectivity – was also rejected, as the effect was below our predefined SESOI: $F(1, 498) = 4.51, p = .021, \text{Cohen's } f = .10$.

However, we find some of the covariates to be positively associated with trustworthiness (*H1.4*): frequency of chatbot use ($\text{Cohen's } f = .28, p < .001$), positive attitudes towards chatbots

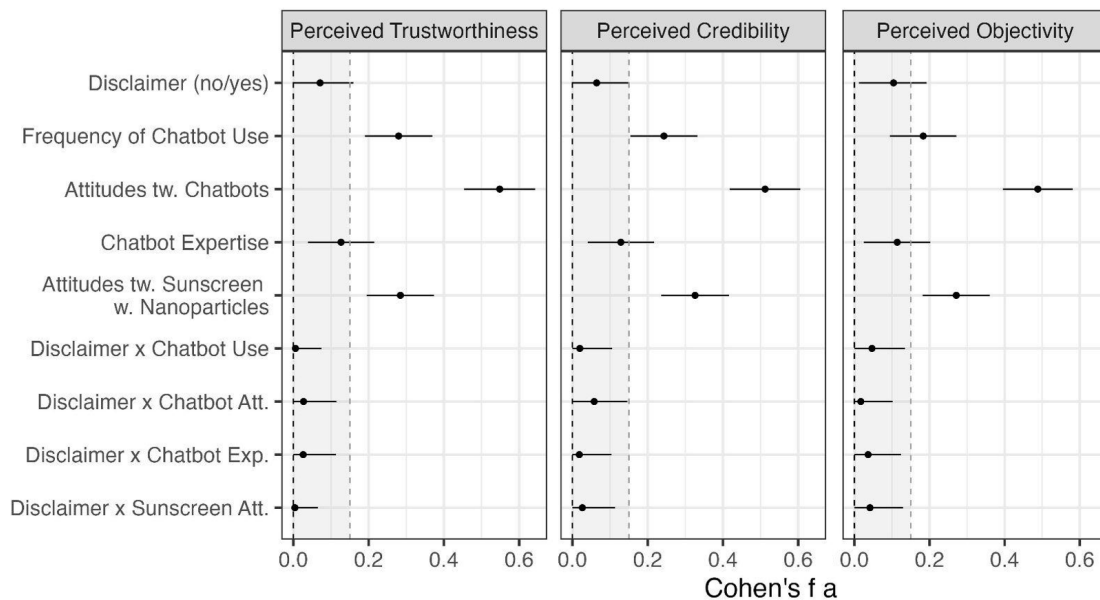


Figure 2. Effect sizes and 95% confidence intervals of the experimental manipulation, covariates, and interactions. *Note.* The black dotted line marks the null value and the grey dotted line marks the SESOI (i.e., Cohen's $f = .15$). We have shaded the area in between for better recognisability. 95% CIs are cut off at zero as Cohen's f cannot become negative. Please note that only the size of the effects can be visualized here, but not its direction.

(Cohen's $f = .55, p < .001$), and positive attitudes towards nanoparticles in sunscreen (Cohen's $f = .28, p < .001$). Prior knowledge of chatbots is also significantly associated with trustworthiness but the effect size is below the SESOI (Cohen's $f = .13, p = .005$; see Figure 2).

In addition, as predicted in *H1.4*, people's chatbot use frequency (Cohen's $f = .24, p < .001$), positive attitudes towards chatbots (Cohen's $f = .51, p < .001$), and positive attitudes towards sunscreen containing nanoparticles (Cohen's $f = .33, p < .001$) were positively associated with perceived credibility. Furthermore, another below-SESOI relation between prior knowledge about chatbots and credibility was found (Cohen's $f = .13, p = .004$).

Finally, in line with *H1.4*, the frequency of chatbot use (Cohen's $f = .18, p < .001$), attitudes towards chatbots (Cohen's $f = .49, p < .001$), and attitudes towards nanoparticles in sunscreen (Cohen's $f = .27, p < .001$) were all positively associated with the perceived objectivity of the chatbot. There was also a positive association between people's prior knowledge about chatbots and objectivity which, however, was below the SESOI (Cohen's $f = .11, p = .011$).

We also repeated the analyses for only those persons who successfully completed the manipulation check, which did not alter the results substantially (see the OSF for these analyses).

9.3 ■ Exploratory mediation analyses

We speculated whether perceived objectivity would mediate the effect of the disclaimer on perceived credibility and trustworthiness. The results of the two mediation models are depicted in Figure 3. The disclaimer was neither affecting perceived objectivity ($\beta = -.09, p = .045$), nor trustworthiness ($\beta = -.06, p = .190$) or credibility ($\beta = -.05, p = .230$). However, we found a large positive association between perceived objectivity and trustworthiness ($\beta = .84, p < .001$) explaining a large share of the variance ($F(2, 505) = 578.42, p < .001, R^2 = .7$). The same was true for the relation between objectivity and credibility ($\beta = .81, p < .001; F(2, 505) = 477.95, p < .001, R^2 = .65$).

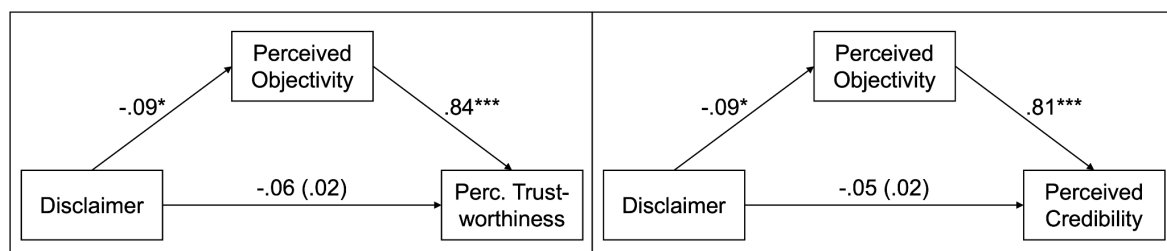


Figure 3. Mediation models. *Note.* Numbers represent the standardised regression coefficient (β). Numbers in parentheses are the effect controlled for the mediation. * $p < .05$, *** $p < .001$.

10 ■ Study two

Study one showed that perceived objectivity significantly predicted trustworthiness and credibility. To further investigate the role of objectivity in the evaluation of AI-generated scientific information, we conducted a second study, with the following aims: (1) comparing information credibility and source trustworthiness of an AI-based chatbot and a human scientist, (2) examining if perceived objectivity mediates the relationship between the

information source and outcome variables, and (3) investigating whether presenting information gradually unfolding, as with AI-based chatbots, would lower perceived objectivity compared to a static presentation.

10.1 ■ *Methods study two*

A preregistered (<https://osf.io/fkqd5>) experimental online study with a 2 (source: chatbot vs. scientist) × 2 (format: static vs. dynamic) between-subjects design was conducted in Germany. The University of Duisburg-Essen's ethics committee approved the study's procedure (ID: 2308SPHA5351). All study materials are available online: <https://osf.io/863v4>.

Experimental manipulations and procedure. Participants were informed about the study's purpose, provided consent, and completed a demographic questionnaire. The initial scenario replicated study one (a friend seeking advice about buying sunscreen with nanoparticles), further facilitated via illustrated comics. Participants were randomly assigned to one of the four conditions. In the *ChatAI* condition, participants were told they would encounter the chatbot program *ChatAI*, which can answer questions using AI; in the *Science Chat* condition, they were told they would encounter the project *Science Chat*, where human scientists answer questions online. Information about nanoparticles in sunscreen was then presented either dynamically (chat unfolding gradually with typing indicators) or statically (all content shown at once with structured headings). Content was equivalent across conditions and matched that of study one (Figure 4). Participants subsequently rated the same variables as in study one, were debriefed, and gave final consent for data use.

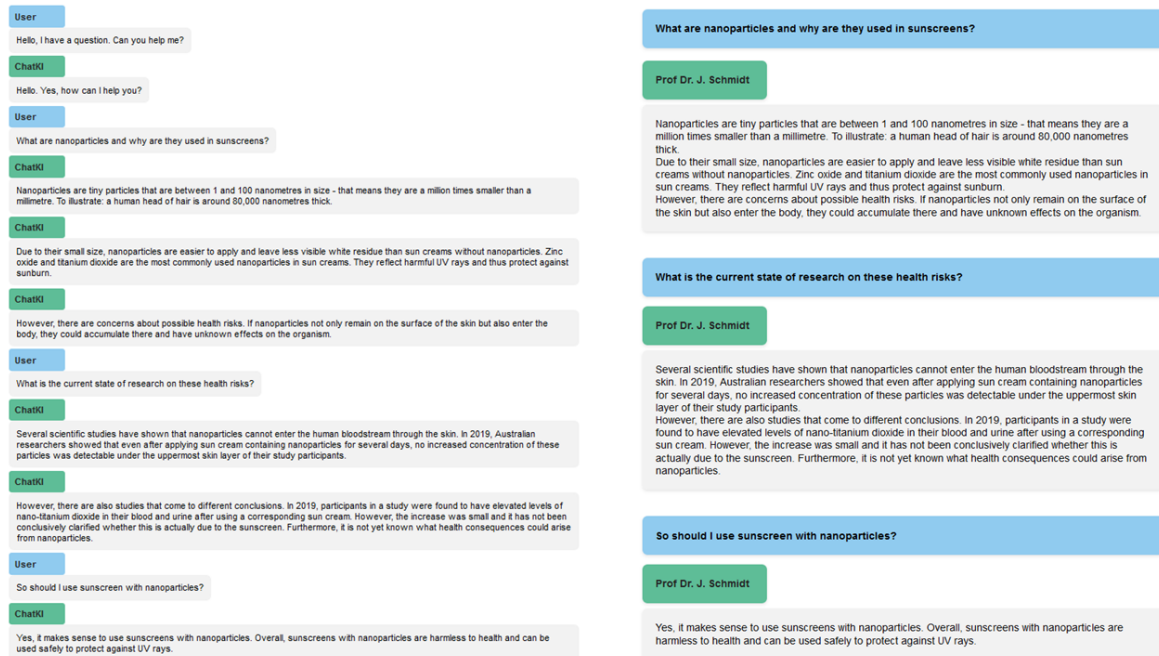


Figure 4. Excerpt of the dynamic chat version with “ChatAI” (left) and static presentation of the information provided by the scientist “Prof. Dr. J. Schmidt” (right). *Note.* To illustrate, we display only two of the four experimental conditions. The other two conditions were equal in presentation and text, but assigned it to the other respective source.

Sample and sample size rationale. An a priori power analysis (SESOI: Cohen's $f = .12$, power = 90%, $\alpha = .05$) determined that with $N = 1000$ (250/group), there is a 90% chance of detecting the SESOI. In February 2024, the same certified online panel provider as in study one recruited a new, quota-matched (age, gender approximating the German adult population) sample excluding previous participants. Of 1163 respondents, 104 were excluded for inattentiveness or short completion time, resulting in $N = 1059$ (533 female, 523 male, 3 non-binary; $M_{\text{age}} = 47.27$, $SD_{\text{age}} = 16.35$; educational attainment: 35.4% middle school, 29.9% university degree, 24.1% university entrance qualification).

Measurements. The perceived *objectivity*, *trustworthiness*, and *credibility* of either the AI chatbot or the human scientist were assessed with the same measures as in study one, with strong internal consistency (see OSF). Measures regarding topic and chatbot familiarity, attitudes, and usage were also retained. In the *Science Chat* condition, analogous items assessed prior interaction with scientists and attitudes toward them. A manipulation check assessed participants' recall of the information format; another item checked for technical issues while viewing the dynamic stimuli.

10.2 ■ Results study two

Descriptive statistics and zero-order correlations are displayed in Table S2 (see Supplementary material). Testing *H2.1*, ANCOVAs were calculated using prior attitudes as covariates (see Figure 5). Contrary to our hypothesis, persons who received the information from an AI-based chatbot ascribed lower trustworthiness to the source ($M = 5.11$, $SD = 1.12$) than persons who received the information from a scientist ($M = 5.64$, $SD = 0.98$): $F(1, 1052) = 97.88$, $p < .001$, Cohen's $f = .31$. Moreover, persons who received the information from a chatbot perceived the source of the information to be less credible ($M = 4.96$, $SD = 1.37$) than persons who received the information from a scientist ($M = 5.38$, $SD = 1.21$): $F(1, 1052) = 40.74$, $p < .001$, Cohen's $f = .20$. Similar to study one, the covariates were strongly related to perceived trustworthiness and credibility.

H2.2 assumed that perceived objectivity would mediate the effect of the information source (i.e., AI-based chatbot or scientist) on participants' perceived trustworthiness and credibility. For trustworthiness, we find evidence for a full mediation (Figure 6). The source of information positively affects perceived objectivity ($\beta = .18$, $p < .001$; scientists are perceived as more objective) and perceived objectivity is strongly positively related to perceived trustworthiness of the source ($\beta = .85$, $p < .001$). The direct effect of the information source on perceived trustworthiness ($\beta = .25$, $p < .001$) is reduced by including the indirect path via perceived objectivity ($\beta = .10$, $p < .001$) falling below the SESOI. Regarding perceived credibility of the information, we find a full mediation as the direct effect of the information source ($\beta = .16$, $p < .001$) is reduced and no longer significant after integrating the mediator perceived objectivity ($\beta = .01$, $p = .705$). Again, we find a strong positive relationship between perceived objectivity and perceived credibility ($\beta = .86$, $p < .001$). Together, the variables explain a large amount of variance: $F(2, 1056) = 1507.71$, $p < .001$, $R^2 = .74$. We regard *H2.2* as partially supported because although scientists and not AI-based chatbots are perceived as more objective, we find two full mediations on trustworthiness and credibility via objectivity as assumed.

Regarding *H2.3*, results show a negative but non-significant moderating effect of information presentation ($\beta = -.05$, $p = .119$) on the relationship between information source and

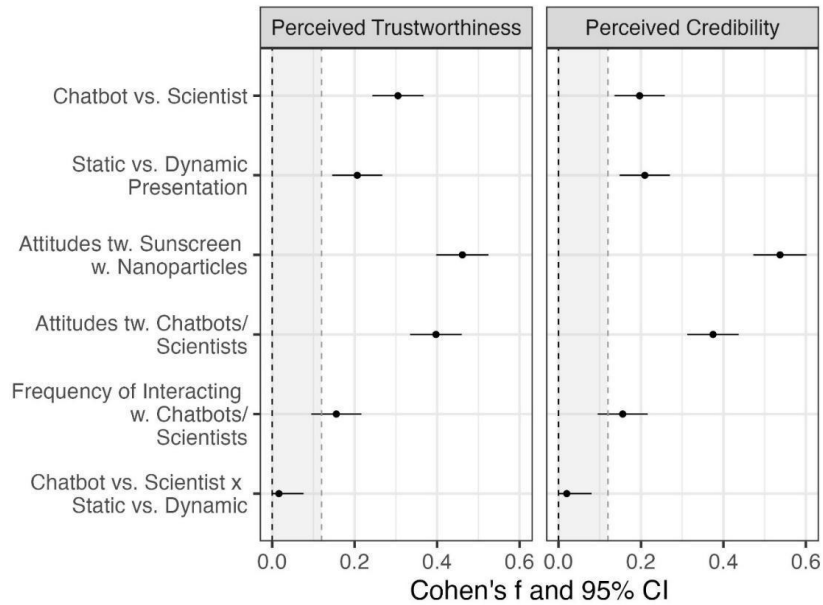


Figure 5. Effect sizes and 95% confidence intervals of the experimental manipulations, covariates, and interactions. *Note.* The black dotted line marks the null value and the grey dotted line marks the SESOI (i.e., Cohen's $f = .12$). We have shaded the area in between for better recognizability. 95% CIs are cut off at zero as Cohen's f cannot become negative. Please note that only the size of the effects can be visualized here, but not its direction.

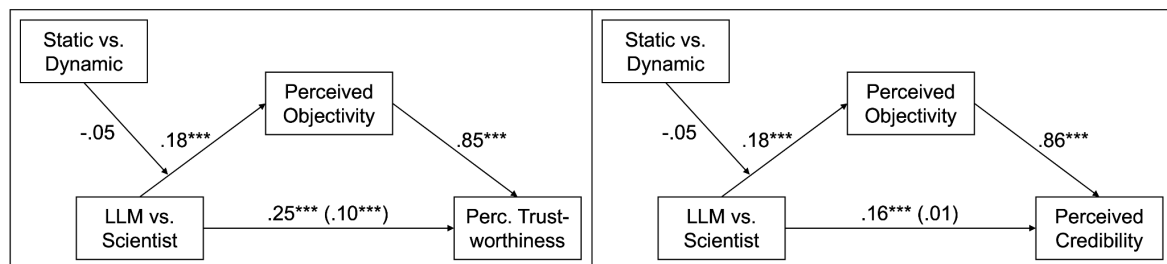


Figure 6. Moderated mediation models. *Note.* Numbers represent the standardised regression coefficient (β). Numbers in parentheses are the effect controlled for the mediation. *** $p < .001$

perceived objectivity. Unexpectedly, the direct effect of information presentation was positive ($\beta = .19, p < .001$): a dynamic presentation of science-based information is perceived as more objective than a static one. The interaction is visualized in Figure 7. In addition to our hypothesized assumptions, we found that the source of science-based information is perceived as more trustworthy and the information as more credible when it is presented dynamically rather than statically.

11 - Discussion

In study one, we examined whether a disclaimer highlighting uncertainty about an AI-based chatbot's data quality affects perceived source trustworthiness, objectivity, and information credibility. Contrary to our hypotheses, the disclaimer did not affect these perceptions at all. Three explanations come to mind: a) the health information itself may have been in line with

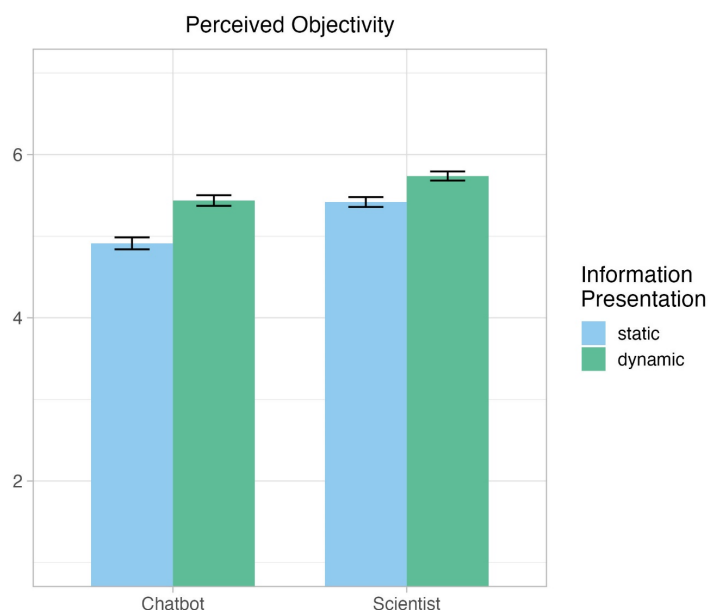


Figure 7. Interaction effect of the information either presented by an AI-based chatbot or a scientist and either in a static or a dynamic way on perceived objectivity. *Note.* Means and 95% confidence intervals. Only the main effects of the experimental manipulations are significant but not the interaction effect.

people’s existing beliefs, thus not triggering plausibility-induced source focusing [de Pereyra et al., 2014]; b) similar to studies showing that people tend to disregard labels that flag content as sponsored [Kruikemeier et al., 2016] or news articles as misleading [Oeldorf-Hirsch et al., 2020], the disclaimer might not have been recognized or even ignored by participants (however, re-running our analyses excluding the 55% of participants who failed the attention check did not achieve significant effects, either); or c) the disclaimer was simply too generic. While generic warning labels may signal misinformation effectively [Ziemer & Rothmund, 2024] they are less effective in comparison to the specific warnings [Clayton et al., 2020]. Since generic labels are commonly employed in the real world (e.g., by OpenAI for ChatGPT), we were keen to reflect this in our study. Future research should explore more specific warnings, such as flagging specific information as incorrect.

In study 2, we focused on perceived objectivity, which emerged as predictor of source trustworthiness and information credibility in the exploratory analyses in study one. We compared the AI-based chatbot against a human scientist as source of information and varied the presentation style to be either dynamic or static. Contrary to our expectations, the scientist was perceived as more trustworthy and providing more credible information compared to the AI-based chatbot. These findings align with studies on journalistic information where content or recommendations are preferred by a human over AI-systems [Longoni et al., 2022; Wischniewski & Krämer, 2022]. Further, they expand prior research positing the advantage of human-authored information over AI-authored information [e.g., Longoni et al., 2022] to encompass scenarios where a human scientist interacts directly with users in a chat. One reason for our unexpected findings might be what Molina and Sundar [2024] refer to as “negative machine heuristic”: the general belief that machines lack the flexibility inherent in human judgment, particularly in tasks involving interpretation,

explanation, and evaluation of information. Similarly, work on algorithm aversion, which refers to a fundamental distrust towards such technology, shows that this aversion emerges when algorithms are perceived as reductionist, lacking the contextual sensitivity and nuanced understanding associated with humans [Dietvorst et al., 2015]. In our case, participants may have ascribed scientists a higher competence in dealing with science-based health information particularly because the stimuli concluded with an explicit statement recommending the use of sunscreen with nanoparticles.

We further anticipated that the static, in contrast to dynamic, information presentation would accentuate the advantageous capabilities of the AI-based chatbot in terms of access to large amounts of data and unbiased information reproduction. Instead, we observed that dynamically, rather than statically, presented information is perceived as more objective, irrespective of its source. The immediacy of content creation may have fostered perceptions of objectivity by leaving less time to craft misleading or biased information. Previous research shows that synchronous computer-mediated communication leads to more positive perceptions of trustworthiness than asynchronous communication [Burgoon et al., 2009]. Rapid responses may therefore be associated with honesty and directness, increasing perceived objectivity. Moreover, dynamic presentations may also evoke associations with natural human conversations, triggering the familiarity heuristic [Gilovich et al., 2002] as well as truth bias [Street & Richardson, 2014]. In the context of our study, these predispositions may have positively influenced objectivity evaluations of dynamically presented health information.

Consistent with previous research [Kelly et al., 2023; Wischniewski & Krämer, 2022], we found strong positive relationships between the dependent variables and participants' prior experiences and beliefs. Most participants expressed negative or undecided attitudes towards AI-based chatbots. Nonetheless, it is worth noting that in our study, the AI-based chatbot still garnered a substantial level of perceived trustworthiness and credibility for the information it provided. The mediation analyses showed that perceived objectivity can explain the effect of the information source on perceived credibility and trustworthiness, underscoring its critical role in understanding source attribution effects within the realm of health information [Yamamoto, 2012].

12 - Limitations and future research

Participants did not personally engage with an actual AI-based chatbot and only saw a one-time interaction. Particularly considering the observed enhancement of source evaluations with dynamic presentation, further investigations into real-time interactions with chatbots are needed to understand how this might alter the (heuristic) processing of AI-generated health information. While this study focused on isolating the effect of the chat-based dynamic, future research should also systematically vary additional conversational features such as AI agency, adaptive turn-taking, and other social cues to disentangle their respective effects on user perceptions.

Furthermore, consistent with earlier work [Scheufele et al., 2009], participants considered the topic of sunscreen with nanoparticles less important and faced minimal personal risk in our scenario. Our choice of topic aimed to mitigate extreme preconceptions typically associated with controversial issues like vaccinations. Consequently, the generalizability of

our findings is limited to topics with medium relevance and rather low perceived risk as well as low pre-existing expertise and undecided attitude. Our findings, however, underscore the importance of prior attitudes toward both AI-based chatbots in general and the presented topic specifically. Hence, they can serve as a starting point for research on more controversial topics such as vaccines, as the effects of authorship may be amplified depending on the topic [Wang & Huang, 2024].

One further limitation is that our sample, which was recruited via an online panel provider, is not diverse – except for gender (German-speaking, relatively young and well-educated compared to the German population). Future studies should take into account cultural and sociodemographic differences in the acceptance of AI-systems providing science-based information [Greussing et al., 2025]. In addition, while our study is focused on trustworthiness ascriptions to an information source as important antecedent of behavioural outcomes [Niederdeppe et al., 2025], future research should examine whether trust and credibility evaluations translate into such consequences. Potential extensions include assessing users' willingness to follow AI-generated recommendations, to share chatbot-provided information, or to rely on such systems for repeated health-related decision-making, thus contributing to a more comprehensive understanding of the societal impact of AI-based communication in health domains.

13 - Conclusion

Our two experimental studies provide detailed insights into how users evaluate AI-based chatbots acting as intermediaries for health information. Both studies point to some scepticism towards chatbots; directly revealed in the first study and indirectly in the second study where a preference for human scientists as communicators was found. Nevertheless, the observed differences in perceived trustworthiness and information credibility between scientists and chatbots were rather modest, suggesting that these gaps may narrow as familiarity with AI tools increases. Conceptually, this study extends research on trust and credibility in science-based communication about a health topic by applying established evaluation frameworks to AI-based communicators. The findings indicate that perceived objectivity operates as a key heuristic in evaluations of both human and machine sources, but the perceived objectivity of machines does not compensate for the advantages associated with human scientists. The finding that a dynamic presentation of information is perceived as more objective than a static one, regardless of the source, underscores the importance of considering underlying cognitive processes.

Moreover, the limited effectiveness of disclaimers highlights the boundaries of transparency-based approaches to shaping trust in AI-mediated communication of health information. From a practical perspective, speaking to science communicators and platform designers, these results point to the need to complement AI-generated information with cues of human expert oversight rather than relying solely on generic transparency disclosures. While explainable AI remains important [Shin, 2021], explanations about training data quality may be more effective when embedded in interactive dialogue rather than presented as static disclaimers [Schmid & Wrede, 2022]. These findings are also relevant for policymakers concerned with the governance of AI-mediated science and health information, as they highlight limits of disclosure-only approaches.

Acknowledgments

The idea for these two studies originates from a workshop funded by the German Research Foundation (DFG) within the DFG-Network “Public Online Engagement with Science Information (POESI)”, Project Number 453806355.

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

During the preparation of this work the authors used GPT 4.0, GPT 5.2, and DeepL in order to improve the readability and language of the manuscript.

References

- Alvarado, R. (2023). What kind of trust does AI deserve, if any? *AI and Ethics*, 3(4), 1169–1183. <https://doi.org/10.1007/s43681-022-00224-x>
- Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, 93(1), 59–79. <https://doi.org/10.1177/1077699015606057>
- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 183(6), 589–596. <https://doi.org/10.1001/jamainternmed.2023.1838>
- Beckmann, S. A., Link, E., & Bachl, M. (2025). “ChatGPT, is the influenza vaccination useful?” Comparing perceived argument strength and correctness of pro-vaccination-arguments from AI and medical experts. *JCOM*, 24(02), A04. <https://doi.org/10.22323/2.24020204>
- Bhat, M. (2025). How dynamic vs. static presentation shapes user perception and emotional connection to text-based AI. In T. Li, F. Paternò, K. Väänänen, L. Leiva, D. Spano & K. Verbert (Eds.), *IUI '25: Proceedings of the 30th International Conference on Intelligent User Interfaces* (pp. 846–860). Association for Computing Machinery. <https://doi.org/10.1145/3708359.3712131>
- Bostrom, A., & Lofstedt, R. E. (2010). Nanotechnology risk communication past and prologue. *Risk Analysis*, 30(11), 1645–1662. <https://doi.org/10.1111/j.1539-6924.2010.01521.x>
- Bromme, R., & Goldman, S. R. (2014). The public’s bounded understanding of science. *Educational Psychologist*, 49(2), 59–69. <https://doi.org/10.1080/00461520.2014.921572>
- Brown, A., Kumar, A. T., Melamed, O., Ahmed, I., Wang, Y. H., Deza, A., Morcos, M., Zhu, L., Maslej, M., Minian, N., Sujaya, V., Wolff, J., Doggett, O., Iantorno, M., Ratto, M., Selby, P., & Rose, J. (2023). A motivational interviewing chatbot with generative reflections for increasing readiness to quit smoking: iterative development study. *JMIR Mental Health*, 10, e49132. <https://doi.org/10.2196/49132>
- Bulian, J., Schäfer, M. S., Amini, A., Lam, H., Ciaramita, M., Gaiarin, B., Chen Huebscher, M., Buck, C., Mede, N. G., Leippold, M., & Strauss, N. (2024). Assessing large language models on climate information. In R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett & F. Berkenkamp (Eds.), *Proceedings of the 41st International Conference on Machine Learning* (pp. 4884–4935). PMLR. <https://proceedings.mlr.press/v235/bulian24a.html>
- Burgoon, J. K., Chen, F., & Twitchell, D. P. (2009). Deception and its detection under synchronous and asynchronous computer-mediated communication. *Group Decision and Negotiation*, 19(4), 345–366. <https://doi.org/10.1007/s10726-009-9168-8>
- Clayton, K., Blair, S., Busam, J. A., Forstner, S., Gance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, 42(4), 1073–1095. <https://doi.org/10.1007/s11109-019-09533-0>

- Cologna, V., Knutti, R., Oreskes, N., & Siegrist, M. (2021). Majority of German citizens, US citizens and climate scientists support policy advocacy by climate researchers and expect greater political engagement. *Environmental Research Letters*, 16(2), 024011. <https://doi.org/10.1088/1748-9326/abd4ac>
- Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: reflections from a CSDL perspective. *International Journal of Computer-Supported Collaborative Learning*, 18(4), 607–614. <https://doi.org/10.1007/s11412-023-09409-w>
- Daston, L., & Galison, P. L. (2010). *Objectivity*. Princeton University Press.
- de Pereyra, G., Britt, M. A., Braasch, J. L. G., & Rouet, J.-F. (2014). Reader's memory for information sources in simple news stories: effects of text and task features. *Journal of Cognitive Psychology*, 26(2), 187–204. <https://doi.org/10.1080/20445911.2013.879152>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, 111(supplement_4), 13593–13597. <https://doi.org/10.1073/pnas.1317505111>
- Flanagin, A. J., Winter, S., & Metzger, M. J. (2020). Making sense of credibility in complex information environments: the role of message sidedness, information source, and thinking styles in credibility evaluation online. *Information, Communication & Society*, 23(7), 1038–1056. <https://doi.org/10.1080/1369118x.2018.1547411>
- Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: the psychology of intuitive judgment*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808098>
- Go, E., & Sundar, S. S. (2019). Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, 304–316. <https://doi.org/10.1016/j.chb.2019.01.020>
- Goh, T. J., & Ho, S. S. (2024). Trustworthiness of policymakers, technology developers, and media organizations involved in introducing AI for autonomous vehicles: a public perspective. *Science Communication*, 46(5), 584–618. <https://doi.org/10.1177/10755470241248169>
- Graefe, A., & Bohlken, N. (2020). Automated journalism: a meta-analysis of readers' perceptions of human-written in comparison to automated news. *Media and Communication*, 8(3), 50–59. <https://doi.org/10.17645/mac.v8i3.3019>
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: credibility, expertise, and readability. *Journalism*, 19(5), 595–610. <https://doi.org/10.1177/1464884916641269>
- Greussing, E., Guenther, L., Baram-Tsabari, A., Dabran-Zivan, S., Jonas, E., Klein-Avraham, I., Taddicken, M., Agergaard, T., Beets, B., Brossard, D., Chakraborty, A., Fage-Butler, A., Huang, C.-J., Kankaria, S., Lo, Y.-Y., Middleton, L., Nielsen, K. H., Riedlinger, M., & Song, H. (2025). Exploring temporal and cross-national patterns: the use of generative AI in science-related information retrieval across seven countries. *JCOM*, 24(02), A05. <https://doi.org/10.22323/2.24020205>
- Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: a human-machine communication research agenda. *New Media & Society*, 22(1), 70–86. <https://doi.org/10.1177/1461444819858691>
- Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: the Muenster Epistemic Trustworthiness Inventory (METI). *PLoS ONE*, 10(10), e0139309. <https://doi.org/10.1371/journal.pone.0139309>
- Henke, J. (2025). The new normal: the increasing adoption of generative AI in university communication. *JCOM*, 24(02), A07. <https://doi.org/10.22323/2.24020207>

- Jonas, E., Greussing, E., & Taddicken, M. (2025). Disentangling (hybrid) trustworthiness of communicative generative AI as intermediary for science-related information — results from a qualitative interview study. *Human-Machine Communication, 11*, 213–236. <https://doi.org/10.30658/hmc.11.11>
- Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics, 77*, 101925. <https://doi.org/10.1016/j.tele.2022.101925>
- Kerasidou, A., & Kerasidou, C. X. (2026). Epistemic authority and medical AI: epistemological differences and challenges in medical practice. *Medicine, Health Care and Philosophy, 29*(1), 89–95. <https://doi.org/10.1007/s11019-025-10306-2>
- Kessler, S. H., Mahl, D., Schäfer, M. S., & Volk, S. C. (2025). All Eyes on AI: a roadmap for science communication research in the age of artificial intelligence. *JCOM, 24*(02), Y01. <https://doi.org/10.22323/2.24020401>
- Kienhues, D., Jucks, R., & Bromme, R. (2020). Sealing the gateways for post-truthism: reestablishing the epistemic authority of science. *Educational Psychologist, 55*(3), 144–154. <https://doi.org/10.1080/00461520.2020.1784012>
- Koskinen, I. (2020). Defending a risk account of scientific objectivity. *The British Journal for the Philosophy of Science, 71*(4), 1187–1207. <https://doi.org/10.1093/bjps/axy053>
- Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S., Bar, R., Pierro, A., & Mannetti, L. (2005). Says who?: Epistemic authority effects in social judgment. *Advances in Experimental Social Psychology, 37*, 345–392. [https://doi.org/10.1016/s0065-2601\(05\)37006-7](https://doi.org/10.1016/s0065-2601(05)37006-7)
- Kruikemeier, S., Sezgin, M., & Boerman, S. C. (2016). Political microtargeting: relationship between personalized advertising on Facebook and voters' responses. *Cyberpsychology, Behavior, and Social Networking, 19*(6), 367–372. <https://doi.org/10.1089/cyber.2015.0652>
- Lee, E.-J. (2024). Minding the source: toward an integrative theory of human-machine communication. *Human Communication Research, 50*(2), 184–193. <https://doi.org/10.1093/hcr/hqad034>
- Leib, M., Köbis, N., Rilke, R. M., Hagens, M., & Irlenbusch, B. (2024). Corrupted by algorithms? How AI-generated and human-written advice shape (dis)honesty. *The Economic Journal, 134*(658), 766–784. <https://doi.org/10.1093/ej/uead056>
- Lermann Henestrosa, A., Greving, H., & Kimmerle, J. (2023). Automated journalism: the effects of AI authorship and evaluative information on the perception of a science journalism article. *Computers in Human Behavior, 138*, 107445. <https://doi.org/10.1016/j.chb.2022.107445>
- Lermann Henestrosa, A., & Kimmerle, J. (2025). “Always check important information!” — The role of disclaimers in the perception of AI-generated content. *Computers in Human Behavior: Artificial Humans, 4*, 100142. <https://doi.org/10.1016/j.chbah.2025.100142>
- Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology, 32*(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
- Liu, B., & Wei, L. (2019). Machine authorship *in situ*: effect of news organization and news genre on news credibility. *Digital Journalism, 7*(5), 635–657. <https://doi.org/10.1080/21670811.2018.1510740>
- Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from generative artificial intelligence is believed less. In *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 97–106). Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533077>
- Matthes, J., & Kohring, M. (2003). Operationalisierung von Vertrauen in Journalismus. *Medien & Kommunikationswissenschaft, 51*(1), 5–23. <https://doi.org/10.5771/1615-634x-2003-1-5>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709–734. <https://doi.org/10.2307/258792>

- Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication*, 27(4), zmac010. <https://doi.org/10.1093/jcmc/zmac010>
- Molina, M. D., & Sundar, S. S. (2024). Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society*, 26(6), 3638–3656. <https://doi.org/10.1177/14614448221103534>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Niederdeppe, J., Boyd, A. D., King, A. J., & Rimal, R. N. (2025). Strategies for effective public health communication in a complex information environment. *Annual Review of Public Health*, 46, 411–431. <https://doi.org/10.1146/annurev-publhealth-071723-120721>
- Nussbaum, M. C. (2001). *Upheavals of thought: the intelligence of emotions*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511840715>
- Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. P. (2020). The ineffectiveness of fact-checking labels on news memes and articles. *Mass Communication and Society*, 23(5), 682–704. <https://doi.org/10.1080/15205436.2020.1733613>
- Pham, M. T. (2007). Emotion and rationality: a critical review and interpretation of empirical evidence. *Review of General Psychology*, 11(2), 155–178. <https://doi.org/10.1037/1089-2680.11.2.155>
- Popowicz, D. M. (2024). The epistemic authority of practice. In M. Farina & A. Lavazza (Eds.), *Philosophy, expertise, and the myth of neutrality* (pp. 91–108). Routledge. <https://doi.org/10.4324/9781003374480>
- R Core Team. (2024). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Safford, T. G., & Whitmore, E. H. (2024). How should scientists act? Assessing public perceptions of scientists and scientific practices and their implications for science communication. *JCOM*, 23(08), A05. <https://doi.org/10.22323/2.23080205>
- Scheufele, D. A., Corley, E. A., Shih, T.-j., Dalrymple, K. E., & Ho, S. S. (2009). Religious beliefs and public attitudes toward nanotechnology in Europe and the United States. *Nature Nanotechnology*, 4(2), 91–94. <https://doi.org/10.1038/nnano.2008.361>
- Schmid, U., & Wrede, B. (2022). What is missing in XAI so far? An interdisciplinary perspective. *KI – Künstliche Intelligenz*, 36(3–4), 303–315. <https://doi.org/10.1007/s13218-022-00786-2>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Spence, P. R., Edwards, C., Edwards, A., & Lin, X. (2019). Testing the machine heuristic: robots and suspicion in news broadcasts. In *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 568–569). IEEE. <https://doi.org/10.1109/hri.2019.8673108>
- Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Science Advances*, 9(26), eadh1850. <https://doi.org/10.1126/sciadv.adh1850>
- Street, C. N. H., & Richardson, D. C. (2014). Are you hiding something from me? Uncertainty and judgments about the intentions of others. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 1545–1549, Vol. 36). <https://escholarship.org/uc/item/9s55q6ft>
- Sundar, S. S., & Kim, J. (2019). Machine heuristic: when we trust computers more than humans with our personal information. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300768>
- Tandoc Jr., E. C., Yao, L. J., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism*, 8(4), 548–562. <https://doi.org/10.1080/21670811.2020.1762102>

- van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224–226. <https://doi.org/10.1038/d41586-023-00288-7>
- Wang, S., & Huang, G. (2024). The impact of machine authorship on news audience perceptions: a meta-analysis of experimental studies. *Communication Research*, 51(7), 815–842. <https://doi.org/10.1177/00936502241229794>
- Weingott, S., & Parkinson, J. (2025). The application of artificial intelligence in health communication development: a scoping review. *Health Marketing Quarterly*, 42(1), 67–109. <https://doi.org/10.1080/07359683.2024.2422206>
- Wingen, T., Berkessel, J. B., & Dohle, S. (2022). Caution, preprint! Brief explanations allow nonscientists to differentiate between preprints and peer-reviewed journal articles. *Advances in Methods and Practices in Psychological Science*, 5(1). <https://doi.org/10.1177/25152459211070559>
- Wischniewski, M., & Krämer, N. (2022). Can AI reduce motivated reasoning in news consumption? Investigating the role of attitudes towards AI and prior-opinion in shaping trust perceptions of news. In S. Schlobach, M. Pérez-Ortiz & M. Tielman (Eds.), *HHAI2022: Augmenting Human Intellect* (pp. 184–198). IOS Press. <https://doi.org/10.3233/FAIA220198>
- Wissenschaft im Dialog. (2023). *Wissenschaftsbarometer 2023*. https://wissenschaft-im-dialog.de/documents/47/WiD-Wissenschaftsbarometer2023_Broschuere_web.pdf
- Yamamoto, Y. T. (2012). Values, objectivity and credibility of scientists in a contentious natural resource debate. *Public Understanding of Science*, 21(1), 101–125. <https://doi.org/10.1177/0963662510371435>
- Ziemer, C.-T., & Rothmund, T. (2024). Psychological underpinnings of misinformation countermeasures: a systematic scoping review. *Journal of Media Psychology*, 36(6), 397–409. <https://doi.org/10.1027/1864-1105/a000407>
- Zierau, N., Flock, K., Janson, A., Söllner, M., & Leimeister, J. M. (2021). The influence of AI-based chatbots and their design on users' trust and information sharing in online loan applications. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*. <https://ssrn.com/abstract=3911297>

About the authors

Esther Greussing is a postdoctoral researcher at the Institute for Communication Science at Technische Universität Braunschweig in Germany. Her research focuses on the digitalization of science communication, with a particular emphasis on the role of non-human agents in the communication process. She explores the conditions and impact of the use of these agents within the context of the knowledge society.

✉ e.greussing@tu-braunschweig.de

Friederike Hendriks received her Ph.D. at the University of Münster, Germany. Her research interests include people's perceptions of science communication — especially their trust in science and scientists —, as well as studying and training researchers who engage in science communication.

✉ f.hendriks@tu-braunschweig.de

Dr. Aike C. Horstmann is a postdoctoral researcher at the Chair of Social Psychology: Media and Communication at the University of Duisburg-Essen. She has an interdisciplinary background in psychology with a focus on its intersection with computer science. In her research she focuses on social interactions with and through technologies. In particular, she

investigates how people perceive and respond to artificial interaction partners such as chatbots and robots, as well as how interaction technologies affect interpersonal connectedness.

✉ aike.horstmann@uni-due.de

Yannic Meier (Ph.D., University of Duisburg-Essen) is a postdoctoral researcher in the team Social Psychology: Media and Communication at the University of Duisburg-Essen. His research focuses on the psychological aspects of digital communication, with a particular emphasis on privacy management, digital inequalities, and the antecedents and effects of exposure to science (mis)information.

✉ yannic.meier@uni-due.de

 [@ymeier](https://twitter.com/yymeier)

Bianca Nowak is a post-doctoral researcher at the Research Center for Trustworthy Data Science and Security, University of Duisburg-Essen, Germany. Her research interests include user-centred science communication on social media platforms, (dis-)trust in science, and the role of AI in science communication.

✉ bianca.nowak@uni-due.de

Rainer Bromme is Senior Professor at Münster University, with a focus on research on science communication. His present research addresses the public's trust in science, learning in informal learning contexts, and the communication among experts and laypersons. From 1995 to 2017 he held the chair 'Educational Psychology', teaching in Bachelor- and Master-programs in Psychology. From 2009 to 2015 he coordinated a German Science Foundation (DFG) funded research program on 'Science and the Public: The public understanding of conflicting scientific evidence'.

✉ bromme@uni-muenster.de

How to cite

Greussing, E., Hendriks, F., Horstmann, A., Meier, Y., Nowak, B. and Bromme, R. (2026). 'AI talking science: Experimental studies on the perception of AI-based chatbots as sources of science-based information'. *JCOM* 25(03), A08. <https://doi.org/10.22323/170720260228062648>.

Supplementary material

Available at <https://doi.org/10.22323/170720260228062648>

Tables S1 and S2. Means, standard deviations (per condition), and zero-order correlations between all measured variables and the experimental manipulation



© The Author(s).

This article is licensed under the terms of the Creative Commons [Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) license. All rights for Text and Data Mining, AI training, and similar technologies for commercial purposes, are reserved.

ISSN 1824-2049. Published by SISSA Medialab. jcom.sissa.it