A JOURNAL BY SISSA MEDIALAB

JCom

SPECIAL ISSUE
Science Communication in the Age of Artificial Intelligence

ARTICLE

Balancing realism and trustworthiness: AI avatars in science communication

Jasmin Baake, Josephine B. Schmitt and Julia Metag

Abstract

AI-generated avatars in science communication offer potential for conveying complex information. However, highly realistic avatars may evoke discomfort and diminish trust, a key factor in science communication. Drawing on existing research, we conducted an experiment (n = 491) examining how avatar realism and gender impact trustworthiness (expertise, integrity, and benevolence). Our findings show that higher realism enhances trustworthiness, contradicting the Uncanny Valley effect. Gender effects were dimension-specific, with male avatars rated higher in expertise. Familiarity with AI and institutional trust also shaped trustworthiness perceptions. These insights inform the design of AI avatars for effective science communication while maintaining public trust.

Keywords

AI tools in science communication; Representations of science and technology; Women in science

Received: 31st October 2024
Accepted: 1st March 2025
Published: 14th April 2025

1 ▪ Introduction

The rise of generative AI tools in science communication presents new opportunities, boosting efficiency and creativity in conveying scientific topics [De Angelis et al., 2023; Schäfer, 2023]. AI enhances scalability and creativity, making information more accessible and engaging [Stavesand & Schröder, 2024]. However, while AI technology has the capability to provide information and support understanding, it is equally crucial to examine its limitations and potential to misinform [Klein-Avraham et al., 2024]. Ensuring public trust in science communication requires transparency and ethical considerations in AI-driven communication [Kusters et al., 2020].

AI-generated content, even when scientifically accurate, may become a target for conspiracy theorists in being manipulated or misrepresented to promote false narratives, potentially undermining trust in scientific facts and science in general [Babiker et al., 2024; Rutjens & Većkalov, 2022]. Balancing these potentials and risks is crucial for AI's role in science communication [Neuberger et al., 2022].

One example of the use of generative AI tools in science communication is synthetically generated video avatars on TikTok, where historical figures like Albert Einstein or Marie Curie are brought back to life (Figure 1). Combining research results with compelling storytelling, these video formats reach up to two million impressions [Maskedteller, 2023]. Such AI-driven avatars are already an integral part of the editorial plan of educational initiatives [Stavesand & Schröder, 2024]. AI-based video generators enable the creation of these

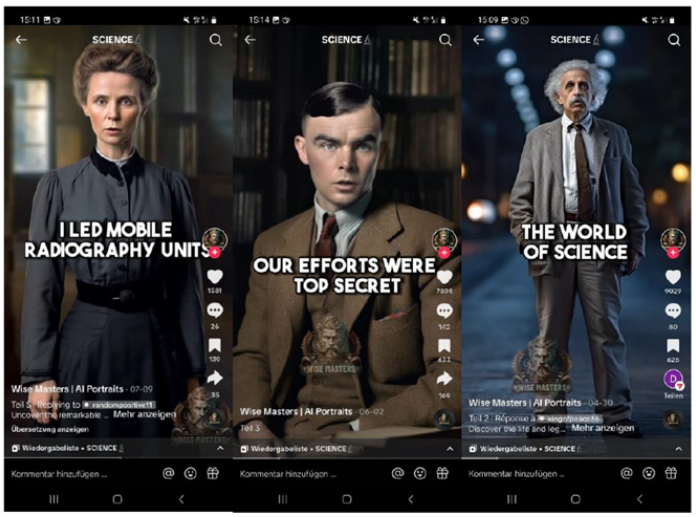

Figure 1. Video avatars of Marie Curie, Nikola Tesla and Albert Einstein on TikTok (channel: @masked-teller).

knowledge-sharing and engaging video avatars — within seconds. However, concerns have been raised about viewer skepticism stemming from the involvement of AI in content creation [Vaccari & Chadwick, 2020]. These concerns often center around trustworthiness and authenticity of AI representations. Additionally, the misuse of synthetic videos, for example as deepfakes, highlights challenges such as disinformation, misinformation, and public trust issues associated with synthetic media [Gräfe, 2022; Kieslich et al., 2024].

But unlike deceptively realistic deepfakes, avatars have an inherent characteristic: they imitate something real without claiming to be [MacDorman & Chattopadhyay, 2016]. Even if the human-like AI avatars impress with their realistic portrayal, there is something artificial about the speakers and their mouth movements or eye blinks that is often easily recognizable and can unsettle recipients [Schwind et al., 2018]. Perceived artificiality, combined with subtle inconsistencies in anthropomorphic features, can provoke unease and skepticism [Thaler et al., 2020]. In science communication, where credibility and clarity are paramount, such uncertainties could affect trust as essential component in effectively conveying factual and complex information [Bromme, 2020].

Until now, the impact of AI-generated (synthetic) formats on trust in science remains largely speculative. As a socio-technical innovation, AI is reshaping the framing of human-like representations in online environments. Research shows that trust in AI-generated content varies based on factors such as topic, application area, and presentation style [Došenović et al., 2021; Graefe et al., 2018]. While studies on the trust perception of AI-generated content have mainly been carried out on texts and in the field of automated journalism [Longoni et al., 2022], there is a lack of comparable studies of synthetic video content in the context of science communication. With AI-generated avatars becoming more prevalent in science communication, a key question arises: What factors influence the trustworthiness of these AI-driven science communicators?

Trust — and trustworthiness — seem to be strongly dependent on the physical appearance of the trustee [Duarte et al., 2012]. Research on virtual representations indicates that anthropomorphic entities can help establishing trust [Gratch et al., 2016; Kulms & Kopp, 2019]. Conversely, it has been shown that unrealistic avatars, such as those seen in cartoons, are often perceived as more credible, promoting positive interactions [Di Natale et al., 2021]. Since such stylized avatars are familiar in popular culture [Stein & MacDorman, 2024], they are likely to be perceived as more trustworthy than highly realistic, human-like avatars. Whilst in science communication, the use of AI avatars has considerable potential to impart knowledge efficiently, it may also harbor the risk of arousing mistrust due to the synthetic nature of the avatars.

Gender — which is related to physical appearance — plays an additional critical role in shaping trust perceptions of science communicators. Research has demonstrated that gender stereotypes influence the perceived trustworthiness of scientific experts [Knobloch-Westerwick et al., 2013; König & Jucks, 2019]. Moreover, negative stereotypes about women's scientific expertise not only influence perceptions of their abilities but may also lead to the devaluation of the fields in which they participate, shaping overall perceptions of scientific disciplines [Light et al., 2022]. It seems further essential to investigate the role of gender since the training data used for AI applications is presumably biased with regards to gender [Buolamwini, 2024]. These gaps highlight the need for a systematic investigation of how AI-generated avatars with different genders can convey scientific information effectively without compromising trust. Understanding these dynamics helps to create more effective, inclusive, and ethical AI avatars for science communication and beyond.

Drawing on existing research in AI, Human-Computer-Interaction (HCI), and science communication [e.g., Glikson & Woolley, 2020; Graefe et al., 2018; König & Jucks, 2019; Rogers et al., 2023], the present study focuses on two key predictors of trustworthiness: the

avatar's level of anthropomorphism and gender. The question guiding the investigation is therefore twofold: *How do the degree of **realism (1)** and portrayed **gender (2)** of AI-generated, human-like avatars affect the perceived trustworthiness of these science communicators?*

2 ▪ How Realism, Gender, and Trustworthiness Relate

Exploring trust in realistic, anthropomorphic appearances inevitably leads to the concept of the "Uncanny Valley". This effect describes a non-linear relationship between human-likeness and emotional response. While entities that are either highly stylized or fully human-like (following the similarity attraction effect), tend to be well-received, those in an intermediate, "almost-human" range evoke discomfort and distrust [Mori et al., 2012; Waddell, 2018]. This phenomenon arises from a cognitive incongruence between expected and actual appearance or behavior of an entity [MacDorman & Chattopadhyay, 2016]. However, the underlying mechanisms remain debated. Evolutionary theories suggest it functions as a threat-avoidance mechanism, where deviations from expected human morphology signal potential danger (ibid.). Cognitive psychologists postulate that uncertainty in object categorization causes the effect [Kawabe et al., 2017]. A body of research in HCI suggests that maintaining a consistent level of realism between human morphological features is essential for the emergence of the effect [MacDorman & Chattopadhyay, 2016]. The different explanations reflect an ongoing debate resurfacing as AI technologies become more powerful [Stein & MacDorman, 2024].

Previous research has shown that highly realistic avatars might trigger discomfort and affect trust in the information source [Ho & MacDorman, 2010; Schwind et al., 2018]. Conversely, stylized avatars could facilitate better parasocial interactions and credibility [Di Natale et al., 2021]. These insights into the "Uncanny Valley" have influenced design principles in popular culture, where audiences typically encounter robots and animated characters as "googly-eyed, charming cartoons" [Stein & MacDorman, 2024, p. 1]. Consequently, audiences may be more familiar and comfortable with such stylized portrayals of technology. Within the changing socio-technical framework, the question arises: Which humanlike characteristics contribute to AI avatars, with their fundamentally 'unreal' yet 'artificially intelligent' appearances, being perceived as trustworthy?

The mixed findings in existing research suggest that careful consideration is necessary when deciding between photorealistic or stylized appearances. The perception of human-likeness, particularly in highly anthropomorphic avatars, appears to depend significantly on the degree of realism. This is supported by studies on computer animations using professionally rendered 3D representations of anthropomorphic entities [Schwind et al., 2018]. Thus, realism serves as the most appropriate operationalization of human-likeness in the context of synthetic and anthropomorphic video avatars. Findings on 3D avatars suggest that a higher degree of realism correlates with an increased sense of discomfort among viewers [MacDorman & Chattopadhyay, 2016]. Conversely, unrealistic characters that maintain a consistently stylized appearance tend to be better accepted [Schwind et al., 2018]. Research on the trustworthiness of virtual agents supports these conclusions while highlighting the importance of task-specific contexts [McDonnell et al., 2012]: more realistic agents are perceived as more trustworthy in formal settings, such as medical tasks, whereas stylized agents perform better in informal social settings [Ring et al., 2014].

External science communication, where non-scientists form a significant part of the audience, is progressively framed as casual interaction [Fähnrich & Schäfer, 2020; Finkler & Leon, 2019]. Stylized, approachable formats, particularly in the online video realm, are often employed to foster emotional engagement through simplified messages that resonate with non-experts [Finkler & Leon, 2019; Reif et al., 2020]. This suggests that science communication via AI avatars operates in a more informal, accessible setting, where a less photorealistic appearance might be beneficial. Similarly, in the field of natural language processing, experts argue that synthesis errors, or "hallucinations," in generative AI might paradoxically enhance trust by making AI appear more human [Heaven, 2023]. Against this background, we hypothesize:

> *H1:* Highly stylized AI avatars communicating science are perceived as more trustworthy compared to highly realistic AI avatars.

Besides the avatar's realism, gender is a key factor in the perception of science communicators. Particularly in STEM fields, different genders are associated with stereotypes that influence the evaluation of scientists [Eaton et al., 2020; Gheorghiu et al., 2017]. Male scientists are often perceived as "highly competent but less warm-hearted", while their female colleagues are often seen as "less competent but more warm-hearted" [Reif et al., 2020, p. 193]. Media representations reinforce the gender-specific stereotypes, shaping trustworthiness perceptions of scientists [Reif et al., 2020; Jarreau et al., 2019]. Further, taking gender appearances of science communicators' avatars into account is critical, as AI-generated representations are likely shaped by biased training data that can reinforce traditional stereotypes of scientific competence [Buolamwini, 2024]. Therefore, the portrayal of scientist's gender in AI-generated video content must be considered in discussions of perceived trustworthiness. Based on this, we assume:

> *H2:* AI avatars perceived as male communicating science are evaluated as more trustworthy compared to AI avatars perceived as female.

In the context of a changing socio-technical landscape, the interplay between gender-specific media bias, perceived realism of synthetic media, and trustworthiness becomes more apparent. Existing media content, where female scientists are systematically underrepresented, likely forms a substantial portion of the data sets used to train generative AI models [Buolamwini, 2024; Criado-Perez, 2019]. Consequently, this introduces a bias in AI-generated, human-like portrayals [Nightingale & Farid, 2022]. Therefore, the AI-generated avatars might be perceived differently in terms of their degree of realism, depending on the gender of the science communicator they represent. An interaction between the avatar's gender and degree of realism is anticipated:

> *H3:* The degree of realism and gender of AI-generated video avatars interact in their effect on perceived trustworthiness, with the strongest effect occurring for stylized avatars perceived as male compared to highly realistic avatars perceived as female.

According to Hendriks et al. [2015], the perceived trustworthiness of scientific experts is composed of three dimensions: expertise, integrity, and benevolence. This framework allows for a more detailed understanding of the trustworthiness evaluation of AI-generated avatars that communicate scientific topics. While general trustworthiness perceptions may align with gender stereotypes [Gheorghiu et al., 2017], findings on the subdimensions remain inconsistent. Some studies indicate that female scientist representations can enhance perceived competence, challenging assumptions of male dominance in expertise evaluations [Fiske et al., 2018; Jarreau et al., 2019]. However, gendered effects on benevolence and integrity vary, with some studies suggesting female scientists are perceived as warmer, while others report no significant differences [Jarreau et al., 2019; Reif et al., 2020]. Additionally, studies suggest that the importance of each subdimension can vary depending on the context and type of scientific communication [e.g., Besley et al., 2021; Könneker, 2024]. As the effects of science communicators' gender on epistemic trustworthiness have been studied for different visual appearances but not for AI-generated avatars, we investigate:

> *RQ1:* How do the three dimensions of epistemic trustworthiness (expertise, integrity, and benevolence) vary as a function of an AI avatar's perceived gender?

Realistic anthropomorphic avatars may enhance perceptions of competence, yet research in HCI typically assesses avatars' trustworthiness on a unidimensional level [e.g., Klein-Avraham et al., 2024]. However, when realistic avatars communicate science, epistemic trustworthiness — which encompasses expertise, integrity, and benevolence — might provide a more suitable framework for evaluation [Hendriks et al., 2015]. It remains unclear in which way the degree of realism shapes trustworthiness in these subdimensions. Given that AI-generated avatars as science communicators have not yet been examined in detail using this multidimensional trustworthiness framework, we ask:

> *RQ2:* How do the three dimensions of epistemic trustworthiness (expertise, integrity, and benevolence) vary as a function of an AI avatar's degree of realism?

3 ▪ Method

To test the influence of realism and gender of anthropomorphic AI entities on viewers' trustworthiness evaluations, this study employs a 2×2 between-subjects design. Participants viewed AI-generated video avatars presenting scientific findings. The study included two pretests prior to the main online-experiment: a quantitative pretest (n = 481) and a qualitative pretest (think-aloud interviews, n = 8).

3.1 ▪ *Manipulations*

For the study we created four one-minute-long video stimuli using the AI-based video generator HeyGen.[1]

1. The paid "Creator" version was used: https://www.heygen.com/.

Avatar appearance and voice. The videos showed either a female or male anthropomorphic avatar in a highly realistic or stylized form. To maintain consistency and minimize confounding factors, all avatars were designed with similar characteristics based on the classification of anthropomorphic entities by Ring et al. [2014]. The highly realistic avatars (A1 and A2) were rendered with detailed skin textures, while the highly stylized avatars (A3 and A4) were generated from images, resulting in a more cartoon-like appearance (Figure 2). The same voice, generated by *Elevenlabs'* free text-to-speech software,[2] was used for both male and female avatars within each condition to ensure consistency in audio.

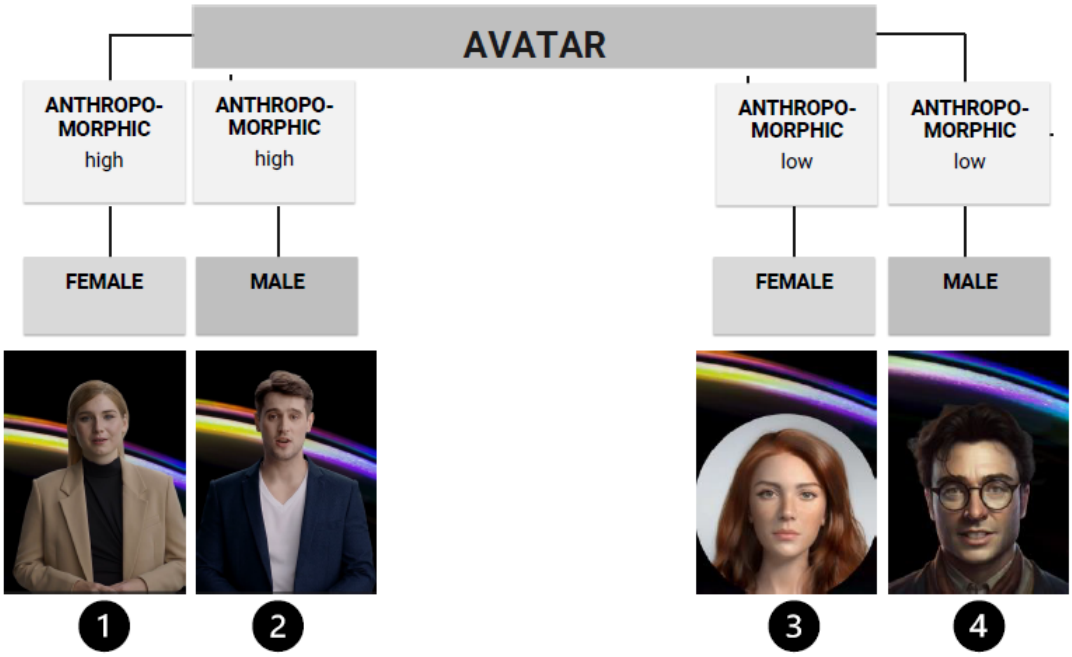

Figure 2. Stimuli.

Text and topic. The same text was used across all avatar conditions. This text was adapted from a YouTube video by German science communicator Mai Thi Nguyen-Kim,[3] whose clear and accessible communication style was deemed suitable for the communication of scientific messages [Reif et al., 2020; Ruzi et al., 2021]. The script for the video incorporated real research findings on color blindness treatment through gene therapy [Mancuso et al., 2009] to preserve external validity (see appendix in the supplementary material). A STEM-related subject was selected to better observe the potential influence of gender on avatar perception, given the underrepresentation of women in these fields [UNESCO Institute for Statistics, 2019]. To reduce potential bias, individual researchers were not named as sources in the videos. Instead, the research was presented as the work of a "research team" ensuring the focus remained on the avatars' visual characteristics.

2. https://elevenlabs.io/.
3. A one-minute text excerpt from the YouTube video https://www.youtube.com/watch?v=r0jXfwPQW9k&t=17s [MAITHINK X, 2018].

3.2 ▪ *Measures*

The degree of realism was operationalized using three item pairs on a 5-point Likert scale from MacDorman and Chattopadhyay [2016]: *computer-animated — real*, *replica — original*, and *digitally copied — authentic*. Item pairs were averaged and summarized as "Realism Scale" with higher mean values indicating greater perceived realism. Additionally, participants were asked to indicate the extent to which they perceived the video avatar as female to gauge perceived gender.

Trustworthiness is measured using the Muenster Epistemic Trustworthiness Inventory (METI) [Hendriks et al., 2015]. This instrument included 14 item pairs across its three subscales (expertise, benevolence, and integrity), which are evaluated using semantic differentials on a 5-point Likert scale.[4]

Reliability tests demonstrated high internal consistency across the scales used, with Cronbach's alpha values exceeding 0.9. The METI scale's criterion validity was confirmed by a strong positive correlation ($r = .838$, $p < .001$) with a separately measured global trustworthiness item (*non-trustworthy — trustworthy*). Figure 2 provides an overview of the key constructs directly related to the hypotheses. Additional variables are discussed in the following section.

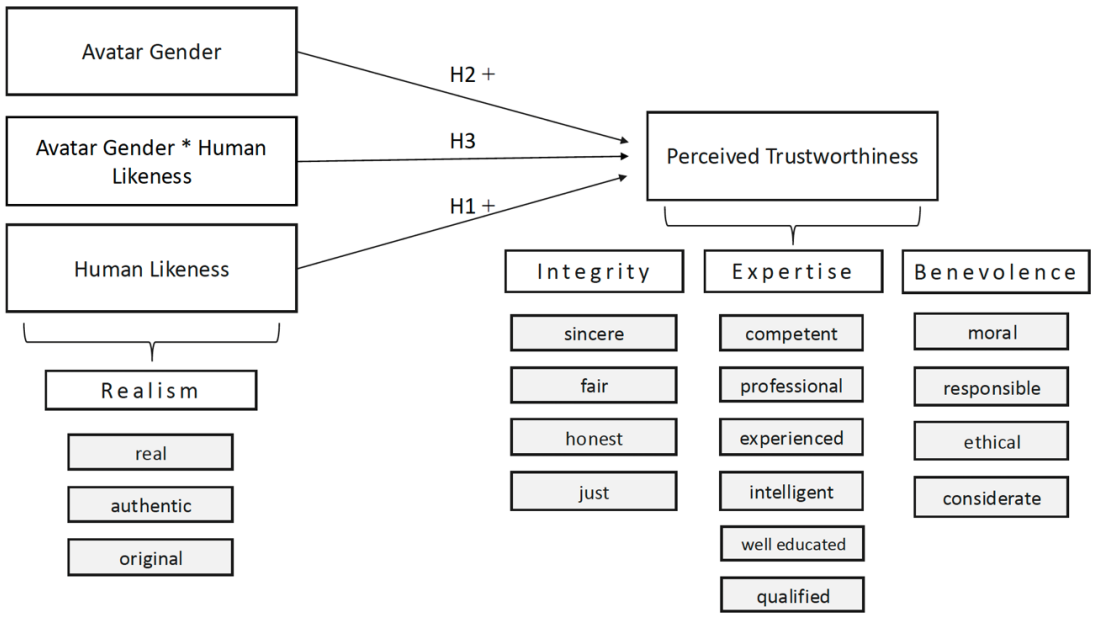

Figure 3. Key constructs and their measures.

3.3 ▪ *Control variables*

Previous research indicates that trust in and familiarity with science as well as exposure to scientific information may affect the perceived trustworthiness of science communicators

4. The item pairs are shown in their positive form but were measured as bipolar adjective pairs.

[Došenović et al., 2021; Wissenschaft im Dialog, 2023]. Participants were asked about their general trust and interest in science [Wissenschaft im Dialog, 2023], media usage habits, and topic involvement before the treatment. The measures are:

- Topic involvement: "How do you rate your knowledge in the field of genetics?" (*very low-very high*)
- Trust in science: "How much do you trust science and research?" (*trust fully-no trust at all*)
- Interest in science: "How high is your interest in scientific topics in the following fields[5]*?*" (3 Items; *very low-very high*).

Interest in science, along with the METI and realism scales, was computed using mean scores, with higher mean values indicating greater levels of interest, perceived trustworthiness, or realism, respectively. For more detailed analyses, the mean scores of the individual METI subdimensions — expertise, benevolence, and integrity — were calculated and reported separately. Table 1 provides an overview of all means and standard deviations.

Table 1. Descriptive statistics for study variables. *Note.* All variables were measured on a 5-point Likert scale, where 1 indicated the lowest and 5 the highest rating. *Variables reflect the mean across all items within each respective scale.

Variable	n	M	SD	α
Trustworthiness				
- Expertise*	485	3.77	0.93	.94
- Integrity*	480	3.72	0.92	.91
- Benevolence*	480	3.64	0.93	.90
Realism*	491	2.48	1.37	.95
Topic involvement	479	3.08	1.17	.
Trust in science	474	3.79	0.94	.
Interest in science*	486	3.24	0.93	.98

Two pretests were conducted prior to the main study to check the quality of the stimuli and the reliability of the scales.

3.4 ▪ *Quantitative pretest*

The quantitative pretest was conducted in December 2023 with 481 participants from an online open-access panel (50.1% female, 49.9% male). The average age was 43 years (*SD* = 13.2), ranging from 18 to 64. A total of 50.1% of participants had acquired a technical diploma or higher. The pretest consisted of 20 items, administered after another survey on digital transformation research. The scales (see *Measures*) were found to be reliable, with high internal consistency (α > .89), making them suitable for use in the main survey without any modifications.

We further aimed to check at an early stage whether the described manipulation of the AI-generated video avatars needed to be adapted. Participants were randomly assigned to

5. Engineering (e.g., Computer Science, Technology), Life Sciences (e.g., Medicine, Biology), Natural Sciences (e.g., Chemistry, Physics).

one of the four treatment groups. The further procedure was comparable to that in the main study reported below. The manipulation check showed that the perceived realism of the realistic avatars was significantly higher than that of the stylized avatars (Table 2), although the mean difference of 0.43 scale points was not substantial and the effect size rather low ($t(479) = 3.6$, $p < .001$).

Table 2. Manipulation check (pretest). Items measuring avatars' realism: "The person speaking in the video about science appears to me as…" (1 = *computer-animated*; 5 = *real* | 1 = *replica*; 5 = *original* | 1 = *digitally copied*; 5 = *authentic*). Item pairs were averaged as Realism Scale (see *Measures*).

	N	*M*	*SD*	*T*	*df*	*p*	*Cohen's d*
Highly realistic (A1, A2)	239	2.61	1.32	3.6	479	< .001	.33
Stylized (A3, A4)	242	2.18	1.34				

The results of the manipulation check underscore the need for enhancing the consistency of the anthropomorphic features in the highly realistic avatars. The findings suggest incorporating additional control variables and conducting a second qualitative pretest to pinpoint specific improvements in the video stimuli and refine the overall structure of the online questionnaire.

3.5 ▪ *Qualitative pretest*

Eight think-aloud interviews were conducted in January 2024. Participants, recruited through convenience sampling, had a balanced gender distribution (50% female, 50% male), with an age range from 18 to 58 years (M_{age} = 42.0, SD_{age} = 22.8). The interviews were recorded and documented. The aim of the qualitative pretest was twofold: to gain initial insights into the trustworthiness assessments and to test whether the topic of red-green color blindness is sufficiently relevant to influence trust judgments.

The qualitative pretest revealed three key findings. First, participants tended to select the middle of the scale rather than the "Don't know" option when unsure about certain trustworthiness items, prompting us to make the "Don't know" button more visible. Second, participants found it difficult to assess the integrity dimension (e.g., sincerity, fairness), with feedback suggesting these attributes were not suitable for evaluating the video avatars, leading to textual adjustments in the stimulus material. Lastly, the avatars' portrayal as researchers was perceived as unrealistic, prompting changes to present them as science communicators instead, alongside improvements in voice synchronization and audio quality.

3.6 ▪ *Sample*

In the main study, 567 participants were recruited via *Dynata's* open-access online panel. Quotas were set for age, gender, and region (federal states) based on German national census data. The survey was conducted from February 22, 2024, to March 5, 2024. The data was cleaned of outliers, resulting in a final sample of 491 cases (M_{age}= 43, SD_{age}=12.6).[6]

6. To ensure a reliable dataset, data sets from panelists whose total time spent in the online experiment was more than one standard deviation below the average were excluded.

While 50% identified as female, 49% identified as male. The remaining participants chose not to disclose their gender identity. Most participants were well-educated, 52.61% of them claim that they acquired a technical diploma or above.

3.7 ▪ *Procedure*

After consent was obtained, participants were randomly assigned to one of the four conditions. After the stimulus and the trustworthiness evaluation, participants were asked whether they believed that the expert in the video was AI-generated. This question checked whether participants immediately recognized the artificial nature of the avatar without influencing the trust ratings. To focus on the effects of realism and gender, the videos were not explicitly labeled as "artificial" during the treatment, avoiding potential bias from preconceived notions about AI [MacDorman, 2024]. While some participants might still identify the avatar as artificial, potentially leading to mistrust [Waddell, 2018], this trade-off was deemed necessary to isolate the impact of the avatars' level of realism and gender. Participants' prior knowledge of AI [Došenović et al., 2021], their experiences with synthetic online content, and their personal use of AI-based software were surveyed. The survey concluded with a debriefing that clarified the study's objectives and explained that the AI-generated avatar was not a real person but created for research purposes.

4 ▪ Results

4.1 ▪ *Manipulation checks*

Participants rated highly realistic avatars as significantly more realistic than stylized avatars (M = 2.90, SD = 1.35 vs. M = 2.07, SD = 1.27; $t(489) = 7.00$, $p < .001$).[7] Gender manipulation was also successful, with participants correctly identifying the gender of the avatars (female: M = 4.48, SD = 1.01; male: M = 1.32, SD = 0.94; $t(697.867) = -7.500$, $p < .001$).

4.2 ▪ *Analysis of perceived trustworthiness*

The descriptive analysis indicates that highly realistic avatars receive higher ratings across most dimensions of trustworthiness (Table 3). The largest difference appears in expertise, where the stylized female avatar is rated notably lower (M = 3.44, SD = 1.01) than other AI avatars, while the stylized male avatar stands out as the highest rated avatar (M = 3.91, SD = .85). Regarding benevolence, the stylized female avatar receives the lowest rating (M = 3.41, SD = .94), while the realistic male avatar receives the highest rating (M = 3.81, SD = .91). Integrity ratings show less variation across conditions but tend to be slightly higher for more realistic avatars.

To provide a more detailed understanding of evaluation patterns, further analyses differentiate between the three trustworthiness dimensions — expertise, integrity, and benevolence. Preliminary, three 2×2 ANOVAs were conducted to explore main and interaction effects (see appendix in the supplementary material). To account for potential confounding variables, subsequent ANCOVAs were conducted including the control variables. As ANCOVA results were largely consistent with the ANOVA results, while providing a more

7. Items measuring avatars' realism: "The person speaking in the video about science appears to me as…" (1 = *computer-animated*; 5 = *real* | 1 = *replica*; 5 = *original* | 1 = *digitally copied*; 5 = *authentic*).

Table 3. Means and Standard Deviations for the Dependent Variables by Avatar-Gender and Degree of Realism (Stylized vs. Highly Realistic). Note. Standard deviations are presented in parentheses.

Avatar-Gender	**Female**		**Male**	
Realism	*n*	Stylized	Realistic	Stylized	Realistic
Trustworthiness					
Expertise	485	3.44 (1.01)	3.83 (.92)	3.91 (.85)	3.88 (.84)
Integrity	480	3.50 (.98)	3.81 (.85)	3.77 (.90)	3.83 (.92)
Benevolence	480	3.41 (.94)	3.78 (.87)	3.57 (.96)	3.81 (.91)

robust estimate by controlling for covariates, only ANCOVAs are reported below for hypothesis testing. The statistical model included avatar gender and realism as independent factors and controlled for prior AI knowledge, trust in science, AI software use, prior experience with AI video content, topic involvement, participants' education level, gender, and age as covariates (see Table 4).

All tests were two-tailed, with an alpha level set at 0.05. We interpreted partial eta squared (η^2) effect sizes based on Cohen's [1992].

Expertise. We found a significant main effect of avatar gender, $F(1, 385) = 7.41$, $p = .007$, $\eta^2 = .019$, indicating that male avatars were perceived as more competent than female avatars. Realism had a significant effect, $F(1, 385) = 4.40$, $p = .037$, $\eta^2 = .011$, with highly realistic avatars rated higher in expertise than stylized ones. The interaction effect between avatar gender and realism did not reach significance, $F(1, 385) = 3.52$, $p = .061$. Among the covariates, prior AI knowledge, trust in science, and AI software use significantly influenced expertise ratings.

Integrity. A significant main effect of realism was found, $F(1, 380) = 10.2$, $p = .002$, $\eta^2 = .026$, indicating that highly realistic avatars were rated as more trustworthy than stylized ones. The main effect of gender ($p = .108$) and the interaction effect were non-significant ($p = .238$). Several covariates significantly shaped integrity perceptions, including AI knowledge, trust in science, AI software use, topic involvement, education level, and age, suggesting that individual characteristics shape dimension-specific evaluations.

Benevolence. The analysis for benevolence showed the strongest main effect of realism, $F(1, 381) = 16.1$, $p < .001$, $\eta^2 = .041$, with realistic avatars perceived as more benevolent than stylized ones. However, avatar gender had no significant effect, $F(1, 381) = .28$, $p = .592$. Further, no significant interaction effect was found, $F(1, 381) = .05$, $p = .819$. Covariates such as AI knowledge, trust in science, AI software use, and age significantly influenced benevolence ratings.

Taken together, realism significantly increases trustworthiness across all subdimensions, contradicting H1 and suggesting that highly realistic AI avatars do not induce an uncanny valley effect. This finding addresses RQ2, emphasizing that higher realism positively influences all three subdimensions of epistemic trustworthiness, with varying effect sizes. H2 was partially supported, as male avatars were rated higher in expertise, but gender did not affect integrity or benevolence. This answers RQ1, showing that gendered perceptions of competence persist, while other trustworthiness dimensions remain unaffected. Although an interaction effect on expertise appeared in the ANOVA (see appendix in the supplementary material), it did not remain significant when controlling for covariates. Thus, H3 was not supported.

Table 4. Results for ANCOVAs for the Dependent Variables Expertise, Integrity, Benevolence. *Note.* *p < .05, **p < .01, ***p < .001; AV-Gender = Avatar-Gender.

Dependent Variable	Source of Variance	df	F	Partial η^2	p
Trustworthiness					
Expertise	AV-Gender	1	7.41**	.019	.007
	Realism	1	4.40*	.011	.037
	AV-Gender*Realism	1	3.52	.009	.061
	Prior AI Knowledge	1	6.55*	.017	.011
	Trust in Science	1	5.02*	.013	.026
	AI Software Use	1	5.21*	.013	.023
	Prior AI Experience	1	.06	.000	.809
	Topic Involvement	1	1.75	.005	.187
	Education Level	1	3.60	.009	.058
	Gender	1	6.67*	.017	.010
	Age	1	3.06	.008	.081
	Error	385			
Integrity	AV-Gender	1	3.06	.007	.108
	Realism	1	10.2**	.026	.002
	AV-Gender*Realism	1	1.40	.004	.238
	Prior AI Knowledge	1	15.4***	.038	<.001
	Trust in Science	1	14.1***	.036	<.001
	AI Software Use	1	14.2***	.036	<.001
	Prior AI Experience	1	.001	.000	.982
	Topic Involvement	1	5.12*	.013	.024
	Education Level	1	4.78*	.012	.029
	Gender	1	5.24*	.014	.023
	Age	1	7.10**	.018	.008
	Error	380			
Benevolence	AV-Gender	1	.28	.001	.592
	Realism	1	16.1***	.041	<.001
	AV-Gender*Realism	1	.05	.000	.819
	Prior AI Knowledge	1	8.12**	.021	.005
	Trust in Science	1	12.1***	.031	<.001
	AI Software Use	1	10.31**	.026	.001
	Prior AI Experience	1	.34	.001	.529
	Topic Involvement	1	2.45	.006	.117
	Education Level	1	2.28	.006	.132
	Gender	1	1.11	.005	.163
	Age	1	4.94*	.013	.027
	Error	381			

5 ▪ Discussion

The study aimed to explore the role of realism and gender in shaping the perceived trustworthiness of AI-generated avatars in science communication. Contrary to our assumptions (H1), avatars with higher levels of realism were perceived as more trustworthy than their stylized counterparts across all subdimensions (RQ2). The influence of gender on the three sub-dimensions of trustworthiness (H2) was inconsistent. While male avatars were generally perceived as more competent, this advantage did not extend uniformly across all dimensions of trustworthiness, such as integrity and benevolence (RQ1). AI avatars' realism and gender influenced trustworthiness independently rather than interactively (H3). In addition, individual predispositions — namely trust in science, prior knowledge of AI, prior use of AI-software, topic involvement, and age — appear to interact with the effects of the avatar characteristics in shaping perceptions of trustworthiness. This underscores the complexity of trustworthiness perceptions and highlights the need for a more detailed examination of the factors at play.

The findings provide nuanced insights into the complex dynamics of trust formation, revealing five key insights:

No evidence of a descent into the "Uncanny Valley". Overall, synthetic experts were perceived as trustworthy, with ratings slightly above the midpoint of the scale. This suggests that participants attributed a level of judgment to the avatars, aligning with the concept of epistemic trustworthiness [Hendriks et al., 2015; Könneker, 2024]. Contrary to the "Uncanny Valley" hypothesis, which suggests increased discomfort and trust dips at intermediate levels of realism [Mori et al., 2012], our findings suggest, that more realistic AI avatars in science communication are perceived as more trustworthy than their stylized counterparts. This challenges previous research that associates higher realism with unease and reduced trust [MacDorman & Chattopadhyay, 2016; Schwind et al., 2018]. Instead, the results support the similarity-attraction effect, where human-like features evoke greater trust [Waddell, 2018]. These findings align with McDonnell et al. [2012] and Ring et al. [2014], who demonstrated that trust in virtual agents is context-dependent, with realistic agents being more effective in formal settings. Within external science communication, which spans a spectrum from formal educational settings to more casual, accessible interactions [Trench & Bucchi, 2010], realistic AI avatars appear to be generally well-accepted. Although avatars must be distinguished from deepfakes, which have crossed the "Uncanny Valley" as deceptive copies [Nightingale & Farid, 2022], the findings indicate that a highly stylized appearance is not essential to maintain trust, suggesting that AI-generated avatars may be farther from the "Uncanny Valley" than previously assumed.

The Uncanny Valley effect can also be understood as a protective mechanism fostering critical engagement with AI tools. Research suggests that the discomfort triggered by highly human-like but artificial agents can serve as a cognitive safeguard, prompting users to scrutinize and evaluate AI-generated information more carefully [Rosenthal-Von der Pütten et al., 2019]. In our study, however, there was no evidence of such a protective mechanism in the evaluation of realistic AI avatars of science communicators. Another possible explanation is that the degree of realism was not high enough to evoke the uncanniness associated with the effect [Song & Shin, 2024]. This interpretation is consistent with the moderate realism ratings of the avatars in the manipulation check, suggesting that their artificial nature remained perceptible, thereby avoiding the perceived ambiguity that normally triggers this protective mechanism.

Importance of communication context. Previous research on virtual agents and automated journalism has emphasized the significance of topic and context in shaping trustworthiness perceptions [Graefe et al., 2018; Ring et al., 2014]. In line with these studies, the context of science communication appears to play a crucial role in our findings. Science communicators operate within an environment where trust in science as an institution remains relatively high [Bromme, 2020]. Within this framework, human-like AI-generated avatars are more likely to evoke trust. This suggests that lay audiences may respond more positively to these AI avatars, enhancing their perceived trustworthiness. By reflecting earlier findings on the context-specific effects of realistic appearance [McDonnell et al., 2012; Ring et al., 2014], this study reinforces the notion that realism is particularly effective in domains where institutional trust, like science communication, is already strong.

Building on these insights, the use of AI to create realistic avatars of science communicators extends beyond enhancing scalability and creativity. It underscores the socio-cultural and ethical dimensions of applying AI-based technologies [Schäfer & Wessler, 2020; Stein et al., 2023], particularly their potential to shape trustworthiness perceptions in synthetic video content. While realistic science communication avatars may enhance trustworthiness, creating synthetic content indistinguishable from humans — characteristic of deepfakes — raises broader ethical concerns about misuse in knowledge dissemination. For science communicators, finding a balance between leveraging the benefits of AI-generated appearances and avoiding the risks of misleading synthetic content, such as deepfakes, is vital to maintain public trust and uphold the integrity of scientific communication.

Integrity and benevolence linked to realism, expertise to gender. When examining the subdimensions of trustworthiness, gender had a significant impact on expertise. Male avatars, even when stylized, were consistently perceived as more competent than their female counterparts, indicating that gender biases persist [König & Jucks, 2019]. However, the minimal differences in trustworthiness ratings between female realistic avatars and male stylized avatars indicate that gender alone may not be a decisive factor in overall trustworthiness perceptions. This contrasts with the evaluations of integrity and benevolence, where avatar's gender had no significant influence, and realism played a more substantial role. It seems that more human-like avatars are perceived as having better intentions and being more moral, which are crucial factors in trust attribution [Könneker, 2024]. This indicates that the influence of avatars' gender and realism on trustworthiness is dimension-specific: realism primarily affects evaluations of integrity and benevolence, while gender is more influential in the assessment of expertise. The latter result is consistent with previous findings on gender-stereotypical confidence judgements among science communicators, according to which the male-read gender representation has a more 'competence-enhancing' effect than the female-read gender representation [König & Jucks, 2019; Reif et al., 2020]. Our results further align with research on human science communicators, which demonstrates that socio-cognitive facial traits often "trump" demographic factors such as scientist's age, gender, and ethnicity in shaping trustworthiness perceptions [Gheorghiu et al., 2017]. The stereotypical thought patterns of competence also seem to persist when the science communicators are presented as AI-generated avatars.

Impact of individual AI experiences. The significant effects of covariates indicate that individual predispositions shape trustworthiness perceptions of AI-generated science communicators. Specifically in terms of the evaluation of avatars' integrity, the use of AI-based software and self-assessed prior knowledge of AI could have impacted the

observed effects of realism. Individuals who frequently encounter AI-generated content may perceive anthropomorphic entities differently. Given the rapid but uneven diffusion of AI in society, exposure to AI content varies, leading to differing opinions and preconceptions about AI [Meckel & Steinacker, 2024; Stein et al., 2023]. Based on these experiences, opinions or basic preconceptions about AI are formed [e.g. Swart, 2021]. These, in turn, can influence the trustworthiness attributed to a source [Došenović et al., 2021; Stein et al., 2023]. However, as the study only accounted for prior experience with AI videos and did not systematically manipulate transparent AI labeling, definitive conclusions cannot be drawn. The absence of AI labeling for avatars also could have affected trust perceptions. Although labeling was intentionally avoided to prevent framing effects, future studies should investigate how labeling impacts trust.

Trust in science was a strong predictor of integrity and benevolence ratings, suggesting that pre-existing institutional trust moderates trustworthiness judgments of these subdimensions. This underscores the multidimensional nature of epistemic trustworthiness, where attributed expertise, perceived adherence to scientific and societal norms play a crucial role. Individuals with greater trust in science are more likely to attribute judgment and integrity to AI-generated science communicators [Könneker, 2024]. These findings highlight that trustworthiness perceptions of AI avatars are not formed in isolation but mediated by users' trust in science, experiences, and familiarity with AI-generated content [Stein et al., 2023]. Future research should explore how different levels of AI literacy and affinity with science influence the perception of synthetic science communicators.

6 ▪ Limitations and Future Directions

This study provides key insights into the trustworthiness of AI-generated avatars in science communication. However, there are several limitations. Its short-term experimental design may not capture the full complexity of trust formation in real-world AI interactions [Glikson & Woolley, 2020]. Future research should explore long-term trust development and the effects of continuous exposure to AI avatars.

The 2×2 design further limits the scope to four experimental conditions with one video stimulus per condition. While two pretests with manipulation checks confirmed the effectiveness of the independent variable manipulations, relying on only four stimuli — one for each experimental condition — limits the generalizability of the findings and raises the possibility that effects might be influenced by uncontrolled video-specific characteristics rather than the intended variations in gender and stylization. To mitigate this concern, all stimuli were generated using the same AI-based video tool, ensuring standardized modeling, texturing, and rendering to maintain consistency across conditions. However, subtle variations in presentation quality — such as avatars' attractiveness, facial symmetry, or animation smoothness — cannot be entirely ruled out as potential influencing factors. Future studies should extend this approach by incorporating multiple stimuli for each experimental condition to test for robustness across different video presentations — and ideally also topics.

While more realistically rendered AI avatars were perceived more human-like than the stylized avatars, their level of realism was still rather low. Despite the observed trust-enhancing effect of more human-like AI avatars, technical limitations, such as poor facial expressions, reduced perceived authenticity. As AI technology improves quickly, these

rendering issues are expected to diminish [Glikson & Woolley, 2020]. Despite technical constraints, this study highlights that AI-based avatars, as artifacts of a socio-technical innovation process, are not merely technological advancements but also involve socio-cultural and ethical dimensions that must be integrated into science communication [Schäfer & Wessler, 2020].

While our study focused on gender and realism, other characteristics of AI avatars, such as scientists' age, skin-color and discipline, may also influence trustworthiness perceptions [Gheorghiu et al., 2017; Reif et al., 2020]. These factors were held constant to isolate the effects of gender and realism but warrant further investigation to better understand trustworthiness in diverse contexts.

The METI scale used to measure trustworthiness [Hendriks et al., 2015] showed limitations, with numerous "don't know" responses, indicating it may require adjustments for online contexts involving AI representations. Think-aloud interviews revealed that participants particularly struggled to assess the dimension of integrity. Additionally, the study's focus on a more neutral topic limits generalizability to more controversial areas where trust in AI avatars may vary [Bromme, 2020]. Future research should also examine how the interaction between source, medium, and message influences the credibility of AI-generated science communicators.

A further limitation of this study is its cultural context, Germany, where AI skepticism is relatively high [Rühle, 2021], potentially influencing AI avatar evaluations. Future research should examine whether our findings generalize to countries with more positive attitudes toward AI. Moreover, the topic of color blindness, rooted in natural sciences, may have interacted with gender biases, as women in STEM often face skepticism regarding their expertise [Eaton et al., 2020], which could shape how female-presenting avatars are judged. This suggests that the scientific domain of the communication topic itself may interact with gendered perceptions of trustworthiness. Future research should explore how different scientific disciplines and topics influence the reception of AI-generated avatars, ensuring a broader understanding of the factors that shape their perceived trustworthiness.

7 ▪ Conclusions

This study investigated the impact of anthropomorphism and gender on the perceived trustworthiness of AI-generated avatars in science communication. Key findings revealed that more realistic avatars were perceived as more trustworthy than their stylized counterparts, challenging the "Uncanny Valley" concept in this context. The results emphasize the significance of the communication context, indicating that AI avatars resembling human experts are more likely to be trusted within a more expertise-oriented framework like science communication. The findings highlight the complex interplay between realism and gender, noting that the male avatar was perceived as more competent in the lower-realism condition.

The findings provide science communicators with practical advice on the key elements to focus on when designing AI avatars that capture public interest without compromising trust. On a broader level, this study offers preliminary insights into the socio-technical framing of

AI-generated anthropomorphic avatars in science communication. As generative AI models advance and public familiarity with AI tools and synthetic content grows, future research must delve deeper into the ethical, social, and technological implications of AI representations. In the field of science communication continued research is essential to ensure the effective and ethical integration of AI-based content into scientific discourse, keeping pace with the rapidly changing socio-technical environment. Understanding these dynamics is crucial to ensuring that AI-generated avatars enhance, rather than undermine, public trust in science.

Acknowledgments

We sincerely thank the two anonymous reviewers for their helpful feedback to improve the manuscript. This article is based on a master's thesis, and the trust and support of the team at the Center for Advanced Internet Studies (CAIS) played a vital role in its development.

References

Babiker, A., Alshakhsi, S., Al-Thani, D., Montag, C., & Ali, R. (2024). Attitude towards AI: potential influence of conspiracy belief, XAI experience and locus of control. *International Journal of Human–Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2024.2401249

Besley, J. C., Lee, N. M., & Pressgrove, G. (2021). Reassessing the variables used to measure public perceptions of scientists. *Science Communication*, *43*, 3–32. https://doi.org/10.1177/1075547020949547

Bromme, R. (2020). Informiertes Vertrauen: Eine psychologische Perspektive auf Vertrauen in Wissenschaft. In *Wissenschaftsreflexion* (pp. 105–134). Brill | mentis.

Buolamwini, J. (2024). *Unmasking AI: my mission to protect what is human in a world of machines*. Random House.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Criado-Perez, C. (2019). *Invisible women — exposing data bias in a world designed for men*. Penguin Random House.

De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of Large Language Models: the new AI-driven infodemic threat in public health. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4352931

Di Natale, A. F., Triberti, S., Sibilla, F., Imperato, C., Villani, D., Mancini, T., & Riva, G. (2021). Behind a digital mask: users' subjective experience of animated characters and its effect on source credibility. *Interacting with Computers*, *33*, 499–510. https://doi.org/10.1093/iwc/iwab030

Došenović, P., Kieslich, K., & Keller, B. (2021). *Meinungsmonitor Künstliche Intelligenz. Methodensteckbrief: Monitor- und Sonderbefragung*. https://www.cais-research.de/wp-content/uploads/Methodensteckbrief2.pdf

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: the role of appearance in peer-to-peer lending. *Review of Financial Studies*, *25*, 2455–2484. https://doi.org/10.1093/rfs/hhs071

Eaton, A. A., Saunders, J. F., Jacobson, R. K., & West, K. (2020). How gender and race stereotypes impact the advancement of scholars in STEM: professors' biased evaluations of physics and biology post-doctoral candidates. *Sex roles*, *82*, 127–141. https://doi.org/10.1007/s11199-019-01052-w

Fähnrich, B., & Schäfer, M. S. (2020). Wissenschaftskommunikation zwischen Gesellschafts-, Wissenschafts- und Medienwandel. *Publizistik*, *65*, 515–522. https://doi.org/10.1007/s11616-020-00623-2

Finkler, W., & Leon, B. (2019). The power of storytelling and video: a visual rhetoric for science communication. *JCOM*, *18*, A02. https://doi.org/10.22323/2.18050202

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2018). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. In *Social cognition* (pp. 162–214). Routledge.

Gheorghiu, A. I., Callan, M. J., & Skylark, W. J. (2017). Facial appearance affects science communication. *Proceedings of the National Academy of Sciences*, *114*, 5970–5975. https://doi.org/10.1073/pnas.1620542114

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. *Academy of Management Annals*, *14*, 627–660. https://doi.org/10.5465/annals.2018.0057

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: credibility, expertise and readability. *Journalism*, *19*, 595–610. https://doi.org/10.1177/1464884916641269

Gräfe, H.-C. (2022). Deepfakes. *Journalistikon*. https://journalistikon.de/deepfakes/

Gratch, J., DeVault, D., & Lucas, G. (2016). The benefits of virtual humans for teaching negotiation. In D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer & A. Leuski (Eds.), *Intelligent virtual agents* (pp. 283–294). Springer International Publishing. https://doi.org/10.1007/978-3-319-47665-0_25

Heaven, W. D. (2023). Artificial Intelligence. Geoffrey Hinton tells us why he's now scared of the tech he helped build. *MIT Technology Review*. https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/

Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: the Muenster Epistemic Trustworthiness Inventory (METI) (J. M. Wicherts, Ed.). *PLOS ONE*, *10*, e0139309. https://doi.org/10.1371/journal.pone.0139309

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*, 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Jarreau, P. B., Cancellare, I. A., Carmichael, B. J., Porter, L., Toker, D., & Yammine, S. Z. (2019). Using selfies to challenge public stereotypes of scientists. *PLOS ONE*, *14*, e0216625. https://doi.org/10.1371/journal.pone.0216625

Kawabe, T., Sasaki, K., Ihaya, K., & Yamada, Y. (2017). When categorization-based stranger avoidance explains the uncanny valley: a comment on MacDorman and Chattopadhyay (2016). *Cognition*, *161*, 129–131. https://doi.org/10.1016/j.cognition.2016.09.001

Kieslich, K., Helberger, N., & Diakopoulos, N. (2024). My future with my chatbot: a scenario-driven, user-centric approach to anticipating AI impacts. *The 2024 ACM Conference on Fairness, Accountability and Transparency*, 2071–2085. https://doi.org/10.1145/3630106.3659026

Klein-Avraham, I., Greussing, E., Taddicken, M., Dabran-Zivan, S., Jonas, E., & Baram-Tsabari, A. (2024). How to make sense of generative AI as a science communication researcher? A conceptual framework in the context of critical engagement with scientific information. *JCOM*, *23*, A05. https://doi.org/10.22323/2.23060205

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, *35*, 603–625. https://doi.org/10.1177/1075547012472684

König, L., & Jucks, R. (2019). When do information seekers trust scientific information? Insights from recipients' evaluations of online video lectures. *International Journal of Educational Technology in Higher Education*, *16*. https://doi.org/10.1186/s41239-019-0132-7

Könneker, C. (2024). Vertrauensbildende Wissenschaftskommunikation. *Wissenschaftskommunikation.de*. https://www.wissenschaftskommunikation.de/vertrauensbildende-wissenschaftskommunikation-74745/

Kulms, P., & Kopp, S. (2019). More human-likeness, more trust? The effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. *Proceedings of Mensch und Computer 2019*, 31–42. https://doi.org/10.1145/3340764.3340793

Kusters, R., Misevic, D., Berry, H., Cully, A., Le Cunff, Y., Dandoy, L., Díaz-Rodríguez, N., Ficher, M., Grizou, J., Othmani, A., Palpanas, T., Komorowski, M., Loiseau, P., Moulin Frier, C., Nanini, S., Quercia, D., Sebag, M., Soulié Fogelman, F., Taleb, S., … Wehbi, F. (2020). Interdisciplinary research in artificial intelligence: challenges and opportunities. *Frontiers in Big Data*, *3*. https://doi.org/10.3389/fdata.2020.577974

Light, A. E., Benson-Greenwald, T. M., & Diekman, A. B. (2022). Gender representation cues labels of hard and soft sciences. *Journal of Experimental Social Psychology*, *98*, 104234. https://doi.org/10.1016/j.jesp.2021.104234

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022). News from generative artificial intelligence is believed less. *2022 ACM Conference on Fairness, Accountability and Transparency*, 97–106. https://doi.org/10.1145/3531146.3533077

MacDorman, K. F. (2024). Does mind perception explain the uncanny valley? A meta-regression analysis and (de)humanization experiment. *Computers in Human Behavior: Artificial Humans*, *2*, 100065. https://doi.org/10.1016/j.chbah.2024.100065

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146*, 190–205. https://doi.org/10.1016/j.cognition.2015.09.019

MAITHINK X. (2018, July 12). *So habt ihr Farben noch nie gesehen* [Youtube video]. https://www.youtube.com/watch?v=r0jXfwPQW9k

Mancuso, K., Hauswirth, W. W., Li, Q., Connor, T. B., Kuchenbecker, J. A., Mauck, M. C., Neitz, J., & Neitz, M. (2009). Gene therapy for red-green colour blindness in adult primates. *Nature*, *461*, 784–787. https://doi.org/10.1038/nature08401

Maskedteller. (2023, April 30). *TikTok*. https://www.tiktok.com/@maskedteller

McDonnell, R., Breidt, M., & Bülthoff, H. H. (2012). Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, *31*, 1–11. https://doi.org/10.1145/2185520.2185587

Meckel, M., & Steinacker, L. (2024). *Alles überall auf einmal: Wie Künstliche Intelligenz unsere Welt verändert und was wir dabei gewinnen können*. Rowohlt Buchverlag.

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*, 98–100. https://doi.org/10.1109/mra.2012.2192811

Neuberger, C., Weingart, P., Schildhauer, T., Fähnrich, B., Wormer, H., Jarren, O., Passoth, J.-H., & Wagner, G. G. (2022). *Gute Wissenschaftskommunikation in der digitalen Welt: Politische, ökonomische, technische und regulatorische Rahmenbedingungen ihrer Qualitätssicherung*. Berlin-Brandenburg Academy of Sciences; Humanities. https://pure.mpg.de/rest/items/item_3390072/component/file_3390073/content

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*, e2120481119. https://doi.org/10.1073/pnas.2120481119

Reif, A., Kneisel, T., Schäfer, M., & Taddicken, M. (2020). Why are scientific experts perceived as trustworthy? Emotional assessment within TV and YouTube videos. *Media and Communication*, *8*, 191–205. https://doi.org/10.17645/mac.v8i1.2536

Ring, L., Utami, D., & Bickmore, T. (2014). The right agent for the job? The effects of agent visual appearance on task domain. In T. Bickmore, S. Marsella & C. Sidner (Eds.), *Intelligent virtual agents* (pp. 374–384). Springer International Publishing. https://doi.org/10.1007/978-3-319-09767-1_49

Rogers, Y., Sharp, H., & Preece, J. (2023). *Interaction design — beyond human-computer interaction*. John Wiley & Sons Inc.

Rosenthal-Von der Pütten, A. M., Krämer, N. C., Maderwald, S., Brand, M., & Grabenhorst, F. (2019). Neural mechanisms for accepting and rejecting artificial social partners in the uncanny valley. *Journal of Neuroscience*, *39*, 6555–6570. https://doi.org/10.1523/JNEUROSCI.2956-18.2019

Rühle, L. (2021). Zwei von fünf Verbrauchern weltweit sind besorgt wegen künstlicher Intelligenz. *YouGov*. https://yougov.de/technology/articles/39712-zwei-von-funf-verbrauchern-weltweit-sind-besorgt-w

Rutjens, B. T., & Većkalov, B. (2022). Conspiracy beliefs and science rejection. *Current Opinion in Psychology*, *46*, 101392. https://doi.org/10.1016/j.copsyc.2022.101392

Ruzi, S. A., Lee, N. M., & Smith, A. A. (2021). Testing how different narrative perspectives achieve communication objectives and goals in online natural science videos (S. Triberti, Ed.). *PLOS ONE*, *16*, e0257866. https://doi.org/10.1371/journal.pone.0257866

Schäfer, M. S. (2023). The notorious GPT: science communication in the age of artificial intelligence. *JCOM*, *22*, Y02. https://doi.org/10.22323/2.22020402

Schäfer, M. S., & Wessler, H. (2020). Öffentliche Kommunikation in Zeiten künstlicher Intelligenz: Warum und wie die Kommunikationswissenschaft Licht in die Black Box soziotechnischer Innovationen bringen sollte. *Publizistik*, *65*, 307–331. https://doi.org/10.1007/s11616-020-00592-6

Schwind, V., Wolf, K., & Henze, N. (2018). Avoiding the uncanny valley in virtual character design. *Interactions*, *25*, 45–49. https://doi.org/10.1145/3236673

Song, S. W., & Shin, M. (2024). Uncanny valley effects on chatbot trust, purchase intention and adoption intention in the context of E-commerce: the moderating role of avatar familiarity. *International Journal of Human-Computer Interaction*, *40*, 441–456. https://doi.org/10.1080/10447318.2022.2121038

Stavesand, M., & Schröder, K. (2024). TikTok und die Ära der KI-Avatare. *Wissenschaftskommunikation.de*. https://www.wissenschaftskommunikation.de/tiktok-und-die-aera-der-ki-avatare-74573/

Stein, J.-P., & MacDorman, K. F. (2024). After confronting one uncanny valley, another awaits. *Nature Reviews Electrical Engineering*, *1*, 276–277. https://doi.org/10.1038/s44287-024-00041-w

Stein, J.-P., Messingschlager, T., & Hutmacher, F. (2023). Künstliche Intelligenz. In M. Appel, F. Hutmacher, C. Mengelkamp, J.-P. Stein & S. Weber (Eds.), *Digital ist besser?! Psychologie der Online- und Mobilkommunikation* (pp. 247–260). Springer. https://doi.org/10.1007/978-3-662-66608-1_17

Swart, J. (2021). Experiencing algorithms: how young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society*, *7*. https://doi.org/10.1177/20563051211008828

Thaler, M., Schlogl, S., & Groth, A. (2020). Agent vs. avatar: comparing embodied conversational agents concerning characteristics of the uncanny valley. *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, 1–6. https://doi.org/10.1109/ichms49158.2020.9209539

Trench, B., & Bucchi, M. (2010). Science communication, an emerging discipline. *JCOM*, *09*, C03. https://doi.org/10.22323/2.09030303

UNESCO Institute for Statistics. (2019). *Women in Science*. https://uis.unesco.org/en/topic/women-science

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty and trust in news. *Social Media + Society*, *6*. https://doi.org/10.1177/2056305120903408

Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, *6*, 236–255. https://doi.org/10.1080/21670811.2017.1384319

Wissenschaft im Dialog. (2023). *Wissenschaftsbarometer. Eine repräsentative Bevölkerungsumfrage zu Wissenschaft und Forschung* [Science barometer: A representative survey on science and research]. https://wissenschaft-im-dialog.de/documents/47/WiD-Wissenschaftsbarometer2023_Broschuere_web.pdf

About the authors

Jasmin Baake is a research associate at the Center for Advanced Internet Studies (CAIS) in Bochum, Germany. She holds an MA in Communication Science from the University of Münster and is currently pursuing her PhD in Human-Computer Interaction. Her research focuses on using co-creative approaches to ensure that underrepresented groups, particularly working-class individuals, are actively included in the development of trustworthy AI systems.

jasmin.baake@cais-research.de @jasminbaake

Josephine B. Schmitt is the scientific coordinator at the Center for Advanced Internet Studies (CAIS), where she focuses on developing innovative concepts for interdisciplinary collaboration in digitalization research. Her research further examines the content, dissemination, and impact of online hate speech, extremist propaganda, and political information and educational content. To engage a broader community with her research, she does science communication in various formats.

Josephine.schmitt@cais-research.de @jottbees

Julia Metag is Professor at the Department of Communication at the University of Muenster. She earned her doctoral degree from the University of Muenster. Her research is centered on science communication and political communication, with a specific emphasis on the audience perspective and digital media environments. She is Co-PI of the Science Barometer Switzerland and currently Co-PI of a project on communication by private universities as well as PI in the DFG-funded research unit DISELMA — Digital Media in Chronic Disease Self-Management.

julia.metag@uni-muenster.de @juliametag

How to cite

Baake, J., Schmitt, J. B. and Metag, J. (2025). 'Balancing realism and trustworthiness: AI avatars in science communication'. *JCOM* 24(02), A03. https://doi.org/10.22323/2.24020203.

Supplementary material

Available at https://doi.org/10.22323/2.24020203

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