# Effects of framing and identity cues in science communication with and about AI

**Daniel Silva Luna, Helena Bilandzic, Martin Bürger**

**Angaben zur Veröffentlichung / Publication details:**

Silva Luna, Daniel, Helena Bilandzic, and Martin Bürger. 2026. "Effects of framing and identity cues in science communication with and about AI." *Media and Communication* 14: 11350. https://doi.org/10.17645/mac.11350.

Media and Communication
2026 • Volume 14 • Article 11350
https://doi.org/10.17645/mac.11350

ARTICLE

Open Access Journal

# Effects of Framing and Identity Cues in Science Communication With and About AI

Daniel Silva Luna, Helena Bilandzic, and Martin Bürger

Department of Media, Knowledge and Communication, University of Augsburg, Germany

**Correspondence:** Daniel Silva Luna (daniel.silva.luna@uni-a.de)

**Submitted:** 15 September 2025 **Accepted:** 11 February 2026 **Published:** 21 April 2026

**Issue:** This article is part of the issue "Exploring Engagement With Complex Information: Perspectives on Generative AI as an Information Intermediary" edited by Monika Taddicken (TU Braunschweig), Esther Greussing (TU Braunschweig), Evelyn Jonas (TU Braunschweig), Ayelet Baram-Tsabari (Technion—Israel Institute of Technology), and Inbal Klein-Avraham (Technion—Israel Institute of Technology), fully open access at https://doi.org/10.17645/mac.i509

## Abstract

As AI increasingly participates in science communication, it is unclear how people evaluate AI as a source of scientific information. This study examines how message framing and identity cues shape public evaluations of communicative AI and whether these effects differ when AI is encountered through reading or direct interaction. Two preregistered online experiments in Germany contrasted science communication *about* AI (reading a news-style article) with science communication *with* AI (interacting with a chatbot), manipulating risk versus progress framing and human-like versus machine-like cues. In an article-based context (Experiment 1, $N$ = 862), progress framing increased trust in AI, while machine-like wording further improved trust. In an interactive context (Experiment 2, $N$ = 868), framing shaped evaluations indirectly by reducing fear, while human-like cues increased social presence and parasocial connection, producing indirect gains across key outcomes. Across both experiments, higher AI competence was associated with more positive evaluations. Overall, the findings show that framing and design cues exert modest but systematic effects that depend on the communicative format.

## Keywords

artificial intelligence; framing; identity cues; machine heuristic; science communication; social presence; trust

## 1. Introduction

Communicative AI is now part of the everyday ecology of science communication. Chatbots, voice assistants, and generative systems are used to explain health, environment, and technology to non-expert publics (science communication with AI), while AI itself has become a regular subject of media reporting (science communication about AI; Biyela et al., 2024; Kessler et al., 2025). Despite this rapid uptake, we still know little about how audiences evaluate AI as a science communicator, whether they judge it trustworthy and credible, consider it useful, intend to use it, and whether they develop parasocial attachment (Greussing et al., 2025). These judgements matter, because public reliance on AI to deliver scientific information may alter how people relate to science, experts, and media institutions (Schäfer, 2023; Schäfer et al., 2024; Silva Luna et al., 2025).

We approach these evaluations as a communicative problem shaped both by what is said and by how it is signalled. From a content perspective, framing theory holds that the way an issue is presented structures interpretation and shapes which responses appear appropriate (Entman, 1993). In reporting on AI, progress frames—corresponding to benefit or gain framing in the broader communication literature—emphasise innovation, efficiency, and accessibility, whereas risk frames foreground potential harms such as misinformation, privacy loss, or erosion of human control (Brause et al., 2023, 2024; Roe & Perkins, 2023). These framings are not neutral descriptions. They direct attention, guide emotion, and influence whether people approach AI with openness or caution (Berendt, 2019). Studies confirm that benefit frames foster acceptance and trust, while risk frames increase scepticism (Bingaman et al., 2021; Ho & Cheung, 2024; Pataranutaporn et al., 2023), making framing a primary lever for influencing how AI is judged as a source of scientific information.

From a design perspective, evaluations are shaped by identity cues and interface features that position the AI communicator. The "Computers Are Social Actors" (CASA) paradigm shows that people apply social rules to computers once they display even minimal human-like signals, often without conscious awareness (Nass & Moon, 2000). Sundar's MAIN model (Modality, Agency, Interactivity, Navigability) extends this logic by highlighting how interface affordances—such as modality of communication or signals of agency—cue systematic judgements about credibility and engagement (Sundar, 2008). Two families of cues are particularly relevant. Human-like anthropomorphic cues, such as a name, avatar, or warm conversational tone, increase perceptions of social presence, the sense of interacting with a responsive other, which fosters parasocial interaction and perceived usefulness (Gambino et al., 2020; Nowak & Biocca, 2003). Machine-like cues, such as technical naming or unemotional precision, highlight computational identity and position AI as systematic, impartial, and accurate, which can strengthen credibility where objectivity is valued (Sundar & Kim, 2019; Yang & Sundar, 2024). In short, these design cues are not merely decorative. They guide audiences toward evaluating AI either through a relational lens or through expectations of mechanistic information delivery.

While framing and cue effects have often been studied separately, little research has examined their interaction (e.g., Gerend & Sias, 2009; J. Wang & Peng, 2023). In the context of media communication about AI, however, audiences encounter them together, as media messages are almost always embedded in a design context, where textual and visual cues signal how the agency and character of the systems should be interpreted (Brewer et al., 2025; Bunz & Braghieri, 2022). Understanding their joint effect is therefore essential for capturing how people actually encounter AI in science communication and form judgements about it.

The influence of frames and cues is often indirect, carried by discrete emotions that orient attention and evaluation (Nabi, 2007) and by heuristic shortcuts that streamline judgment under uncertainty (Sundar, 2008). In the case of AI, risk frames are linked to heightened fear and threat perception, whereas progress frames evoke hope and openness to innovation (Bilandzic et al., 2020). These affective states can, in turn, shape outcomes such as trust, credibility, and behavioural intention (Nabi et al., 2018). In parallel, human-like cues can trigger social presence—the impression of interacting with a responsive other—which fosters engagement and parasocial interaction (Gambino et al., 2020; Kim et al., 2013; Toader et al., 2020; Tsai et al., 2021). Similarly, machine-like cues can activate the machine heuristic, the expectation that machines are systematic and accurate, which bolsters credibility and trust in the information provided (Sundar & Kim, 2019; Yang & Sundar, 2024). Emotions and heuristics thus represent parallel routes through which communicative features shape evaluations of AI as a science communicator.

Audience characteristics provide an additional layer of complexity. Evaluations of AI are not made in a vacuum but are filtered through prior knowledge, experiences, and dispositions (J. D. Lee & See, 2004). Research on human–machine interaction shows that trust and prior experience shape approval of AI technologies and the heuristics applied in their evaluation (Hoff & Bashir, 2015; Molina & Sundar, 2024). Studies of AI literacy and public engagement further suggest that individuals who are more familiar with AI, who use it frequently, or who hold positive attitudes towards it could be better equipped to assess new applications critically and less reliant on surface cues (Bewersdorff et al., 2025; Gedik et al., 2025; Greussing et al., 2025). To integrate these dimensions, recent work proposes the concept of AI competence, defined as the ability to identify, use, and evaluate AI in line with ethical standards (B. Wang et al., 2023). We adopt this construct to capture people's accumulated experiences and dispositions towards AI, operationalising competence as a composite of attitudes, literacy, and usage. This integrated construct captures baseline openness to AI, reflecting a general orientation rather than any single component, though aggregation may obscure finer distinctions (Hoff & Bashir, 2015; J. D. Lee & See, 2004).

Beyond individual differences, evaluations of communicative AI are also shaped by the *format* of the encounter. Encounters with AI in science communication range from reading a news article about AI (science communication about AI) to directly conversing with a chatbot (science communication with AI; Kessler et al., 2025; Schäfer, 2023). Here, we use the term *format* broadly to refer not only to the medium of communication (e.g., article versus chatbot), but also to the degree of interactivity and attributed agency implied by the encounter. These formats differ in their communicative features and evaluative expectations: One-way, text-based news exposure tends to privilege mechanistic appraisals of accuracy and reliability (Chen et al., 2024; Grabe et al., 2000), whereas interactive exchanges foreground relational judgements of responsiveness and warmth (Go & Sundar, 2019; Tsai et al., 2021). As a result, the same framing or identity cues may carry different psychological weight depending on whether AI is encountered as a topic of reporting or as an interaction partner (Reeves & Nass, 1996). Our study explicitly compares these settings, recognising communicative format as a boundary condition for how audiences interpret and evaluate AI as a science communicator.

Our outcomes reflect these different layers of evaluation. Trust indexes a willingness to rely on the system and to feel safe doing so (Jian et al., 2000). Credibility refers to perceived accuracy and information quality (Appelman & Sundar, 2016). Perceived usefulness and intention to use capture the system's expected instrumental value—whether it helps people learn and perform better—and the behavioural willingness to adopt and recommend such systems (Chin et al., 2008). In interactive contexts, we further consider parasocial interaction: the one-sided sense of relationship with a media agent characterised by things like perceived responsiveness, warmth, and familiarity (Jin, 2010). Parasocial outcomes are particularly salient for AI because they signal both heightened engagement and the possibility of relational attachment (Maeda & Quan-Haase, 2024). Together, these outcomes capture epistemic, relational, and behavioural dimensions of how communicative AI is evaluated.

This study advances research on communicative AI in three respects. First, it specifies *processes* by linking framing theory with CASA/MAIN, identifying emotional and heuristic pathways through which communicative features influence evaluation. Second, it situates these processes in *different communicative contexts* by testing frame–cue interactions, reflecting how content and design features co-occur in real-world communication. Third, it addresses *boundaries* by examining how outcomes vary with audience competence and communicative format, providing a fuller account of when and for whom these features matter.

We preregistered the following hypotheses:

> H1 (framing): Progress (vs. risk) frames will increase trust, credibility, usefulness, intention to use, and—in interactive settings—parasocial interaction.
>
> H2 (identity cues): Human-like cues will increase usefulness, intention to use, and (in Experiment 2) parasocial interaction, while machine-like cues will increase trust and credibility.
>
> H3 (competence): AI competence will positively predict all outcomes and moderate the effects of framing and identity cues.
>
> H4 (interaction): Cue benefits will be strongest under progress framing and weaker under risk framing. Specifically, human-like cues will most strongly increase usefulness, intention to use, and (in Experiment 2) parasocial interaction under progress; machine-like cues will most strongly increase trust and credibility under progress.
>
> H5 (pathways): Framing effects will be mediated by fear and hope, and cue effects by social presence and the machine heuristic, consistent with a parallel mediation model.

Across two experiments—one article-based and one interactive—we test these hypotheses using outcomes central to evaluating AI as a science communicator. In Experiment 1, participants read a framed news-style article describing an AI system with embedded identity cues. In Experiment 2, participants read a framed article and then interacted with a chatbot designed with human-like or machine-like cues. Both studies measured discrete emotions, heuristic activation, and evaluations of trust, credibility, usefulness, and intention to use, with parasocial connection included in the interactive setting only.

## 2. Methods

### 2.1. Design and Preregistration

Two preregistered online experiments examined how media framing and AI identity cues shape public responses to communicative AI in science communication. Both studies employed between-subjects designs with random assignment implemented in SoSci Survey. Experiment 1 used a 2 (framing: risk vs. progress) × 3 (identity cues: human-like, machine-like, no cue) design. Experiment 2 used a 2 (framing: risk vs. progress) × 2 (identity cues: human-like, machine-like) design. A no-cue condition was omitted in Experiment 2 because it cannot be meaningfully operationalised in interactive settings: Even minimal chatbots necessarily convey cues through their interface (e.g., name, avatar, conversational style). Retaining only human-like and machine-like conditions therefore allowed us to contrast the two theoretically specified cue families in a realistic interaction context. The preregistrations specified hypotheses, exclusion criteria, sample size planning, and analytic strategy prior to data collection. Preregistration documents, analysis code, and study data are available via the Open Science Framework (OSF; https://osf.io/cwqjt).

The protocol was approved by the Ethics Committee of the University of Augsburg and cleared by this university's Data Protection Office. All procedures complied with GDPR. Participants provided informed consent electronically before participation and were debriefed at the end. Data were stored on secure university servers.

### 2.2. Participants and Sampling

Participants were recruited via Bilendi, a German-language online access panel. Eligibility criteria were: age $\geq$ 18, residence in Germany, sufficient German proficiency, and no prior participation. Quotas were set for gender (male/female), age (18–35, 36–59, 60+), and education (with/without Abitur), yielding 12 quota cells per experiment. Quotas were largely achieved (see Table 1). A randomisation error in SoSci Survey produced modest oversampling in specific cells (Experiment 1: men aged 36–59 without Abitur; Experiment 2: women aged 36–59 with Abitur and women aged 60+ with Abitur). We retained these additional cases because they met preregistered quality criteria, thereby increasing statistical precision without compromising the design.

Target sample size was $N$ = 800 per study, determined through an a priori power analysis in G*Power for multiple regression with nine predictors ($f^2$ = .02, $\alpha$ = .05, power = .80; minimum $N$ = 791). This provided adequate power for main and interaction effects (H1–H4), though mediation models (H5) were likely underpowered for small indirect effects.

Recruitment occurred sequentially in August 2025. In Experiment 1, 909 participants completed the survey; after excluding 47 for straightlining—defined as giving the same response across five or more completed item blocks—862 remained (94.8%). In Experiment 2, 1,048 participants completed the survey; after excluding 180 for non-engagement with the chatbot (i.e., no message sent to the chatbot) or straightlining, 868 remained (82.8%). Random assignment was implemented automatically by SoSci Survey, and participants were blind to both the study hypotheses and the existence of experimental conditions. A $\chi^2$ goodness-of-fit test indicated no deviation from equal allocation across framing × cue cells (Experiment 1: $\chi^2(5) = 1.64$, $p = .90$; Experiment 2: $\chi^2(3) = 5.99$, $p = .11$). The original preregistration specified exclusion for straightlining on more than one block, which would have excluded 73.5% (Experiment 1) and 75.7% (Experiment 2) of respondents, severely compromising power. We revised the criterion to five or more blocks after Experiment 1 recruitment but before analysing any data from either experiment, and applied it consistently to both.

Two pre-tests ($N \approx 50$) with university students assessed clarity, realism, and functionality of the materials prior to data collection. Feedback was used to refine translations, improve clarity of instructions, adjust emotional tone, and ensure chatbot usability.

**Table 1.** Sample by quotas.

| Panel A. Experiment 1. | | | | | |
|---|---|---|---|---|---|
| Age group | Male/high education | Male/low education | Female/high education | Female/low education | Total |
| 18–35 | 67 (7.8%) | 72 (8.4%) | 73 (8.5%) | 68 (7.9%) | 280 (32.5%) |
| 36–59 | 66 (7.7%) | 94 (10.9%) | 69 (8.0%) | 70 (8.1%) | 299 (34.7%) |
| 60+ | 72 (8.4%) | 68 (7.9%) | 67 (7.8%) | 76 (8.8%) | 283 (32.8%) |
| Total | 205 (23.8%) | 234 (27.1%) | 209 (24.2%) | 214 (24.8%) | 862 (100.0%) |
| Panel B. Experiment 2. | | | | | |
| Age group | Male/high education | Male/low education | Female/high education | Female/low education | Total |
| 18–35 | 85 (9.8%) | 61 (7.0%) | 62 (7.1%) | 58 (6.7%) | 266 (30.6%) |
| 36–59 | 63 (7.3%) | 57 (6.6%) | 104 (12.0%) | 65 (7.5%) | 289 (33.3%) |
| 60+ | 63 (7.3%) | 53 (6.1%) | 136 (15.7%) | 61 (7.0%) | 313 (36.1%) |
| Total | 211 (24.3%) | 171 (19.7%) | 302 (34.8%) | 184 (21.2%) | 868 (100.0%) |

### *2.3. Procedure and Stimuli*

After providing consent and demographic information, participants completed baseline measures of AI competence (attitudes, literacy, usage) and science interest, after which they were randomly assigned to conditions. At least one instructed-response attention check was embedded in the survey.

In Experiment 1, participants read a short science news article adapted from *Medical Tribune* (Söchtig, 2023), reporting on a study in which an AI chatbot provided higher-quality answers to patient questions than physicians. The article implemented a framing manipulation, distinguishing between progress and risk portrayals of AI. The progress frame emphasised benefits, societal utility, and AI's supportive role, whereas the risk frame highlighted uncertainty, harm, and accountability concerns. We use the term *progress frame* to align with narrative approaches in science communication (Bilandzic et al., 2020). Analytically, the manipulation captures the same positive–negative valence contrast that underlies benefit–risk framing. Identity cues were embedded in the description of the chatbot and operationalised through wording: human-like (empathy, warmth), machine-like (precision, objectivity), or no cue (control). Afterwards, participants completed the dependent measures in fixed order: discrete emotions (fear, hope, frustration, fascination), heuristics (social presence, machine heuristic), and outcomes (trust, credibility, usefulness, intention). Fear and hope served as confirmatory mediators; frustration and fascination were preregistered as exploratory. Mean completion time was 9 minutes 4 seconds.

In Experiment 2, participants first read a news article adapted from the German newspaper *taz* (Ronzheimer, 2023) about a Bundestag report on AI and labour shortages, framed again as progress or risk. They then completed the discrete emotion measures (fear, hope, frustration, fascination). Measuring emotions at this point, after the framed article but before the chatbot interaction, allowed us to capture the affective response to framing and to model it as an antecedent of subsequent chatbot evaluations. After this, an instruction introduced participants to a new topic: brain–computer interfaces. After a short primer, they were asked to interact with a chatbot to learn more about brain–computer interfaces; to sustain engagement, participants were told they would later answer a comprehension question. Chatbot identity cues were manipulated multimodally. In the human-like condition, the chatbot was named Kai, displayed a human avatar, and used a warm conversational style; in the machine-like condition, it was called InfoBot, displayed a chip icon, and used concise, factual language. Following the interaction, participants completed the heuristic and outcome measures as in Experiment 1, plus parasocial interaction. Mean completion time was 14 minutes 25 seconds.

All experimental stimuli used in both studies are provided in the Supplementary File.

In both experiments, participants were debriefed, provided with researcher contact information and further reading, and compensated through Bilendi.

### 2.4. Measures

All multi-item constructs were adapted from validated or widely used scales, translated into German and tailored to the context of AI as a science communicator. Translations were produced by bilingual researchers and pre-tested for clarity. Most items were measured on 7-point Likert-type scales (1 = *strongly disagree*, 7 = *strongly agree*) and averaged after reverse-coding where necessary. The number of items per scale and Cronbach's $\alpha$ reliabilities are reported in Table 2, with $\alpha \geq .70$ considered acceptable. The measured constructs were grouped as follows:

- Primary outcomes: Trust (Jian et al., 2000), credibility (Appelman & Sundar, 2016), perceived usefulness and intention to use (Chin et al., 2008), and parasocial interaction (Experiment 2 only; Jin, 2010).
- Mediators: Emotions (fear, hope, frustration, fascination; Harmon-Jones et al., 2016) and heuristic processing (social presence [K. M. Lee & Nass, 2005]; machine heuristic [Yang & Sundar, 2024]).
- Moderators/covariates: AI attitudes (Artificial Intelligence Attitudes Scale [AIAS-4]; Grassini, 2023), AI literacy (Artificial Intelligence Literacy Scale [AILS]; B. Wang et al., 2023), and AI usage (Greussing et al., 2025). Demographics (age, gender, education) also informed quotas.

All scale items are provided in the Supplementary File.

An AI Competence Index was computed as the mean of attitudes, literacy, and usage (mean-centred for interaction analyses). This preregistered composite reflects the idea that competence integrates dispositions, knowledge, and experience in everyday AI encounters (B. Wang et al., 2023). Analytically, the index is intended to model general baseline differences in how positively or negatively AI is evaluated, rather than to isolate the distinct causal roles of attitudes, literacy, or usage. Reliability was acceptable in both experiments (Experiment 1: $\alpha = .765$, $\omega_t = .783$; Experiment 2: $\alpha = .769$, $\omega_t = .783$), with moderate intercorrelations among components (Experiment 1: *rs* = .47–.62; Experiment 2: *rs* = .53–.58). Raw-mean and *z*-scored indices were near-identical ($r = .997$), and disaggregated models showed no problematic multicollinearity (max VIF $\leq$ 1.84).

**Table 2.** Measures and reliability (standardised Cronbach's $\alpha$).

| Scale | $k$ | Experiment 1 std $\alpha$ | Experiment 2 std $\alpha$ |
|---|---|---|---|
| AI attitudes | 4 | 0.852 | 0.861 |
| AI literacy | 12 | 0.870 | 0.884 |
| AI usage | 9 | 0.828 | 0.856 |
| fear | 3 | 0.892 | 0.886 |
| hope | 3 | 0.901 | 0.911 |
| social presence | 5 | 0.862 | 0.889 |
| machine heuristic | 4 | 0.852 | 0.867 |
| credibility | 3 | 0.930 | 0.941 |
| trust | 6 | 0.896 | 0.881 |
| intention | 4 | 0.957 | 0.970 |
| usefulness | 6 | 0.946 | 0.954 |
| parasocial | 4 | | 0.840 |

Notes: $k$ denotes the number of items included in each scale; frustration and fascination were measured as exploratory emotions; reliabilities and descriptive statistics are reported in the OSF supplementary materials.

### 2.5. Analysis Plan

Confirmatory analyses were conducted separately for each experiment in line with the preregistered analytic strategy. For hypotheses H1–H4, we estimated one general linear model (GLM) per outcome, with predictors for framing, AI identity cue, the mean-centred AI Competence Index, and all two-way interactions among these terms. Three-way interactions were not specified. Categorical predictors were dummy-coded, with risk framing as the reference category and, respectively, the no-cue condition in Experiment 1 and the machine-like condition in Experiment 2, as cue references.

Analyses were conducted across predefined families of outcomes to account for multiple testing (Experiment 1: trust, credibility, perceived usefulness, intention to use; Experiment 2: the same outcomes plus parasocial interaction). Unadjusted $p$-values are reported throughout, and Bonferroni-adjusted values are indicated where they alter substantive interpretation. This family-wise approach reflects the theoretical assumption that the outcomes capture related but non-identical dimensions of how communicative AI is evaluated, spanning epistemic (trust, credibility), instrumental (usefulness, intention), and—where applicable—relational (parasocial interaction) judgements. Hypotheses were evaluated at the level of outcome patterns rather than individual coefficients, and interpreted as partially supported when effects emerged for some but not all outcomes within a family. This logic was specified prior to data collection and guided both model estimation and inference.

Model assumptions (linearity, homoscedasticity, normality of residuals, and multicollinearity) were assessed for all GLMs. To reduce sensitivity to heteroscedasticity, HC3 (heteroscedasticity-consistent estimator, type 3) robust standard errors were used throughout. Full statistical output and diagnostics—including Type-II ANOVA statistics, estimated marginal means, Bonferroni-corrected pairwise comparisons, and model diagnostic information—are available via OSF (https://osf.io/cwqjt).

To test H5, we estimated separate parallel mediation models for each outcome using *lavaan* (R version 4.5.0; package version 0.6). Indirect effects were assessed using 5,000 bias-corrected bootstrap samples with 95% confidence intervals. Two sets of mediation pathways were specified: (a) framing → fear and hope → outcomes, and (b) AI identity cues → social presence and the machine heuristic → outcomes, with parasocial interaction included as an outcome in Experiment 2. All tests were two-tailed with $\alpha = .05$. Mediation models were powered to detect medium indirect effects and may be underpowered for small effects; null mediation results are therefore interpreted with caution.

Cases excluded under preregistered quality criteria were removed prior to analysis, and the resulting datasets contained no missing values. Mediation outputs, robustness checks, and sensitivity analyses are available via OSF (https://osf.io/cwqjt). Sensitivity analyses further indicated that the reported results were stable across diagnostic variants. Preregistered exploratory analyses, including moderation by science interest and effects of frustration and fascination, are reported via OSF (https://osf.io/cwqjt).

## 3. Results

### *3.1. Experiment 1*

The article-based experiment (Table 3) revealed modest framing effects overall. Progress framing increased trust compared with risk ($B = 0.43$, 95% CI [0.17, 0.69], $p = .001$), an effect robust to Bonferroni correction. Gains for usefulness ($B = 0.35$, 95% CI [0.07, 0.62], $p = .015$) and intention ($B = 0.34$, 95% CI [0.03, 0.64], $p = .030$) were positive but did not survive correction, while credibility was unaffected ($p = .058$). Taken together, these results provide partial support for H1, concentrated on trust.

Design cues exerted a stronger and more consistent influence than framing. Relative to the no-cue control, machine-like cues significantly enhanced trust ($B = 0.48$, 95% CI [0.23, 0.74], $p < .001$), usefulness ($B = 0.45$, 95% CI [0.18, 0.73], $p = .001$), and intention ($B = 0.41$, 95% CI [0.11, 0.71], $p = .008$), with a smaller improvement in credibility ($B = 0.28$, 95% CI [0.00, 0.56], $p = .048$) that did not survive Bonferroni. Human-like cues produced weaker and less consistent benefits: modest gains for trust ($B = 0.28$, 95% CI [0.03, 0.54], $p = .028$) and intention ($B = 0.31$, 95% CI [0.01, 0.61], $p = .043$) that did not survive correction, while credibility and usefulness were unaffected. These patterns support H2 more clearly for machine-like cues than for human-like cues.

AI competence did not moderate the effects of framing or cues; all interactions with competence were null, apart from one small effect for credibility ($B = -0.22$, 95% CI [−0.41, −0.04], $p = .020$), which was not robust. By contrast, competence showed large, positive main effects across outcomes: trust ($B = 0.76$, 95% CI [0.62, 0.89], $p < .001$); credibility ($B = 0.66$, 95% CI [0.51, 0.80], $p < .001$); usefulness ($B = 0.60$, 95% CI [0.46, 0.75], $p < .001$); and intention ($B = 1.01$, 95% CI [0.85, 1.17], $p < .001$). Participants who were higher in competence were consistently more positive about AI in science communication. Thus, H3 was supported in terms of baseline elevation, but not in terms of buffering sensitivity to frames or cues.

**Table 3.** GLM coefficients (OLS): Experiment 1.

| Section | Predictor | trust | | | credibility | | | usefulness | | | intention | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $B$ | 95% CI | $p$ | $B$ | 95% CI | $p$ | $B$ | 95% CI | $p$ | $B$ | 95% CI | $p$ |
| Main effects | Framing: progress (vs risk) | 0.43** | [0.17, 0.69] | 0.001 | 0.27 | [−0.01, 0.56] | 0.058 | 0.35* | [0.07, 0.62] | 0.015 | 0.34* | [0.03, 0.64] | 0.030 |
| | Cue: human (vs none) | 0.28* | [0.03, 0.54] | 0.028 | 0.16 | [−0.12, 0.44] | 0.254 | 0.20 | [−0.07, 0.48] | 0.150 | 0.31* | [0.01, 0.61] | 0.043 |
| | Cue: machine (vs none) | 0.48*** | [0.23, 0.74] | < .001 | 0.28* | [0.00, 0.56] | 0.048 | 0.45** | [0.18, 0.73] | 0.001 | 0.41** | [0.11, 0.71] | 0.008 |
| | AI competence (centred) | 0.76*** | [0.62, 0.89] | < .001 | 0.66*** | [0.51, 0.80] | < .001 | 0.60*** | [0.46, 0.75] | < .001 | 1.01*** | [0.85, 1.17] | < .001 |
| Interactions | Cue: human × AI competence (centred) | −0.04 | [−0.21, 0.13] | 0.649 | −0.22* | [−0.41, −0.04] | 0.020 | −0.09 | [−0.28, 0.09] | 0.330 | 0.10 | [−0.10, 0.30] | 0.337 |
| | Cue: machine × AI competence (centred) | −0.05 | [−0.22, 0.13] | 0.603 | −0.11 | [−0.30, 0.09] | 0.285 | −0.06 | [−0.25, 0.13] | 0.568 | −0.01 | [−0.22, 0.20] | 0.926 |
| | Framing: progress × AI competence (centred) | −0.00 | [−0.15, 0.14] | 0.955 | 0.00 | [−0.15, 0.16] | 0.953 | −0.02 | [−0.18, 0.13] | 0.755 | 0.06 | [−0.11, 0.23] | 0.475 |
| | Framing: progress × Cue: human | −0.40* | [−0.76, −0.05] | 0.027 | −0.24 | [−0.63, 0.15] | 0.232 | −0.36 | [−0.74, 0.03] | 0.070 | −0.32 | [−0.75, 0.10] | 0.134 |
| | Framing: progress × Cue: machine | −0.48** | [−0.85, −0.12] | 0.009 | −0.22 | [−0.62, 0.18] | 0.271 | −0.61** | [−1.00, −0.22] | 0.002 | −0.53* | [−0.96, −0.10] | 0.015 |

Notes: Entries are unstandardised coefficients ($B$), 95% confidence intervals, and $p$ values; reference categories: framing = risk, cue = none; AI competence is mean-centred; included terms in all models: framing, cue, AI competence, and all two-way interactions; significance flags: * $p < .05$, ** $p < .01$, *** $p < .001$; model fit—trust: $R^2 = 0.328$, adj. $R^2 = 0.320$, RMSE = 1.092 | credibility: $R^2 = 0.190$, adj. $R^2 = 0.181$, RMSE = 1.201 | usefulness: $R^2 = 0.195$, adj. $R^2 = 0.187$, RMSE = 1.178 | intention: $R^2 = 0.425$, adj. $R^2 = 0.419$, RMSE = 1.294.

Finally, H4 was not supported. Frame × cue interactions indicated that cue benefits were weaker under progress rather than stronger. Machine-like cues lost effectiveness under progress, reducing gains in trust ($B$ = −0.48, 95% CI [−0.85, −0.12], $p$ = .01, Bonferroni $p$ = .04) and usefulness ($B$ = −0.61, 95% CI [−1.00, −0.22], $p$ < .001, Bonferroni $p$ = .01). A similar trend appeared for intention ($B$ = −0.53, 95% CI [−0.96, −0.10], $p$ = .02), though this did not survive correction (Bonferroni $p$ = .06). Human-like cues also showed a smaller trust benefit under progress ($B$ = −0.40, 95% CI [−0.76, −0.05], $p$ = .03), but this too did not remain significant after correction (Bonferroni $p$ = .11). These patterns suggest that cues mattered most under risk, where mechanistic or human-like signals may have provided reassurance; under progress, cues appeared redundant.

In mediation, framing effects were not carried by emotion. Neither fear nor hope accounted for the progress advantage on trust, as all indirect confidence intervals included zero, pointing to a direct, non-affective route. By contrast, heuristic pathways explained cue effects in line with preregistered expectations. Machine-like phrasing increased activation of the machine heuristic, which in turn raised trust ($B$ = 0.14, 95% CI [0.04, 0.24], $p$ = .006); usefulness ($B$ = 0.16, 95% CI [0.05, 0.27], $p$ = .006); intention ($B$ = 0.13, 95% CI [0.04, 0.23], $p$ = .009); and credibility ($B$ = 0.22, 95% CI [0.07, 0.38], $p$ = .006). Human-like phrasing instead increased social presence, which mediated gains in trust ($B$ = 0.11, 95% CI [0.04, 0.20], $p$ = .005); intention ($B$ = 0.17, 95% CI [0.06, 0.29], $p$ = .005); credibility ($B$ = 0.07, 95% CI [0.02, 0.12], $p$ = .006); and usefulness ($B$ = 0.07, 95% CI [0.02, 0.13], $p$ = .012). These specific heuristic routes were the only consistent mediators; cross-paths (e.g., human-like cues via the machine heuristic) were non-significant. Total indirect effects confirmed that both cue families had reliable mediated effects on key outcomes (e.g., human-like via social presence on trust, $B$ = 0.17, 95% CI [0.01, 0.32]; machine-like via the machine heuristic on credibility, $B$ = 0.22, 95% CI [0.03, 0.41]). Some indirect gains were offset by small negative direct paths, which helps explain why cue effects appeared weaker in the GLMs. Overall, H5 was partially supported: Cue effects were transmitted through their respective heuristic mechanisms, but framing effects were not mediated by fear or hope.

### 3.2. Experiment 2

In the interactive setting (Table 4) where participants wrote to a chatbot, framing effects were negligible. Progress relative to risk produced no reliable changes in trust ($B$ = 0.11, 95% CI [−0.11, 0.32], $p$ = .332); credibility ($B$ = 0.15, 95% CI [−0.11, 0.41], $p$ = .245); usefulness ($B$ = 0.15, 95% CI [−0.08, 0.38], $p$ = .208); intention ($B$ = 0.14, 95% CI [−0.15, 0.44], $p$ = .338); or parasocial interaction ($B$ = 0.04, 95% CI [−0.15, 0.23], $p$ = .666). These results indicate that H1 was not supported: Framing did not directly alter how participants evaluated the chatbot.

By contrast, cue effects yielded one robust result for H2. Human-like design—operationalised through a name, avatar, and warm conversational style—substantially increased parasocial interaction compared to machine-like design ($B$ = 0.54, 95% CI [0.34, 0.74], $p$ < .001), a medium-to-large effect that remained robust after correction. For trust, credibility, usefulness, and intention, however, differences between human-like and machine-like cues were negligible (all $p$s ≥ .18). In other words, H2 was supported only for parasocial interaction.

AI competence again emerged as the most consistent predictor of evaluation. Higher competence scores predicted more positive responses across all outcomes: trust ($B = 0.48$, 95% CI [0.35, 0.61], $p < .001$); credibility ($B = 0.27$, 95% CI [0.11, 0.43], $p < .001$); usefulness ($B = 0.45$, 95% CI [0.31, 0.59], $p < .001$); intention ($B = 0.80$, 95% CI [0.62, 0.98], $p < .001$); and parasocial interaction ($B = 0.34$, 95% CI [0.23, 0.46], $p < .001$). Yet competence did not moderate sensitivity to frames or cues: All interaction terms were non-significant after correction. Thus, H3 was only partially supported: Higher competence consistently elevated baseline evaluations, but it did not moderate the effects of framing or cues.

Finally, no evidence emerged for framing × cue interactions. Neither human-like nor machine-like design varied in effect depending on whether the chatbot was introduced with a progress or risk frame (all $ps \geq .33$). In this context, H4 was not supported.

In mediation, two pathways were evident. First, framing influenced outcomes indirectly through fear. Progress framing reduced fear, which in turn improved trust (indirect $B = 0.054$, 95% CI [0.017, 0.098], $p = .009$); credibility ($B = 0.044$, 95% CI [0.013, 0.089], $p = .018$); usefulness ($B = 0.063$, 95% CI [0.019, 0.115], $p = .010$); and intention ($B = 0.059$, 95% CI [0.018, 0.112], $p = .013$). Equivalently, risk framing heightened fear, lowering evaluations across these outcomes. These effects did not extend to parasocial interaction. Hope played no mediating role. Second, cues exerted their influence through social presence. Human-like design increased perceived presence, which carried sizeable indirect gains for trust ($B = 0.261$, 95% CI [0.180, 0.357], $p < .001$); credibility ($B = 0.137$, 95% CI [0.085, 0.201], $p < .001$); usefulness ($B = 0.255$, 95% CI [0.181, 0.351], $p < .001$); intention ($B = 0.443$, 95% CI [0.320, 0.592], $p < .001$); and parasocial interaction ($B = 0.494$, 95% CI [0.381, 0.608], $p < .001$). By contrast, the machine heuristic was inert. Suppression effects were evident for trust, usefulness, and intention: Strong positive indirect pathways via presence were partly cancelled by small negative direct coefficients, leaving total cue effects less pronounced in the GLMs. For parasocial interaction, both indirect and total effects were large and positive, with no suppression. Overall, H5 was supported for the fear pathway (framing) and the social-presence pathway (cues), but not for hope or the machine heuristic.

Three regularities stand out across studies. First, AI competence consistently elevated evaluations across all outcomes but did not alter sensitivity to frames or cues, offering only partial support for H3. Second, the mechanisms through which cues operated shifted with format. In article reading (Experiment 1), machine-like presentation activated a mechanistic route via the machine heuristic, boosting trust, usefulness, and intention, while human-like cues worked more modestly through social presence. In chatbot interaction (Experiment 2), by contrast, human-like presentation activated a relational route via social presence, producing strong parasocial bonding and broad indirect gains, whereas the machine heuristic remained inert. Third, framing effects were weak overall but differed by context: In articles, progress framing directly boosted trust; in chatbot interactions, framing shaped outcomes only indirectly by reducing fear. Hope did not mediate effects in either study. Interactions between framing and cues were counter-theoretical in Experiment 1, where progress dampened cue benefits, and were absent in Experiment 2. Together, these findings indicate that communicative format determines whether audiences evaluate AI through mechanistic or relational pathways, while framing operates chiefly as an affective modifier that reduces fear in interactive use.

**Table 4.** GLM coefficients (OLS): Experiment 2.

| | | trust | | | credibility | | | usefulness | | | intention | | | parasocial | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Section | Predictor | *B* | 95% CI | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* |
| Main effects | Framing: progress (vs risk) | 0.11 | [−0.11, 0.32] | 0.332 | 0.15 | [−0.11, 0.41] | 0.245 | 0.15 | [−0.08, 0.38] | 0.208 | 0.14 | [−0.15, 0.44] | 0.338 | 0.04 | [−0.15, 0.23] | 0.666 |
| | Cue: human (vs machine) | −0.15 | [−0.37, 0.07] | 0.190 | 0.07 | [−0.19, 0.34] | 0.589 | −0.06 | [−0.30, 0.17] | 0.609 | −0.06 | [−0.36, 0.24] | 0.698 | 0.54*** | [0.34, 0.74] | < .001 |
| | AI competence | 0.48*** | [0.35, 0.61] | < .001 | 0.27*** | [0.11, 0.43] | < .001 | 0.45*** | [0.31, 0.59] | < .001 | 0.80*** | [0.62, 0.98] | < .001 | 0.34*** | [0.23, 0.46] | < .001 |
| Interactions | Cue: human × AI competence | 0.03 | [−0.11, 0.18] | 0.667 | 0.18* | [0.01, 0.36] | 0.042 | 0.14 | [−0.02, 0.29] | 0.084 | 0.02 | [−0.18, 0.22] | 0.839 | −0.02 | [−0.15, 0.10] | 0.718 |
| | Framing: progress × AI competence | 0.11 | [−0.03, 0.26] | 0.121 | 0.08 | [−0.09, 0.26] | 0.359 | 0.13 | [−0.03, 0.28] | 0.110 | 0.18 | [−0.02, 0.38] | 0.077 | 0.04 | [−0.09, 0.17] | 0.518 |
| | Framing: progress × Cue: human | −0.04 | [−0.34, 0.26] | 0.769 | −0.16 | [−0.53, 0.20] | 0.385 | −0.10 | [−0.42, 0.22] | 0.529 | −0.14 | [−0.56, 0.27] | 0.496 | −0.13 | [−0.40, 0.13] | 0.330 |

Notes: Entries are unstandardised coefficients (*B*), 95% confidence intervals, and *p* values; reference categories: framing = risk, cue = machine; AI competence is mean-centred; included terms in all models: framing, cue, AI competence, and all two-way interactions; significance flags: * $p < .05$, ** $p < .01$, *** $p < .001$; model fit—trust: $R^2 = 0.219$, adj. $R^2 = 0.213$, RMSE = 1.117 | credibility: $R^2 = 0.095$, adj. $R^2 = 0.088$, RMSE = 1.356 | usefulness: $R^2 = 0.211$, adj. $R^2 = 0.206$, RMSE = 1.197 | intention: $R^2 = 0.274$, adj. $R^2 = 0.269$, RMSE = 1.539 | parasocial: $R^2 = 0.162$, adj. $R^2 = 0.156$, RMSE = 0.995.

## 4. Discussion

This study examined how audiences evaluate communicative AI in science communication across two contexts. By comparing two common encounters—reading a framed news-style article (science communication about AI) and interacting with a chatbot after reading a similar article (science communication with AI)—we tested how framing, identity cues, and audience competence shape evaluations of communicative AI. The results converge on one core finding: Communicative format determines which evaluative routes people use. Articles invited a mechanistic appraisal, where machine-like descriptors and progress framing supported trust, with smaller or less robust effects on other outcomes. Chatbot exchanges invited relational appraisal, where human-like design fostered social presence and parasocial bonding, and framing played a role mainly by lowering fear before the interaction. AI competence consistently raised overall evaluations but did not insulate participants from these communicative influences. Taken together, these effects are modest but consistent, clarifying how format, framing, and design cues jointly shape public views of AI as a science communicator.

In the article context, machine-like descriptions reliably increased trust, usefulness, and intention to use, whereas human-like wording produced weaker and less consistent gains. Mediation shows why: Mechanistic phrasing activated the machine heuristic, which then boosted positive evaluations. This fits the epistemic profile of the task. The article presented AI as a source of scientific information in a health context, and the outcome measures emphasised objectivity, precision, reliability, and responsibility. The stimulus took the form of a news article. News formats already carry expectations of impartiality and factual accuracy, so mechanistic cues added an extra layer of epistemic authority by aligning both with domain norms and with journalistic conventions of neutrality (Grabe et al., 2000). Progress framing also raised trust directly, without mediation through the emotions we measured, suggesting that other pathways were at play—such as expectations about news provision or beliefs about medical reliability. Frames may have supplied ready-made standards for judgment aligned with both journalistic and domain norms.

In the interactive context, leverage shifted to a relational route. A chatbot with a name, avatar, and warm tone elicited strong social presence, which increased parasocial interaction and, indirectly, trust, credibility, usefulness, and intention to use. The machine heuristic was largely inert. This is not a failed replication but rather a shift in communicative conditions: Once people enter a conversation, responsiveness, warmth, and the feeling of being addressed become salient evaluative criteria, and those criteria are precisely what human-like design supplies (Gambino et al., 2020; Nass & Moon, 2000). Suppression effects help reconcile these patterns: Strong positive indirect effects via presence were partly offset by small negative direct paths, yielding near-zero totals outside parasocial interaction. Participants can feel connected yet answer cautiously about whether such a system is useful or safe—especially after exposure to risk-framed content.

The strength of parasocial interaction as an outcome is worth particular note. On the one hand, it represents a powerful engagement mechanism: It deepens attention, supports persuasion, and fosters a sense of responsiveness that may encourage sustained interaction with communicative AI (Matz et al., 2024; Schäfer, 2023). On the other hand, parasociality carries risks. If a system feels too human, it may blur boundaries of responsibility, create unrealistic expectations of empathy, or prompt misplaced reliance (Maeda & Quan-Haase, 2024). For science communication, this double-edged sword is consequential. Relational ease can enhance openness to scientific information, but without visible signals of non-human status, limits of expertise, and accountability structures, it may also foster over-attachment or uncritical trust (Silva Luna et al., 2025). Designers should therefore treat parasociality as a resource to be channelled carefully, combining warmth with clear epistemic boundaries.

Framing behaved differently across formats. In reading, progress raised trust directly. In interaction, direct framing effects disappeared, but progress reduced fear, which then improved trust, credibility, usefulness, and intention to use; risk increased fear and indirectly reduced these outcomes. Hope did not mediate effects in either study. This pattern aligns with ample evidence that discrete, negatively valenced affect is a more diagnostic guide under uncertainty than positive affect (Anderson et al., 2019). It also fits our timeline: Emotions were measured after the framed article and before the chat, so framing set the affective context within which the exchange unfolded (Bilandzic et al., 2020). The practical lever is modest but actionable: Where interaction follows, early fear reduction is a reliable pathway to better evaluations.

When we examine how framing combined with design cues, the interplay was small and inconsistent across settings. Contrary to our preregistration, cue benefits were not amplified under progress framing. In the article study, machine-like descriptions increased trust and usefulness most clearly under risk, with weaker or absent effects under progress; human-like cues showed a similar but smaller attenuation pattern. One plausible explanation is simple redundancy. When progress framing already conveys positive value, additional signals make little difference, whereas under risk, either mechanistic or human-like features can ease concerns and provide reassurance (J. Wang & Peng, 2023). Under risk, both cue families may function as reassurance signals that reduce uncertainty, despite typically operating through different heuristics (Anderson et al., 2019; Nabi et al., 2018). This pattern complicates simple congruence accounts by suggesting that evaluative criteria shift under conditions of perceived risk. We treat this interpretation as exploratory, given the modest interaction effects and brief exposure, but it points to a context-sensitive use of heuristics that warrants further testing.

In the chatbot study, by contrast, no frame–cue interactions emerged. Here, the conversational setting drew attention to relational qualities, and framing affected outcomes only indirectly through fear reduction. Overall, the expectation of a general "congruence bonus"—for example, stronger outcomes when progress and human-like design are paired—was not supported. Instead, cue effectiveness depended on context: Cues mattered under risk in text, but lost salience under progress or in interactive formats.

AI competence showed the same profile in both experiments. As a composite of attitudes, literacy, and usage, it was positively associated with all baseline evaluations: More competent participants trusted AI to a greater degree, found it more credible and useful, reported stronger intentions to use it, and in chat reported higher parasocial interaction (Schäfer et al., 2024). What it did not do was dampen sensitivity to frames or cues. In other words, competence was associated with differences in levels rather than slopes. This complicates the idea that individual differences, such as trust or attitudes, can eliminate reliance on heuristics altogether (Molina & Sundar, 2024). In brief encounters, even experienced users appear to rely on the same quick routes—objectivity signals in text, relational signals in chat—though they begin from a more positive position. Null interactions may also reflect methodological limits, since short and weak exposures and small sample sizes are not ideal for detecting moderating effects (R. Wang & Ware, 2013).

These interpretations need to be read with caution. The two experiments differed simultaneously in format, topic domain, source article, and factorial structure, so the patterns attributed to format could partly reflect these confounds; future work should isolate format while holding domain constant. Exposures were brief, cues deliberately simplified, and interaction restricted to a single exchange, meaning that the study captures early, impression-level evaluations rather than judgements formed through repeated or extended use. Correspondingly, several of the observed effects are modest in size, and future work should examine whether the emotional and heuristic pathways identified here persist, strengthen, or attenuate in multi-turn or longitudinal encounters with communicative AI.

As controlled online experiments, the studies traded ecological richness for causal clarity, and panel-based recruitment introduces concerns about attention, motivation, and device heterogeneity. Materials were German-language and domain-specific, which constrains generalisability in a fast-moving field where public discourse about AI evolves rapidly. The AI competence measure combined attitudes, literacy, and usage, capturing broad orientation at the expense of finer distinctions. Manipulation checks were limited, and pretesting of frame and cue salience was largely qualitative. Finally, outcomes were self-reported and measured immediately after exposure, leaving the durability of effects and their behavioural implications open questions.

Importantly, the findings extend framing work into human–AI communication by showing that frames matter less as direct levers of judgement in reading than as affective priors for interaction (Nabi et al., 2018). The findings also refine CASA and MAIN by demonstrating a context-dependent division of labour between mechanistic and relational heuristics: Textual cues in news-style formats align with objectivity and precision, while embodied cues in live chat align with presence and responsiveness. Which cues matter depends on the norms and expectations of the setting (Nass & Moon, 2000; Sundar, 2008). The findings further qualify a common assumption about literacy as a buffer. Competence predicted higher evaluations, but it does not mute responsiveness to communication features. Finally, the findings identify a practical lever that is often invoked but rarely demonstrated through mediation: Reducing fear before or at the outset of a human–AI interaction improves evaluations, whereas eliciting hope does not.

These implications speak directly to journalists, science communication practitioners, designers, and institutions responsible for AI-based science communication tools. For practice, the guidance is straightforward: Match cues to format and to the evaluative standards implied by the setting. In static explainers and news reports, highlighting objectivity, precision, and responsibility can be effective, and simple machine-like wording may help. In live chat, design should prioritise attentiveness and presence, while also surfacing accuracy, uncertainty, provenance, and liability at key points. High-competence audiences should not be assumed to ignore framing or design cues; rather, they tend to start from more positive baselines while still relying on format-specific shortcuts. In sensitive domains such as health, labour, or neurotechnology, reducing fear early through clear boundaries, privacy and accountability statements, and transparent claims about competence and limits can improve downstream evaluations, even when direct framing effects are absent.

Future work should track these processes over time, disaggregate components of competence, extend tests to additional domains, and incorporate behavioural outcomes. A key next step is to examine hybrid encounters where mechanistic and relational cues are interleaved, and to study how calibration features—such as uncertainty displays, citations, or guardrails—shape both social presence and perceived reliability.

## 5. Conclusion

This study asked how people evaluate communicative AI when it presents science, contrasting news-style reading with live chatbot interaction. The results converge on one insight: Format channels evaluation through different heuristics. Articles elicited mechanistic judgements, boosted by machine-like descriptors and progress framing. Chatbots, by contrast, elicited relational judgements, strengthened by human-like design and carried indirectly through social presence, with framing shaping outcomes only by reducing (or increasing) fear. AI competence was consistently associated with more positive overall evaluations, but did not insulate participants from these communicative influences.

The theoretical point is that evaluations of communicative AI are not governed by universal rules but by the intersection of features with the format of encounter. Mechanistic and relational heuristics divide labour depending on whether AI is read about or interacted with. The practical point is that design should meet audiences where the format already directs their attention: emphasising precision and responsibility in news, while pairing warmth and attentiveness with clear boundaries in live chat.

What follows is a broader challenge. As communicative AI becomes woven into science communication, its impact will hinge less on isolated cues than on how formats, expectations, and safeguards are aligned. Designing systems that harness these heuristics while protecting epistemic standards is where the field now needs to move.

### Acknowledgments

This article was written within the project Wiss-KKI – Wissenschaftskommunikation über und mit kommunikativer Künstlicher Intelligenz (Science Communication About and With Communicative Artificial Intelligence).

### Funding

This research was supported by a grant from the German Federal Ministry of Research, Technology and Space (Bundesministerium für Forschung, Technologie und Raumfahrt [BMFTR]; grant no. 01WK2305B) awarded to Helena Bilandzic.

### Conflict of Interests

The authors declare no conflict of interests.

### Data Availability

The data, analysis code, and additional analytical outputs supporting the findings of this study are available via OSF at https://osf.io/cwqjt

### LLMs Disclosure

ChatGPT-5 (OpenAI) was used for language editing and proofreading of the manuscript. All theoretical framing, study design, interpretation, data analysis, and substantive writing were conducted by the authors.

### Supplementary Material

Supplementary material for this article is available online in the format provided by the authors (unedited).

## References


Anderson, E. C., Carleton, R. N., Diefenbach, M., & Han, P. K. (2019). The relationship between uncertainty and affect. *Frontiers in Psychology*, *10*, Article 2504.

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79.

Berendt, B. (2019). AI for the common good?! Pitfalls, challenges, and ethics pen-testing. *Paladyn, Journal of Behavioral Robotics*, *10*(1), 44–65.

Bewersdorff, A., Hornberger, M., Nerdel, C., & Schiff, D. S. (2025). AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy. *Computers and Education: Artificial Intelligence*, *8*, Article 100340.

Bilandzic, H., Kinnebrock, S., & Klingler, M. (2020). The emotional effects of science narratives: A theoretical framework. *Media and Communication*, *8*(1), 151–163.

Bingaman, J., Brewer, P. R., Paintsil, A., & Wilson, D. C. (2021). "Siri, show me scary images of AI": Effects of text-based frames and visuals on support for artificial intelligence. *Science Communication*, *43*(3), 388–401.

Biyela, S., Dihal, K., Gero, K. I., Ippolito, D., Menczer, F., Schäfer, M. S., & Yokoyama, H. M. (2024). Generative AI and science communication in the physical sciences. *Nature Reviews Physics*, *6*(3), 162–165.

Brause, S. R., Yang, H., Schäfer, M. S., & Zeng, J. (2024). News media imaginaries of artificial intelligence in healthcare: A qualitative analysis across China, Germany, and the United States. *Emerging Media*, *2*(4), 581–610.

Brause, S. R., Zeng, J., Schäfer, M. S., & Katzenbach, C. (2023). Media representations of artificial intelligence: Surveying the field. In S. Lindgren (Ed.), *Handbook of critical studies of artificial intelligence* (pp. 277–288). Edward Elgar Publishing.

Brewer, P. R., Cuddy, L., Dawson, W., & Stise, R. (2025). Artists or art thieves? Media use, media messages, and public opinion about artificial intelligence image generators. *AI & Society*, *40*(1), 77–87.

Bunz, M., & Braghieri, M. (2022). The AI doctor will see you now: Assessing the framing of AI in news coverage. *AI & Society*, *37*(1), 9–22.

Chen, J., Song, C., Xiao, M., & Nie, H. (2024). The richer, the better? Users perception of news credibility of short video news. *Journal of Broadcasting & Electronic Media*, *68*(3), 422–444.

Chin, W. W., Johnson, N., & Schwarz, A. (2008). A fast form approach to measuring technology acceptance and other constructs. *MIS Quarterly*, *32*(4), 687–703.

Entman, R. M. (1993). Framing: Towards clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the Computers Are Social Actors paradigm. *Human-Machine Communication*, *1*, 71–85.

Gedik, N., Işıkoğlu, M. A., & Şendağ, S. (2025). Encompassing AI attitudes: The role of AI literacy and several human and technology-oriented variables. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2025.2523379

Gerend, M. A., & Sias, T. (2009). Message framing and color priming: How subtle threat cues affect persuasion. *Journal of Experimental Social Psychology*, *45*(4), 999–1002.

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.

Grabe, M. E., Zhou, S., Lang, A., & Bolls, P. D. (2000). Packaging television news: The effects of tabloid on information processing and evaluative responses. *Journal of Broadcasting & Electronic Media*, *44*(4), 581–598.

Grassini, S. (2023). Development and validation of the AI Attitude Scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology*, *14*, Article 1191628.

Greussing, E., Guenther, L., Baram-Tsabari, A., Dabran-Zivan, S., Jonas, E., Klein-Avraham, I., Taddicken, M., Agergaard, T. E., Beets, B., Brossard, D., Chakraborty, A., Fage-Butler, A., Huang, C.-J., Kankaria, S., Lo, Y.-Y., Nielsen, K. H., Riedlinger, M., & Song, H. (2025). The perception and use of generative AI for science-related information search: Insights from a cross-national study. *Public Understanding of Science*, *34*(5), 599–615.

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PLOS One*, *11*(8), Article e0159915.

Ho, S. S., & Cheung, J. C. (2024). Trust in artificial intelligence, trust in engineers, and news media: Factors shaping public perceptions of autonomous drones through UTAUT2. *Technology in Society*, *77*, Article 102533.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Jin, S. A. A. (2010). Parasocial interaction with an avatar in Second Life: A typology of the self and an empirical test of the mediating role of social presence. *Presence*, *19*(4), 331–340.

Kessler, S. H., Mahl, D., Schäfer, M. S., & Volk, S. C. (2025). All eyez on AI: A roadmap for science communication research in the age of artificial intelligence. *JCOM: Journal of Science Communication*, *24*(2), Article Y01.

Kim, K. J., Park, E., & Sundar, S. S. (2013). Caregiving role in human–robot interaction: A study of the mediating effects of perceived benefit and social presence. *Computers in Human Behavior*, *29*(4), 1799–1806.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

Lee, K. M., & Nass, C. (2005). Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology*, *7*(1), 31–45.

Maeda, T., & Quan-Haase, A. (2024). When human–AI interactions become parasocial: Agency and anthropomorphism in affective design. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1068–1077). ACM.

Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports*, *14*(1), Article 4692.

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.

Nabi, R. L. (2007). Emotion and persuasion: A social cognitive perspective. In D. R. Roskos-Ewoldsen & J. Monahan (Eds.), *Communication and social cognition: Theories and methods* (pp. 377–398). Lawrence Erlbaum Associates.

Nabi, R. L., Gustafson, A., & Jensen, R. (2018). Framing climate change: Exploring the role of emotion in generating advocacy behavior. *Science Communication*, *40*(4), 442–468.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *12*(5), 481–494.

Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence*, *5*(10), 1076–1086.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Center for the Study of Language and Information; Cambridge University Press.

Roe, J., & Perkins, M. (2023). 'What they're not telling you about ChatGPT': Exploring the discourse of AI in UK news media headlines. *Humanities and Social Sciences Communications*, *10*(1), Article 753.

Ronzheimer, M. (2023, May 5). ChatGPT und Fachkräftemangel—KI hat Bock auf Arbeit. *taz*. https://taz.de/ChatGPT-und-Fachkraeftemangel/!5930895

Schäfer, M. S. (2023). The Notorious GPT: Science communication in the age of artificial intelligence. *Journal of Science Communication*, *22*(2), Article Y02.

Schäfer, M. S., Kremer, B., Mede, N. G., & Fischer, L. (2024). Trust in science, trust in ChatGPT? How Germans think about generative AI as a source in science communication. *Journal of Science Communication*, *23*(9), Article A04.

Silva Luna, D., Broer, I., Bilandzic, H., Taddicken, M., Schuller, B. W., & Bürger, M. (2025). Quality in science communication with communicative artificial intelligence: A principle-based framework. *Public Understanding of Science*, *34*(8), 966–987.

Söchtig, M. (2023, September 6). Beim Beantworten von Patientenfragen ist KI empathischer und besser. *Medical Tribune*. https://www.medical-tribune.de/medizin-und-forschung/artikel/chatbot-schlaegt-aerzte

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 72–100). MIT Press.

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Article 538). ACM.

Toader, D.-C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D., & Rădulescu, A. T. (2020). The effect of social presence and chatbot errors on trust. *Sustainability*, *12*(1), Article 256.

Tsai, W. H. S., Liu, Y., & Chuan, C. H. (2021). How chatbots' social presence communication enhances consumer engagement: The mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing*, *15*(3), 460–482.

Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of Artificial Intelligence Literacy Scale. *Behaviour & Information Technology*, *42*(9), 1324–1337.

Wang, J., & Peng, L. (2023). Striking an emotional chord: Effects of emotional appeals and chatbot anthropomorphism on persuasive science communication. *Science Communication*, *45*(4), 485–511.

Wang, R., & Ware, J. H. (2013). Detecting moderator effects using subgroup analyses. *Prevention Science*, *14*(2), 111–120.

Yang, H., & Sundar, S. S. (2024). Machine heuristic: Concept explication and development of a measurement scale. *Journal of Computer-Mediated Communication*, *29*(6), Article zmae019.


## About the Authors

**Daniel Silva Luna** is a postdoctoral researcher in communication science at the University of Augsburg. His work focuses on science communication, especially emotions and artificial intelligence. He completed his PhD at the University of Otago and has held postdoctoral positions in Antwerp and Karlsruhe.

**Helena Bilandzic** is a professor of communication science at the University of Augsburg. She specialises in media effects and narrative persuasion, focusing on environmental and health communication. She has led multiple DFG-funded projects and publishes widely, using both qualitative and quantitative methods.

**Martin Bürger** is a researcher and doctoral candidate at the Chair for Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg. His work focuses on intelligent systems for health monitoring and their intersection with science communication.