MUNI FSS

JOURNAL OF PSYCHOSOCIAL RESEARCH ON CYBERSPACE
CYBERPSYCHOLOGY

Long, J. A., Xiao, J., McCray, S. S., Ağaoğlu, E., Alajmi, A. M., Akalonu, C. P., & Xu, Y. (2026). Perceptions of AI-generated profile pictures: Effects on quality, attractiveness, and trustworthiness. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 20*(4), Article 3. https://doi.org/10.5817/CP2026-4-3

Perceptions of AI-Generated Profile Pictures: Effects on Quality, Attractiveness, and Trustworthiness

Jacob A. Long[1], Jingyi Xiao[1], Shamira S. McCray[1], Ertan Ağaoğlu[1], Abdullah M. Alajmi[1], Chinwendu P. Akalonu[1], & Yanzhen Xu[2]

[1] School of Journalism and Mass Communications, University of South Carolina, Columbia, SC, USA
[2] Bauer College of Business, University of Houston, Houston, Texas, USA

Abstract

This study investigates perceptions of AI-generated profile pictures and how disclosure of AI use shapes impressions. In a preregistered experiment, U.S. adults (N = 817; ages 18–99, M = 47.83, SD = 18.79; 48% men and 50% women) evaluated what they were told were social media profile pictures. The 2 × 2 factorial design varied whether images, all with a professional headshot style, were AI-generated and whether they were disclosed as AI-generated. Across three outcomes, AI-generated images were rated as higher quality than professional photos of the same subjects, but disclosure that an image was AI-generated reduced perceived quality. On average, AI generation and disclosure did not significantly affect perceptions of the person's social attractiveness or trustworthiness. Exploratory analyses indicated that self-assessed AI knowledge moderated responses to disclosure: lower-knowledge respondents downgraded labeled images and the depicted person, whereas higher-knowledge respondents evaluated labeled images and the person more positively. Taken together, the findings suggest that disclosure of AI use primarily affects evaluations of the image as an object rather than person-level impressions.

Keywords: generative AI; artificial intelligence; online self-presentation; social media; deception

Editorial Record

First submission received:
September 12, 2024

Revisions received:
August 20, 2025
February 11, 2026
May 26, 2026

Accepted for publication:
May 30, 2026

Editor in charge:
Lenka Dedkova

Introduction

Profile photos are salient cues for first impressions across social platforms. They signal identity and professionalism in contexts where quick judgments affect attention and credibility. Consumer-facing generative AI tools now let users produce headshot-like images that mimic professional photography at low or no cost. The use of these services raises a familiar question in computer-mediated impression formation: do medium and disclosure cues (whether an image was generated by AI and whether AI use is labeled) alter evaluations of the image and, in particular, of the person depicted?

Research on selective self-presentation shows that users curate their profiles to affect how they are perceived (Ellison et al., 2006; Hollenbaugh, 2021). Images matter because they convey social information efficiently (Bacev-Giles & Haji, 2017; van der Zanden et al., 2022). Users also know that people they encounter online can stage, select, and retouch images (Fox & Vendemia, 2016; Harris & Bardey, 2019). Warranting theory formalizes how people navigate these concerns by privileging information perceived as less controllable by the target (DeAndrea, 2014; Walther et al., 2009; Walther & Parks, 2002). A caption disclosing that a profile photo was generated by AI is

a process cue that could lower the image's perceived warranting value even when nothing visible in the image distinguishes it from a real photo. Whether such a penalty, if it exists, remains confined to the image or extends to person-level judgments is unclear.

Work on face-based first impressions suggests that medium-level and person-level evaluations can diverge. Trait inferences from faces arise quickly and are organized primarily by a general positive-negative dimension often labeled trustworthiness (Todorov, 2008; Todorov et al., 2008). Attractive faces tend to be evaluated more positively across traits (Dion et al., 1972), but these global impressions are driven mainly by facial features and expressions. Improving production value may make a headshot look more professional and increase perceived quality without necessarily shifting attractiveness or trustworthiness when the same person is depicted. Our question, therefore, is whether AI generation and its disclosure change evaluations at the level of the image and whether any such changes translate into person judgments. We examine these questions in a preregistered factorial experiment that varies an image's origin and its label, focusing on perceived photo quality, social attractiveness, and trustworthiness.

Related Work and Hypotheses

Online Self-Presentation, Authenticity, and AI-Generated Headshots

People curate content in their online profiles to shape how they are perceived when attention is scarce and information is thin (Ellison et al., 2006; Hollenbaugh, 2021). Photographs are consequential because they compress multiple judgments into an overall impression at a glance. Observers routinely make social inferences from a single portrait, even when little else is known (Bacev-Giles & Haji, 2017; van der Zanden et al., 2022). As profile images and avatars became central on social platforms, the use of high-quality headshots has become normative beyond explicitly professional networks. In personal profiles, such portraits can function as status signals and heuristics for competence and credibility. This makes production processes and labels consequential for how profile images are received.

The same affordances that make images useful for impression management also invite questions about authenticity. Users can stage settings, choose flattering angles, and use retouching tools to present more desirable versions of themselves; these practices can be seen as routine curation or, when they diverge too far from the offline self, as deception (Fox & Vendemia, 2016; Harris & Bardey, 2019; McCornack & Levine, 1990). In relational contexts such as online dating, discrepancies between online and offline selves have predictable consequences for perceived credibility and trust (Sharabi & Caughlin, 2019). Observers are aware of these possibilities and often attend to cues about how an image was produced and how much control the target exercised over the content. Evaluations of profile images therefore depend not only on their aesthetic appeal but also on the perceived legitimacy of the information they convey about the person depicted.

Warranting theory offers a means for understanding how receivers weigh such cues. The warranting principle holds that information perceived as less controllable by the target is given greater weight in shaping impressions because it is seen as more reliably connected to the offline self (Walther et al., 2009; Walther & Parks, 2002). Research in this area has developed the construct of warranting value, the (perceived) extent to which a target could have manipulated information (DeAndrea, 2014; DeAndrea & Carpenter, 2018). In the context of social media profiles, cues that appear to originate from others or from third parties can override self-claims when they conflict, suggesting that the source and controllability of information can be as important as its valence (Walther et al., 2009). By the same logic, a process disclosure that a profile photo was generated by AI is a salient warranting cue: it signals high target control over the production of the image. Even when the pixels are indistinguishable from a conventional photograph, such a disclosure could lower the image's warranting value and, in turn, reduce its influence on impressions. Whether any penalty remains confined to the image as an object or extends to judgments of the person is ultimately empirical and depends on how receivers integrate production cues with facial information.

Consumer-facing tools such as Remini offer a workflow in which users upload a small set of photos and receive a portfolio of synthesized headshot-like images at minimal cost (Hagy, 2023; Kudhail, 2023; Weiss, 2023). Conceptually, such images occupy the "maximally optimized" end of a spectrum that runs from casual selfies, to lightly edited images, to professional headshots, to fully AI-generated portraits. Movement along this spectrum generally increases production value while also increasing perceived user control, which can cause skepticism

about how realistic the image is. Contemporary systems can produce portraits that are highly realistic (Miller et al., 2023), making the production process itself, rather than obvious visual artifacts, the primary object of judgment for many observers. In this sense, a disclosure about the provenance of the image may loom larger as a cue than the image itself.

Empirical work on synthetic faces suggests that many, or most, people may struggle to distinguish AI-generated faces from real ones and may sometimes judge synthetic faces as more trustworthy. At the same time, accuracy and impressions vary across tasks and stimulus sets (Bozkir et al., 2025; Nightingale & Farid, 2022), along with the general challenge of researching a fast-improving set of technologies. More broadly, work on face-swapped videos and deepfake detection suggests that sensitivity to synthetic humans depends on modality and task demands (Somoray et al., 2025; Wöhler et al., 2021). These findings show why disclosure cues, rather than visible artifacts, may be especially consequential in profile-photo settings; the generative systems produce increasingly few tell-tale signs and people who use such systems are likely not to use outputs with obvious flaws. The findings also suggest that contemporary generators can produce portraits that meet or exceed viewers' expectations for visual quality, motivating our expectation that AI-generated headshots will be evaluated as higher in photo quality than conventional professional photographs. Unlike photography, which can improve appearance through composition, staging, lighting, and high-quality equipment, AI generation can emulate those advantages while being less constrained by the subject's actual appearance.

Much of what users do with AI-generated portraits resembles the optimization of more familiar self-presentation strategies more so than unambiguous deception. Users who might otherwise rely on casual, lower-production images adopt synthesized portraits to achieve a polished look that would otherwise be costly or unattainable. From a theoretical standpoint, comparing AI-generated portraits to professional photography therefore poses a conservative test. If generative tools can match or exceed the perceived production value of professional photos, any impact on person judgments, if present, would be more plausibly attributed to the production-process cue (i.e., the label and its warranting implications) than to differences in facial information or gross production deficits. This distinction between evaluations of the medium and judgments of the person is necessary to see how authenticity cues work in profile-image contexts and for informing policy debates about labeling synthetic media. Because the stimuli in the present study are professional-looking headshots, the inferences are most applicable to contexts where polished portraits are normative. We do not attempt to generalize to casual selfies or other aesthetics, which may involve somewhat different norms and expectations. By using actually AI-generated images, we are able to separate the effects of disclosure and whatever may or may not be distinctive about AI-generated imagery.

Prior work on professional networking sites indicates that profile photos play an outsized role in early credibility and competence judgments. On LinkedIn, the mere presence of a photo (vs. none) increases perceived competence and social attractiveness (Edwards et al., 2015), and specific visual cues such as smiling, eye contact, and attire shape recruiters' credibility assessments and interview intentions (van der Land et al., 2016). Context matters: when people choose profile images for professional contexts, selections accentuate competence and trust-related impressions more than attractiveness (White et al., 2017). Recruiters' judgments during LinkedIn screening are also vulnerable to biases activated by visual cues in profile photos, such as apparent age, which can spill over to job-suitability ratings (Schellaert et al., 2025). Building on this work, the present study focuses on a different but increasingly relevant cue in professionalized profile imagery: the production process. This lets us test whether process labels function as authenticity/warranting cues that primarily shift appraisals of the image (photo quality) versus person-level first impressions (social attractiveness, trustworthiness).

Related work on disclosure labels suggests small and context-dependent effects, in part because viewers already suspect some manipulation or fail to notice unobtrusive labels (Lewis et al., 2020; Naderer et al., 2022; Tiggemann & Brown, 2018). In profile contexts, an "AI-generated" label may be redundant for some perceivers but carry meaning for others. Our design orthogonally varies true image origin and disclosure to separate reactions to the medium from reactions to labeling and to test whether any penalties attach to the image, the person, or both. Audiences also differ in how they interpret an "AI-generated" label. Past experiences with AI and knowledge of how AI systems work may affect whether the label signals manipulability, familiarity, or shared orientation toward the use of AI. This possibility is consistent with other evidence that perceived similarity is associated with interpersonal attraction and positive evaluation (Montoya et al., 2008). Although preregistered tests focus on average effects, we probe this heterogeneity in exploratory analyses using a self-assessed AI knowledge measure. Taken together, this work suggests that AI headshots should raise perceived production value when the same identities are depicted across conditions. We therefore preregistered the following hypothesis[1]:

H1a: AI-generated profile pictures will be perceived as higher quality compared to real, professional photos depicting the same person.

Face-Based First Impressions and Their Downstream Effects

Trait inferences from faces arise quickly and with a high degree of consensus across observers. In laboratory settings, people form stable impressions after brief exposures to an unfamiliar face, often within a fraction of a second, and additional viewing time tends to increase confidence rather than change the evaluation itself (Willis & Todorov, 2006). These impressions are not arbitrary. A large literature shows that judgments cluster along two primary dimensions. The first reflects general valence, often labeled trustworthiness or warmth, and the second reflects power or dominance (Todorov, 2008; Todorov et al., 2008). When people evaluate faces in terms of specific traits, these ratings are highly correlated and can be positioned within this two-dimensional space. One robust observation is the attractiveness halo. More attractive faces are evaluated more positively on a range of social traits, including likability and perceived social skill, which suggests that observers generalize from a salient positive cue to a broader judgment of the person (Dion et al., 1972). This does not imply that all traits are interchangeable. Rather, it reflects the way people summarize limited information when there is little else to go on. In the context of profile photos, where a face is often the only available cue, it is unsurprising that attractiveness might covary with perceived trustworthiness and related traits. In keeping with this structure, we use the term social attractiveness to denote a composite of attractiveness, likability, and warmth, a global positive evaluation commonly applied under limited-information conditions.

The mechanisms that produce these first impressions depend primarily on the facial information itself. Features such as eye shape, mouth curvature, and the resemblance of neutral expressions to approach or avoidance cues drive evaluations on the valence dimension, while structural cues related to perceived strength support dominance evaluations (Todorov, 2008; Todorov et al., 2008). Dominance is an important part of the face-perception literature, but it is not the focus of this study. Our manipulations concern production method and disclosure, which are more directly tied to perceived authenticity, social appeal, and trust than to perceived power, strength, or threat. For that reason, the preregistered person-level outcomes focus on the broad evaluative side of first impressions. These inferences arise with minimal visual exploration and are relatively insensitive to minor changes in presentation. Production features such as lighting, pose, and background can change how professional an image appears, but they do not typically alter the morphology of the face, though they can emphasize certain features. When the same identities are depicted across conditions, differences in production value might be most apparent in evaluations of the image rather than in judgments of the person. On the other hand, AI image generators may make more profound changes to the appearance of the face in ways relevant to these evaluations without making the subject completely unrecognizable or otherwise looking plainly inauthentic.

At the same time, production features may serve as indirect social signals in some settings. A polished headshot can imply conscientiousness, care, or status, and those inferences could plausibly correlate with competence-related judgments. This is one reason our trustworthiness composite includes competence, dependability, and reliability. Even so, when the person depicted is held constant, theoretical expectations for sizable changes in social attractiveness or trustworthiness based solely on production value are modest. Larger person-level effects would be more likely if the facial content substantially changed alongside production features.

The context and framing of evaluation matter as well. When observers are asked to judge image quality, production features are the target of evaluation. When observers are asked to judge social attractiveness or trustworthiness, the face is the target of evaluation, and production features are background. This distinction aligns with the warranting perspective introduced earlier. A label about how an image was produced is a cue about the medium and its authenticity. It is not a cue that directly alters facial information. We might therefore expect process labels to bear more strongly on perceived image quality and the legitimacy of the photograph than on person-level judgments when the face is held constant across conditions.

Finally, it is important to recognize that even small differences in production value can matter in real-world contexts if they move images across informal thresholds of adequacy. A headshot that looks appropriate for professional use is different from one that appears casual or poorly lit, and such differences can have practical consequences. Our comparison between AI-generated portraits and professional photographs therefore constitutes a conservative test. Both images sit at the high end of production value, which raises the bar for detecting person-level changes. If effects on social attractiveness or trustworthiness are small or absent under these conditions, this is consistent with a face-perception account in which facial information carries most of the

signal for person judgments, while production features shape evaluations of the image as such. The structure of face-based evaluation implies that changes in production value should matter more for judgments of the image than for judgments of the person when facial information is held constant. At the same time, polished images can function as indirect social signals. Accordingly, we preregistered H2a and H3a: AI-generated profile photos will lead the person in the image to be perceived as more socially attractive than when the subject is depicted using a professional photo (**H2a**), and more trustworthy than when the subject is depicted using a professional photo (**H3a**). We expected this as a possible downstream consequence of subtle AI-generated changes to how the person appears in the photo.

Disclosure of AI Use and Algorithm Aversion

Research on disclosure of algorithmic involvement often documents a penalty for the labeled option, even when its objective performance is comparable to human alternatives. In forecasting and decision contexts, people tend to prefer human forecasters over algorithms once they observe errors, and they are slow to return to automated systems even when those systems perform well on average (Dietvorst & Bharti, 2020). In consumer and service settings, identifying an automated interlocutor can depress engagement or purchase behavior while holding the content of the interaction constant, as in studies of customer service chatbots (Luo et al., 2019) and service encounters (Mozafari et al., 2021). In managerial contexts, disclosure that feedback or screening was produced by AI can harm perceptions of fairness or willingness to follow recommendations relative to human sources (Keppeler, 2023; Tong et al., 2021). These findings are often interpreted as a form of algorithm aversion or disclosure penalty, where the label serves as a negative signal independent of output quality.

Results are more mixed in creative and aesthetic domains. Labels sometimes have small or negligible effects on appreciation of creative works, and in some cases audience responses depend on genre or framing. For instance, identifying an AI artist did not reduce appreciation of visual artworks in one study that manipulated perceived authorship while holding the image constant (Messingschlager & Appel, 2025). Evaluations of poetry produced by AI were similar to those of human-authored poems except in first-person emotional formats where readers discounted the AI label more strongly (Raj et al., 2026). In marketing, audiences sometimes exhibit mild negativity toward a disclosure that content was created by AI, yet still prefer the labeled content if it is objectively higher in quality than the alternatives (Zhang & Gosline, 2023). These cases suggest that disclosures do not produce a uniform and consistent penalty, and that the sign and size of the effect depend on how audiences interpret the relevance of the process cue. Recent research continues to document disclosure effects in platform contexts, including short-form video (Chen et al., 2025) where disclosure reduced perceptions of quality without hurting engagement. Related work also emphasizes that labeling and the assumptions about process that audiences attach to an "AI-generated" label are major inputs into their evaluations (Altay & Gilardi, 2024; Wittenberg et al., 2025).

Warranting theory can help to explain these results. Labels that signal high target control over content production are plausibly read as lowering the warranting value of the information, which reduces the weight that receivers give the content when forming impressions (DeAndrea, 2014; Walther et al., 2009; Walther & Parks, 2002). In profile-image contexts, a caption that a headshot was generated by AI indicates a process in which the target can strongly shape what is displayed, which could reduce the perceived authenticity of the image. When the image is the object of evaluation, this should show up as lower perceived quality or appropriateness. Whether such a penalty transfers to person-level judgments is less clear, but plausible given the nature of the evaluations people make in these contexts. In creative appraisals of artifacts such as paintings or poems, the warranting relevance of the production method is itself debatable, which may help explain small or null average effects in those settings.

Disclosure effects also depend on what receivers would have inferred without the label. As noted earlier, many observers already suspect some degree of curation or editing in profile photographs. In that case, labels can be redundant, and small average effects are unsurprising. The salience and specificity of labels matter as well. Prior work on disclaimers for digitally altered images finds that subtle labels often go unnoticed or are not processed as intended, which limits their impact (Lewis et al., 2020; Naderer et al., 2022; Tiggemann & Brown, 2018). Furthermore, if viewers already assume some manipulation in headshots, the incremental impact of a label may be confined to evaluations of the image rather than to judgments of the person.

From a warranting perspective, an "AI-generated" label identifies a process under high target control and should lower the warranting value of the image. We therefore preregistered **H1b**: disclosing that a profile photo was generated by AI will reduce perceived photo quality relative to disclosing a professional photographer. If penalties

carry over from the image to person judgments, disclosure should also reduce attractiveness and trustworthiness. We preregistered H2b and H3b: disclosing AI generation will lead the person in the image to be perceived as less socially attractive than when the profile photo is disclosed as taken by a professional photographer (**H2b**), and less trustworthy than when the profile photo is disclosed as taken by a professional photographer (**H3b**). The same warranting logic suggests that perceivers may treat AI disclosure as a cue that the image was optimized to improve the subject's appearance. Even when disclosed, that optimization could make the person seem less authentic or less trustworthy.

Methods

Preregistration

This study's hypotheses, design, and analysis were preregistered before data collection. There are no deviations from the preregistration to report. The preregistration is available at https://osf.io/b47d6/. Analysis scripts, data, stimuli, and other materials needed to reproduce the reported results are publicly available at https://osf.io/vtn63/.

Sample

Data for this study were collected from December 7, 2023 to December 14, 2023 via online survey with sampling and recruitment performed by Qualtrics. The study protocol was approved by the institutional review board of the University of South Carolina, USA (IRB protocol Pro00133423) and was conducted in accordance with the principles of the Declaration of Helsinki. The sample consisted of 817 U.S. adults with quotas to resemble the U.S. adult population along the lines of age, gender, and race/ethnicity. Participants were recruited from Qualtrics' online panel. The final sample identified as 48.2% men, 50.1% women, and fewer than 1% each identified as non-binary, preferred not to disclose, or used another self-description. On an item that allowed for multiple selections, 67% identified as white, 13% as Black, 18% as Hispanic/Latino, 5% as Asian, 4% as American Indian / Alaska Native, < 1% as Native Hawaiian / Pacific Islander, and 2% as another race. Participants ranged in age from 18 to 99 (M = 47.83, SD = 18.79) and education was 3.53 (Mdn = 3, SD = 1.64) on an item ranging from "Less than high school degree" (1) to "Doctoral degree" (8) with high school graduate coded as 3 and completed 4-year degrees coded as 5. On a measure where "less than $10,000" was coded as 1 and subsequent $10,000 increments taking on a digit higher code up to "$100,000 to $149,000" (11) and "more than $150,000" (12), average self-reported income was 5.28 (Mdn = 5, SD = 3.28), indicating in substantive terms an average between $40,000 and $49,999 annual household income.

Stimuli

To assess perceptions of images produced by generative AI, we compared AI-generated portraits with real photographs of the same subjects. To do so, we relied on the mobile app Remini, which promises users they can "transform your selfies into astonishingly realistic images using AI magic" (Bending Spoons, 2023). This app, which has over 100 million installs on the Google Play Store[2], allows users to upload photos of a person (usually oneself) and use generative AI to produce realistic photos depicting that person in posed portraits, in novel settings, in different outfits, with new hairstyles, and so on.

Note that this feature is distinct from other capabilities branded as AI both by Remini and other apps/services that use various machine learning techniques to retouch user-uploaded images. The AI generation in question and used for this study is not editing an existing photo, but rather training an AI model on images of a subject and generating completely new quasi-photographs. The appeal, at least for some users, is the ability to create images that appear to be genuine photographs but are beyond their personal ability/budget to create via actual photography (Kudhail, 2023).

Research using photos of people as stimuli generally requires the use of multiple variants of each stimulus, given the range of human appearances and factors that may influence perceptions of a person. With this in mind, we used 12 variants of each stimulus. There were 12 AI-generated photos used as stimuli and 12 real photos used as stimuli. To generate these, we gathered sets of stock photos portraying the same subject[3] at least 8 times, the minimum number of images required to train the Remini generative AI model. Images used for training can be

found in the online replication file. A goal in gathering photo subjects for use in stimuli was introducing variation in gender and apparent race/ethnicity to ensure findings are not restricted to an overly specific type of person. Ultimately, we arrived at 12 photo subjects who are depicted both in a real photograph and an AI-generated one. For each subject, care was taken to ensure the AI-generated image was roughly comparable in its composition, facial expression, and so on. That being said, there is no single visual factor that distinguishes AI-generated images from real photographs, so we did not endlessly generate images in service of creating a near-replica of the real photo version of the stimulus. Each pairing of real and AI-generated photo is shown in the Appendix A. The base photographs were stock photos and were not produced by the authors. The stock photos were sourced from Pexels, whose license permits reuse and modification of the images; we verified that this license permits use of the images in the AI-training workflow used here.

Because AI-generated portraits were produced by training on multiple photos of the same subject, generated images generally preserved the identity cues of the stock-photo model. The app produces multiple image outputs by default. Our choices for which to keep were guided by both internal validity concerns (similarity of composition, facial expression, and related aspects of photo relative to the designed real photos) and external validity concerns (whether the image appeared sufficiently realistic and similar to its real-world referent that an ordinary person might reasonably be expected to find it acceptable). We did not formally pretest whether, when shown both images, the study population would recognize them as depicting the same person. Finally, to allow for comparability across stimulus variants and to make the format similar to how most platforms display profile pictures, images were cropped into square aspect ratio, scaled down to a constant 500 pixels, and with the composition emphasizing the subject's face.

Procedure

In the present study's design, there are two manipulated independent variables. One is whether the images presented to participants are labeled as being AI-generated or captured by a professional photographer. The other is where those images actually came from: a generative AI model as described previously or a professional photographer. The study used a 2 × 2 factorial design with repeated measures. Participants were randomly assigned, between subjects, to conditions defined by two factors: whether the image was AI-generated (versus a real professional photograph) and whether the accompanying caption disclosed AI generation (versus crediting a photography agency). Participants were informed that the study was designed to assess how people perceive social media profile pictures.

In each of three trials, participants were shown an image consistent with their assigned condition. The order of these three images was randomized[4] for each participant. The photo was accompanied with a caption that stated either "This image was generated using an artificial intelligence tool based on real photos of the person in the image" or "This photo was taken by Stunning Headshots LLC." The factorial design means that captions were misleading for some participants, allowing for analyses that distinguish effects arising from the image itself versus the belief about its origin. After viewing the image and caption, respondents rated the photo's quality and their perceptions of the person's social attractiveness and trustworthiness. After completing all three trials, participants provided demographic information. Captions were presented as part of the survey materials and were not attributed to the profile owner, the platform, or an independent third party; this framing avoids making deception about the person a salient confound of AI disclosure.

If the stimuli variants indeed differ in the magnitude of their effects, then they would harm statistical power. To alleviate this, in line with best practice recommendations (Slater et al., 2015), we use a repeated measures design. To balance the desire to have participants evaluate multiple photos with the need for a concise survey, participants evaluate 3 of the possible 12 photos. The experimental condition for each participant remains the same across the 3 repetitions; they are simply seeing multiple variants of the stimulus. The variants are selected at random within the condition. We did not conduct separate pretests to equate stimuli on perceived traits. Instead, we (a) sampled across 12 distinct subjects to capture variability in appearance; (b) paired each AI-generated image with a professional photograph of the same subject and broadly similar composition; and (c) treated stimuli as sampled in the analysis by including crossed random effects for photo subject and participant. This strategy reduces reliance on any single stimulus and supports generalization across the sampled identities.

To document the success of randomization, we conducted balance checks across the four experimental cells on participant age, education, income, gender, and race/ethnicity indicators, as well as the self-assessed AI

knowledge item used in exploratory moderation models. Differences were small: omnibus ANOVAs for age, education, and income were not statistically significant after Holm adjustment ($F(3, 813) \leq 1.03$, $p = 1.000$, $\eta^2 \leq .004$). AI knowledge also did not differ after adjustment ($F(3, 799) = 3.40$, Holm-adjusted $p = .174$, $\eta^2 = .013$); this item was measured after the main task. Chi-squared tests indicated no detectable differences in gender and key race/ethnicity indicators ($\chi^2(3) \leq 4.98$, Holm-adjusted $p = 1.000$; Cramer's $V \leq .078$). Of course, any differences observed would be attributable to random variation. Table 1 reports descriptives by experimental cell.

Table 1. *Sample Descriptives by Experimental Condition.*

Condition	*n*	Age *M* (*SD*)	Men	Women	White	Black	Latino	Asian	AI knwl. *M* (*SD*)
Human photo + human label	185	46.43 (18.10)	49.7%	49.2%	67.6%	13.0%	18.4%	5.9%	3.53 (1.80)
AI photo + human label	211	48.73 (18.71)	49.3%	49.8%	62.6%	11.8%	19.0%	7.6%	3.44 (1.92)
Human photo + AI label	202	47.53 (19.17)	47.0%	50.0%	68.3%	13.4%	16.8%	3.5%	3.18 (1.80)
AI photo + AI label	219	48.42 (19.14)	47.0%	51.1%	70.8%	12.3%	18.3%	3.7%	3.75 (1.89)

Note. AI knowledge was measured after the main task.

Measures

Unless otherwise noted, measures were collected after each image trial.

Perceived Photo Quality

Photo quality items were developed for this study to capture perceived production value and appropriateness of profile images in professionalized contexts, drawing on Mortensen et al.'s (2023) photo-credibility scale development work. Participants were presented with statements about the photo and reported their agreement on a 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*). The questionnaire included six photo-evaluation statements: *This is a good photo*; *A skilled photographer took this photo*; *This photo is appropriate for professional settings*; *This photo is an accurate representation of the person in it*; *If the person in the photo was a friend of mine, I would encourage them to use it for their social media profiles*; and *This photo looks like it has been edited*. For the preregistered primary outcome, we operationalized perceived photo quality as the average of four items that focus on global quality and appropriateness for profile use (excluding the *skilled photographer* item and the *looks edited* item, which are more directly tied to the manipulations and to perceived authenticity, respectively).

The resulting measure had a mean of 5.27 (*SD* = 1.27), and reliability was acceptable (ω = .79), as assessed by McDonald's Omega to account for the repeated measures design (Geldhof et al., 2014). As a robustness check, the full six-item set also showed acceptable reliability (ω = .74 within-person; ω = .88 between-person) and yielded the same substantive conclusions in hypothesis tests (Table E2). As additional evidence of structural coherence, one-factor CFAs of the photo-quality items suggested good fit for the preregistered four-item factor ($\chi^2(2) = 10.41$, $p = .005$, CFI = .998, TLI = .994, RMSEA = .042, 90% CI = .019 to .069, SRMR = .007) and acceptable but inferior fit for broader five- and six-item specifications (five items: $\chi^2(5) = 40.07$, $p < .001$, CFI = .994, TLI = .988, RMSEA = .054, 90% CI = .039 to .070, SRMR = .012; six items: $\chi^2(9) = 106.14$, $p < .001$, CFI = .983, TLI = .972, RMSEA = .067, 90% CI = .056 to .079, SRMR = .025).

Person Perceptions

Participants assessed the person depicted in the photo with the prompt *I think the person in the photo is…* followed by a series of adjectives. Response choices ranged from *Not at all* (coded as 1) to *Extremely* (9), with the numerical codes made explicit to respondents. This intensity-scale format allows respondents to express negative or neutral impressions via the lower end of the scale without requiring explicitly negative adjective wording. The trait-adjective format follows common face-perception paradigms in which participants rate targets on brief trait descriptors after minimal exposure (Todorov & Porter, 2014). The adjective items were not intended to introduce a new standardized scale. Rather, the author team selected brief trait adjectives before data collection to adapt face-perception rating procedures to the theoretical concerns of this study: general affiliative evaluation and trust-

related judgments that could plausibly be affected by perceived deception or warranting value. We therefore report item content, reliability, and CFA-based structure checks to support their use in this context.

Social Attractiveness

For social attractiveness, the responses for the adjectives *attractive*, *likable*, and *warm or sympathetic* are averaged (M = 6.34, SD = 1.85). McDonald's Omega reliability for this measure is acceptable (ω = .72).

Trustworthiness

For trustworthiness, responses for five other adjectives—*competent*, *trustworthy*, *honest*, *dependable*, and *reliable*—are averaged (M = 6.25, SD = 1.79). McDonald's Omega reliability is acceptable (ω = .84).

In research on social perception (e.g., the stereotype content model), warmth/trustworthiness and competence are treated as distinct dimensions (Fiske, 2018; Fiske et al., 2002). In face-based first-impression research, trait ratings are typically correlated to a very high degree under limited-information conditions (Todorov, 2008). Given the ambiguity and that our measurements do not map cleanly onto the social perception research framework, we therefore report confirmatory factor analyses (CFA) and robustness checks using alternative factor structures. We retain our preregistered composites for primary hypothesis tests, but we assessed the measurement structure using CFA. The two composites should therefore be read as preregistered, theoretically motivated outcomes rather than as evidence that social attractiveness and trustworthiness are cleanly separable constructs in this kind of minimal-information task.

For the five-item trustworthiness composite, a two-factor model that separates competence-related adjectives (competent, dependable, reliable) from integrity-related adjectives (trustworthy, honest) fit slightly better than a one-factor model (one factor: $\chi^2(5) = 73.70$, $p < .001$, CFI = .994, TLI = .988, RMSEA = .077, 90% CI = .062 to .093, SRMR = .008; two factors: $\chi^2(4) = 11.93$, $p = .018$, CFI = .999, TLI = .998, RMSEA = .029, 90% CI = .011 to 0.049, SRMR = .003), but the two latent factors were extremely highly correlated (r = .97). In a broader CFA of all eight person-perception adjectives, a three-factor model (social attractiveness, competence, integrity) fit slightly better than the preregistered two-factor structure (three factors: $\chi^2(17) = 103.19$, $p < .001$, CFI = .995, TLI = .993, RMSEA = .047, 90% CI = .038 to .056, SRMR = .009; two factors: $\chi^2(19) = 173.49$, $p < .001$, CFI = .992, TLI = .988, RMSEA = .059, 90% CI = .051 to .067, SRMR = .011), but the latent correlations again indicated limited discriminant separation among the constructs (r = .94–.97). Consistent with these results, robustness checks using separate competence and integrity composites yield the same substantive conclusion as the preregistered person-level outcomes (Appendix E).

AI Knowledge

After the main experimental task, participants reported their agreement (M = 3.48, SD = 1.86) on a 7-point scale with the statement, *I know a lot about how AI technologies work*. This item captures their self-perception rather than objective skill or knowledge.

Analytic Strategy

As specified in the preregistration, hypotheses were tested with multilevel linear regression models to account for clustering from the repeated-measures design and stimulus variability. Models included fixed effects for AI generation, caption disclosure (both effect-coded at −0.5/0.5), and their interaction, with crossed random intercepts for participant and photo subject. There are no hypotheses about interactions, but the term is included due to the factorial design and to rule out any unanticipated interactions. This is preferred to repeated-measures ANOVA[5], which is a special case of this analytic approach, because it explicitly models stimulus-level variance and accommodates unbalanced data (Judd et al., 2012). Models are estimated using the "lme4" package for R (Bates et al., 2015) with p values calculated using the Satterthwaite method as implemented in the "jtools" and "lmerTest" R packages (Kuznetsova et al., 2017; Long, 2024). We additionally report an exploratory analysis of whether the effect of manipulated variables depends on familiarity with AI.

To benchmark statistical power, we conducted simulation-based design sensitivity analyses (with R package simr) using the observed variance components from the mixed-effects models. We did not conduct an a priori power

analysis for this mixed-effects design. Instead, we report design sensitivity benchmarks using the fitted variance components to contextualize detectable effect sizes. With this design, main effects of approximately 0.20 points (on the 1–7 photo-quality scale) are detectable with roughly 80% power; slightly larger effects (approximately 0.30) are required for person-level outcomes, reflecting greater between-person variability. Full details are reported in the Appendix D.

Results

We report results in the order of the preregistered hypotheses: H1 (photo quality), H2 (social attractiveness), and H3 (trustworthiness), followed by exploratory analyses. There are 3 models, one with each of the key variables as outcome: perceived photo quality, social attractiveness, and trustworthiness. Table 2 reports details for the preregistered mixed-effects models for each outcome. Participants contributed three repeated ratings each. Coefficients are unstandardized regression coefficients (*b*).

Table 2. *Multilevel Regression Results for Each Dependent Variable.*

	Quality		Social Attractiveness		Trustworthiness	
	Coef. (*SE*)	*p* value	Coef. (*SE*)	*p* value	Coef. (*SE*)	*p* value
Fixed effects						
AI-generated image	0.206 (0.073)	.005	0.112 (0.111)	.313	0.117 (0.114)	.306
AI disclosure	−0.149 (0.073)	.041	−0.020 (0.111)	.856	−0.051 (0.114)	.655
Disclosure x image	0.105 (0.145)	.468	0.339 (0.223)	.129	0.429 (0.228)	.060
Intercept	5.269 (0.083)	< .001	6.337 (0.107)	< .001	6.250 (0.075)	< .001
Random effects						
SD Respondent	0.91		1.46		1.54	
SD Photo subject	0.26		0.32		0.17	
SD Residual	0.85		1.09		0.91	
Model information						
N	2451		2446		2451	
Pseudo-R^2	.56		.65		.75	
AIC	7412.51		8935.42		8368.82	
BIC	7453.14		8976.03		8409.45	

Note. Values for fixed effects are unstandardized regression coefficients with standard errors in parentheses. Random effect variances refer to the intercept term and are reported as standard deviations. Pseudo-R^2 are calculated using the Nakagawa and Schielzeth (2013) method.

Photo Quality (H1a, H1b)

Consistent with H1a, AI-generated images were rated as higher quality than professional photographs of the same subjects (*b* = 0.206, SE = 0.073, *p* = .005). Consistent with H1b, disclosure that an image was AI-generated lowered perceived quality (*b* = −0.149, SE = 0.073, *p* = .041). The interaction between generation and disclosure was not statistically significant (*p* = .468).

Social Attractiveness (H2a, H2b)

Neither AI generation nor disclosure produced a statistically significant change in perceived social attractiveness (AI-generated image: *b* = 0.112, *SE* = 0.111, *p* = .313; AI disclosure: *b* = −0.020, *SE* = 0.111, *p* = .856), contrary to H2a and H2b. The interaction was not statistically significant (*p* = .129).

Trustworthiness (H3a, H3b)

Neither AI generation nor disclosure produced a statistically significant change in perceived trustworthiness (AI-generated image: *b* = 0.117, *SE* = 0.114, *p* = .306; AI disclosure: *b* = −0.051, *SE* = 0.114, *p* = .655), contrary to H3a and H3b. The interaction was not statistically significant (*p* = .060).

Given the high correlations among the person-perception latent constructs, the results for H2 and H3 are best interpreted as preregistered tests of two theoretically motivated ways of summarizing a broader evaluation of the person depicted. They do not provide evidence that AI generation or disclosure thereof reliably changed either component of that broad person-level evaluation, on average.

Exploratory Analyses

As an exploratory analysis, we re-estimated each of the three models including the AI knowledge variable as a covariate and interacting with the experimental conditions. These analyses indicated that AI knowledge consistently moderated responses to disclosure across all three outcomes (all $p < .01$). Among respondents lower in AI knowledge, disclosure of AI usage was associated with lower evaluations of the image and lower person-level judgments. Among respondents higher in AI knowledge, the same label was associated with more positive evaluations of the image and of the person when the image was generated by AI. In other words, those who feel they understand AI systems evaluate AI images and those who use them rather positively while those who do not rate them more negatively. This pattern should be interpreted cautiously, but it suggests that disclosure may carry different social meanings depending on whether AI use feels familiar or distant to the observer. Full details and moderation plots are included in the Appendix B and Appendix C. Additional exploratory analyses of potential moderation by demographic variables yielded consistently null results and are not reported further.

Discussion

In a preregistered 2 × 2 factorial experiment with repeated measures, U.S. adults evaluated three professional-style profile photos that varied by (a) whether the image was AI-generated and (b) whether the caption disclosed AI generation (versus crediting a professional photographer). We tested preregistered hypotheses using mixed-effects regression models with crossed random intercepts for participants and photo subjects. AI-generated images were rated as slightly higher in photo quality than professional photographs of the same identities, whereas an "AI-generated" caption modestly reduced perceived quality. On average, neither manipulation reliably shifted social attractiveness or trustworthiness judgments of the person depicted.

Main Findings

The overall pattern of results is a disconnect between evaluations of the image and evaluations of the person, at least in response to the experimental manipulations. AI generation increased perceived photo quality relative to professional photographs, but disclosure partially offsets this gain. Because there was no detectable interaction between generation and disclosure, these effects combine additively, leaving AI images labeled as AI with a modest net advantage in photo quality over professional photos, whether they are labeled as such or not. This pattern resembles prior work in which an "AI" label reduces evaluations but does not eliminate preferences for higher-quality outputs (Zhang & Gosline, 2023). For person-level impressions, the average effects were small and not statistically distinguishable from zero for both social attractiveness and trustworthiness. Holding identity constant across conditions likely constrained movement in these judgments, which aligns with face-perception accounts that emphasize facial appearance as the primary driver of trait inferences under limited information. This also implies that the AI generation did not alter the subjects' appearance too fundamentally despite making the image perceptibly better according to the tastes of the study participants.

A warranting perspective helps in the interpretation of these findings. A process label that an image was generated by AI can lower perceived warranting value by signaling high target control over what is shown. When the judgment target is the image, production cues are directly relevant and a disclosure penalty is expected. When the judgment target is the person, the appearance of the person is primary and the same disclosure cue may not be at top of mind, especially when identity is held constant across conditions. This helps explain why the disclosure effect appeared clearly for photo quality but not, on average, for social attractiveness or trustworthiness. It also connects this study to broader work showing that AI labels can reduce evaluations without fully overriding the perceived quality of the object being evaluated (Altay & Gilardi, 2024; Raj et al., 2026; Zhang & Gosline, 2023).

AI Knowledge and Audience Heterogeneity

The exploratory AI knowledge results suggest that disclosure effects may depend on how receivers interpret the social meaning of AI use. Respondents lower in self-assessed AI knowledge tended to penalize labeled images and the person depicted, whereas respondents higher in AI knowledge evaluated labeled AI images and the person somewhat more positively. One possible explanation is that lower-knowledge respondents treat an AI label mainly as evidence of manipulation, whereas higher-knowledge respondents treat the same label as more familiar or normative. This interpretation does not depend on respondents seeing AI image generation as technically demanding. Instead, higher-knowledge respondents may have treated AI use as more familiar, acceptable, or similar to their own practices. It therefore may reflect perceived similarity: people who see AI use as familiar may respond more favorably to someone who appears to use it, consistent with evidence that perceived similarity is associated with attraction and positive interpersonal evaluation (Montoya et al., 2008).

This interpretation is speculative. The AI knowledge item was self-assessed, measured after the main task, and did not distinguish between objective knowledge, frequency of AI use, attitudes toward AI, or identification with people who use AI tools. We also did not measure perceived similarity directly. The moderation observation points to a useful boundary condition, but future work should distinguish preregistered measures of AI familiarity, AI attitudes, and identification with AI users from post-exposure perceptions of similarity to the target. Doing so would help determine whether the effect reflects knowledge, comfort with AI, social identification with AI users, or some combination of these.

Measurement Considerations

Although the stereotype content model (Fiske, 2018; Fiske et al., 2002) distinguishes warmth and competence, our CFAs suggest little discrimination among the adjective ratings in this context. A two-factor model of competence-related versus integrity-related adjectives fit slightly better than a one-factor model, but the latent factors were nearly perfectly correlated. We interpret this as evidence that, under brief exposure and minimal context, participants' trait ratings largely reflect a broad evaluation rather than clearly differentiated subconstructs. This is compatible with the stereotype content model, but it suggests that this particular task compressed distinctions that may be more separable in richer communication settings.

The same point applies to the separation between social attractiveness and trustworthiness. We retained these composites because they were preregistered and because the warranting approach made trust-related judgments theoretically important. At the same time, the CFA results suggest they should not be interpreted as if they are obviously distinct latent variables in this study. One way to interpret these findings is that the study tested whether AI generation and disclosure affected the image itself and whether those changes transferred to two summaries of warmth- and valence-related person evaluation. There were no plausible alternate specifications that changed the null experimental results on the person perceptions.

Dominance was not included as an outcome, even though it is an important part of face-perception models. This reflects the scope of the present study rather than a claim that dominance is unimportant to face perception. The study was designed around disclosure and warranting value, which led us to focus on quality, social attractiveness, and trust rather than (for example) perceived power or threat. Because dominance was not measured, we cannot determine whether subtle appearance changes introduced by AI generation affected related perceptions. Future work could extend this design by measuring dominance directly, especially in settings where pose, status cues, facial maturity, or threat-related cues are central to the evaluation.

Limitations and Future Research

This experiment isolates process cues by holding identity constant, but the AI-vs-professional comparison is necessarily dependent on the specific tool, training workflow, and stimulus set. Future work should test whether the same dissociation between image appraisals and person impressions holds when AI images depart more visibly from the target's real appearance or when viewers have stronger viewpoints about what "AI-generated" implies. Disclosure in this study was presented as part of the survey materials and not as a platform label or self-disclosure by the profile owner. Varying the source, timing, and certainty of disclosure (e.g., platform label vs. self-disclosure; simultaneous vs. delayed) would help clarify which forms of labeling transfer from the medium to person-level impressions. At present, it is likely there is substantial use of AI tools for impression management

on online social platforms but it is probably not often disclosed; effects may well be different if users find out AI was used, but not disclosed (or actively concealed).

To contextualize small and null effects, we report simulation-based design sensitivity benchmarks rather than observed post hoc power. With this design, small main effects on photo quality are detectable, whereas somewhat larger (but still substantively small) effects would be required to move person-level outcomes.

Practical Implications

In the context of professionally-oriented profile pictures, the present results suggest that generative AI can improve perceived image quality without substantially shifting first impressions of the person on average. Those who have followed the social trends of using AI in lieu of professional photography (e.g., Kudhail, 2023) may in fact not be facing any social penalty for doing so. At the same time, disclosure may carry different meanings for different audiences, which implies that both self-presentation strategies and labeling policies are likely to have heterogeneous effects. As both social norms around the use of AI and technical capabilities change, it is possible that reactions to such usage change as well.

Conclusion

In a preregistered 2 × 2 experiment, AI-generated headshots were evaluated as slightly higher in photo quality than professional photographs of the same identities, but an "AI-generated" label modestly reduced perceived quality. Person-level impressions (social attractiveness, trustworthiness) were not reliably affected by either manipulation, suggesting that process labels primarily shape evaluations of the image rather than first impressions of the depicted person when identity is held constant. Exploratory analyses indicated that self-assessed AI knowledge qualifies responses to disclosure, pointing to audience heterogeneity as an important boundary condition that deserves more investigation. Future work should vary the source and timing of disclosure and expand beyond professionalized headshots to clarify when labeling cues transfer to person-level impressions.

Footnotes

[1] These hypotheses are reworded relative to the preregistration to better fit the context of the literature review. Their substantive meanings are unchanged.

[2] Apple's App Store does not disclose usage statistics of this kind.

[3] Although one might usually call the people who appear in the stock photos "models," we avoid this term here due to our extensive discussion of the statistical models that power AI services.

[4] As a robustness check, we attached a trial order variable (1–3) to each record; including order as a covariate did not change inferences nor did interacting it with the experimental variables.

[5] In the case of the present study, RM-ANOVA yields near-identical inferential results.

Conflict of Interest

The authors have no conflicts of interest to declare.

Use of AI Services

The authors used the generative AI application Remini to generate the experimental image stimuli, as described in the Methods section. OpenAI's Codex was used to check the manuscript and analytic scripts for errors and omissions. The authors take full responsibility for the content of the article.

Data Availability Statement

Analysis scripts, data, stimuli, and other materials needed to reproduce the reported results are publicly available at https://osf.io/vtn63/. The preregistration is available at https://osf.io/b47d6/.

Authors' Contribution

Jacob A. Long: conceptualization, data curation, formal analysis, funding acquisition, methodology, project administration, supervision, writing—original draft, writing—review & editing. **Jingyi Xiao:** conceptualization, methodology, writing—original draft, writing—review & editing. **Shamira S. McCray:** conceptualization, methodology, writing—original draft, writing—review & editing. **Ertan Ağaoğlu:** conceptualization, methodology, writing—original draft, writing—review & editing. **Abdullah M. Alajmi:** conceptualization, methodology, writing—original draft, writing—review & editing. **Chinwendu P. Akalonu:** conceptualization, methodology, writing—original draft, writing—review & editing. **Yanzhen Xu:** conceptualization, methodology, writing—original draft, writing—review & editing.

Jingyi Xiao, Shamira S. McCray, Ertan Ağaoğlu, Abdullah M. Alajmi, Chinwendu P. Akalonu, and Yanzhen Xu contributed equally, and their order in the author list was determined by random draw. All authors approved the final version of the article.

Acknowledgement

The authors acknowledge the graduate program of the School of Journalism and Mass Communications at the University of South Carolina for its funding of the research.

References

Altay, S., & Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, *3*(10), Article pgae403. https://doi.org/10.1093/pnasnexus/pgae403

Bacev-Giles, C., & Haji, R. (2017). Online first impressions: Person perception in social media profiles. *Computers in Human Behavior*, *75*, 50–57. https://doi.org/10.1016/j.chb.2017.04.056

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bending Spoons. (2023). *Remini – AI photo enhancer* [Computer software]. https://play.google.com/store/apps/details?id=com.bigwinepot.nwdn.international&hl=en

Bozkir, E., Riedmiller, C., Skodras, A. N., Kasneci, G., & Kasneci, E. (2025). Can you tell real from fake face images? Perception of computer-generated faces by humans. *ACM Transactions on Applied Perception*, *22*(2), Article 6. https://doi.org/10.1145/3696667

Chen, H., Wang, P., & Hao, S. (2025). AI in the spotlight: The impact of artificial intelligence disclosure on user engagement in short-form videos. *Computers in Human Behavior*, *162*, Article 108448. https://doi.org/10.1016/j.chb.2024.108448

DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, *24*(2), 186–204. https://doi.org/10.1111/comt.12033

DeAndrea, D. C., & Carpenter, C. J. (2018). Measuring the construct of warranting value and testing warranting theory. *Communication Research*, *45*(8), 1193–1215. https://doi.org/10.1177/0093650216644022

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. https://doi.org/10.1037/h0033731

Edwards, C., Stoll, B., Faculak, N., & Karman, S. (2015). Social presence on LinkedIn: Perceived credibility and interpersonal attractiveness based on user profile picture. *Online Journal of Communication and Media Technologies*, *5*(4), 102–115. https://doi.org/10.29333/ojcmt/2528

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, *11*(2), 415–441. https://doi.org/10.1111/j.1083-6101.2006.00020.x

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fox, J., & Vendemia, M. A. (2016). Selective self-presentation and social comparison through photographs on social networking sites. *Cyberpsychology, Behavior, and Social Networking*, *19*(10), 593–600. https://doi.org/10.1089/cyber.2016.0248

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Hagy, P. (2023, July 19). *Gen Zers are using apps like Remini and Canva to turn selfies into A.I.-generated professional headshots while saving big money*. Fortune. https://fortune.com/2023/07/19/tiktok-genz-millennials-generative-ai-headshots-remini-canva/

Harris, E., & Bardey, A. C. (2019). Do Instagram profiles accurately portray personality? An investigation into idealized online self-presentation. *Frontiers in Psychology*, *10,* Article 871. https://doi.org/10.3389/fpsyg.2019.00871

Hollenbaugh, E. E. (2021). Self-presentation in social media: Review and research opportunities. *Review of Communication Research*, *9*, 80–98. https://doi.org/10.12840/ISSN.2255-4165.027

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Keppeler, F. (2023). No thanks, dear AI! Understanding the effects of disclosure and deployment of artificial intelligence in public sector recruitment. *Journal of Public Administration Research and Theory*, *34*(1), 39–52. https://doi.org/10.1093/jopart/muad009

Kudhail, P. (2023, October 12). *Could an AI-created profile picture help you get a job?* BBC News. https://www.bbc.com/news/business-67054382

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lewis, N., Pelled, A., & Tal-Or, N. (2020). The effect of exposure to thin models and digital modification disclaimers on women's body satisfaction. *International Journal of Psychology*, *55*(2), 245–254. https://doi.org/10.1002/ijop.12572

Long, J. A. (2024). jtools: Analysis and presentation of social scientific data. *The Journal of Open Source Software*, *9*(101), Article 6610. https://doi.org/10.21105/joss.06610

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

McCornack, S. A., & Levine, T. R. (1990). When lies are uncovered: Emotional and relational outcomes of discovered deception. *Communication Monographs*, *57*(2), 119–138. https://doi.org/10.1080/03637759009376190

Messingschlager, T. V., & Appel, M. (2025). Mind ascribed to AI and the appreciation of AI-generated art. *New Media & Society*, *27*(3), 1673–1692. https://doi.org/10.1177/14614448231200248

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A. M., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science*, *34*(12), 1390–1403. https://doi.org/10.1177/09567976231207095

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889–922. https://doi.org/10.1177/0265407508096700

Mortensen, T. M., McDermott, B. P., & Ejaz, K. (2023). Measuring photo credibility in journalistic contexts: Scale development and application to staff and stock photography. *Journalism Practice*, *17*(6), 1158–1177. https://doi.org/10.1080/17512786.2021.1976073

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2021). Trust me, I'm a bot — repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management*, *33*(2), 221–245. https://doi.org/10.1108/JOSM-10-2020-0380

Naderer, B., Peter, C., & Karsay, K. (2022). This picture does not portray reality: Developing and testing a disclaimer for digitally enhanced pictures on social media appropriate for Austrian tweens and teens. *Journal of Children and Media*, *16*(2), 149–167. https://doi.org/10.1080/17482798.2021.1938619

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

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*(8), Article e2120481119. https://doi.org/10.1073/pnas.2120481119

Raj, M., Berg, J. M., & Seamans, R. (2026). The artificial intelligence disclosure penalty: Humans persistently devalue AI-generated creative writing. *Journal of Experimental Psychology: General*, *155*(4), 896–915. https://doi.org/10.1037/xge0001889

Schellaert, M., Oostrom, J. K., & Derous, E. (2025). Ageism on LinkedIn: Discrimination towards older applicants during LinkedIn screening. *Computers in Human Behavior*, *162*, Article 108430. https://doi.org/10.1016/j.chb.2024.108430

Sharabi, L. L., & Caughlin, J. P. (2019). Deception in online dating: Significance and implications for the first offline date. *New Media & Society*, *21*(1), 229–247. https://doi.org/10.1177/1461444818792425

Slater, D. M., Peter, J., & Valkenburg, P. M. (2015). Message variability and heterogeneity: A core challenge for communication research. *Annals of the International Communication Association*, *39*(1), 3–31. https://doi.org/10.1080/23808985.2015.11679170

Somoray, K., Miller, D. J., & Holmes, M. (2025). Human performance in deepfake detection: A systematic review. *Human Behavior and Emerging Technologies*, *2025*(1), Article 1833228. https://doi.org/10.1155/hbe2/1833228

Tiggemann, M., & Brown, Z. (2018). Labelling fashion magazine advertisements: Effectiveness of different label formats on social comparison and body dissatisfaction. *Body Image*, *25*, 97–102. https://doi.org/10.1016/j.bodyim.2018.02.010

Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, *1124*(1), 208–224. https://doi.org/10.1196/annals.1440.012

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404–1417. https://doi.org/10.1177/0956797614532474

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, *42*(9), 1600–1631. https://doi.org/10.1002/smj.3322

van der Land, S. F., Willemsen, L. M., & Wilton, B. G. E. (2016). Professional personal branding: Using a "think-aloud" protocol to investigate how recruiters judge LinkedIn profile pictures. In F. F.-H. Nah & C.-H. Tan (Eds.), *HCI in business, government, and organizations: Ecommerce and innovation* (pp. 118–128). Springer. https://doi.org/10.1007/978-3-319-39396-4_11

van der Zanden, T., Mos, M. B. J., Schouten, A. P., & Krahmer, E. J. (2022). What people look at in multimodal online dating profiles: How pictorial and textual cues affect impression formation. *Communication Research*, *49*(6), 863–890. https://doi.org/10.1177/0093650221995316

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). Sage.

Walther, J. B., Van Der Heide, B., Hamel, L. M., & Shulman, H. C. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using Facebook. *Communication Research*, *36*(2), 229–253. https://doi.org/10.1177/0093650208330251

Weiss, G. (2023, July 18). *TikTokers are flocking to the viral photo app Remini for AI-generated corporate headshots, but some say it's editing their bodies beyond recognition*. Business Insider. https://www.businessinsider.com/tiktokers-flock-remini-viral-photo-app-ai-generated-headshots-2023-7

White, D., Sutherland, C. A. M., & Burton, A. L. (2017). Choosing face: The curse of self in profile image selection. *Cognitive Research: Principles and Implications*, *2*(1), Article 23. https://doi.org/10.1186/s41235-017-0058-3

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wittenberg, C., Epstein, Z., Péloquin-Skulski, G., Berinsky, A. J., & Rand, D. G. (2025). Labeling AI-generated media online. *PNAS Nexus*, *4*(6), Article pgaf170. https://doi.org/10.1093/pnasnexus/pgaf170

Wöhler, L., Zembaty, M., Castillo, S., & Magnor, M. (2021). Towards understanding perceptual differences between genuine and face-swapped videos. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Article 240. https://doi.org/10.1145/3411764.3445627

Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human-GAI collaboration in persuasive content generation. *Judgment and Decision Making*, *18*(1), Article e41. https://doi.org/10.1017/jdm.2023.37

Appendices

Appendix A. AI-Generated and Real Photo Stimuli

Original Photo	AI-Generated
	
Photographer: Andrea Piacquadio (source)	
	
Photographer: Jeffrey Reed (source)	
	
Photographer: Royal Anwar (source)	


Photographer: Pavel Danilyuk (source)


Photographer: Andrea Piacquadio (source)


Photographer: Mizuno Kozuki (source)


Photographer: Dziana Hasanbekava (source)


Photographer: Andrea Piacquadio (source)


Photographer: Andrea Piacquadio (source)


Photographer: Meruyert Gonullu (source)


Photographer: Beth Swart (source)

Photographer: Pavel Danilyuk (source)

Appendix B. Moderation Analysis

Table B1. *Results of 3-Way Moderation Analyses*.

	Perceived photo quality		Social attractiveness		Trustworthiness	
	Coef. (*SE*)	*p* value	Coef. (*SE*)	*p* value	Coef. (*SE*)	*p* value
Intercept	5.26 (0.08)	< .001	6.32 (0.11)	< .001	6.23 (0.08)	< .001
AI Image	0.15 (0.07)	.035	0.06 (0.11)	.615	0.05 (0.11)	.663
AI Label	−0.14 (0.07)	.056	−0.03 (0.11)	.782	−0.05 (0.11)	.661
AI Knowledge	0.08 (0.02)	< .001	0.15 (0.03)	< .001	0.19 (0.03)	< .001
AI Image × AI Label	0.04 (0.14)	.779	0.29 (0.22)	.190	0.33 (0.23)	.147
AI Image × AI Knowledge	−0.02 (0.04)	.552	−0.05 (0.06)	.445	−0.02 (0.06)	.751
AI Label × AI Knowledge	0.06 (0.04)	.155	0.04 (0.06)	.555	0.04 (0.06)	.568
Image × Label × Knowl.	0.27 (0.08)	< .001	0.37 (0.12)	.002	0.34 (0.12)	.006
N	2,409		2,408		2,409	
R^2 Marg.	.034		.034		.049	
R^2 Cond.	.557		.655		.750	
AIC	7285.754		8776.030		8189.875	
BIC	7349.411		8839.682		8253.531	

Note. Models are multilevel linear regression models. Coefficients are unstandardized.

Appendix C

Figure C1. *Visualization of Moderation of Effects on Perceived Quality by AI Knowledge.*

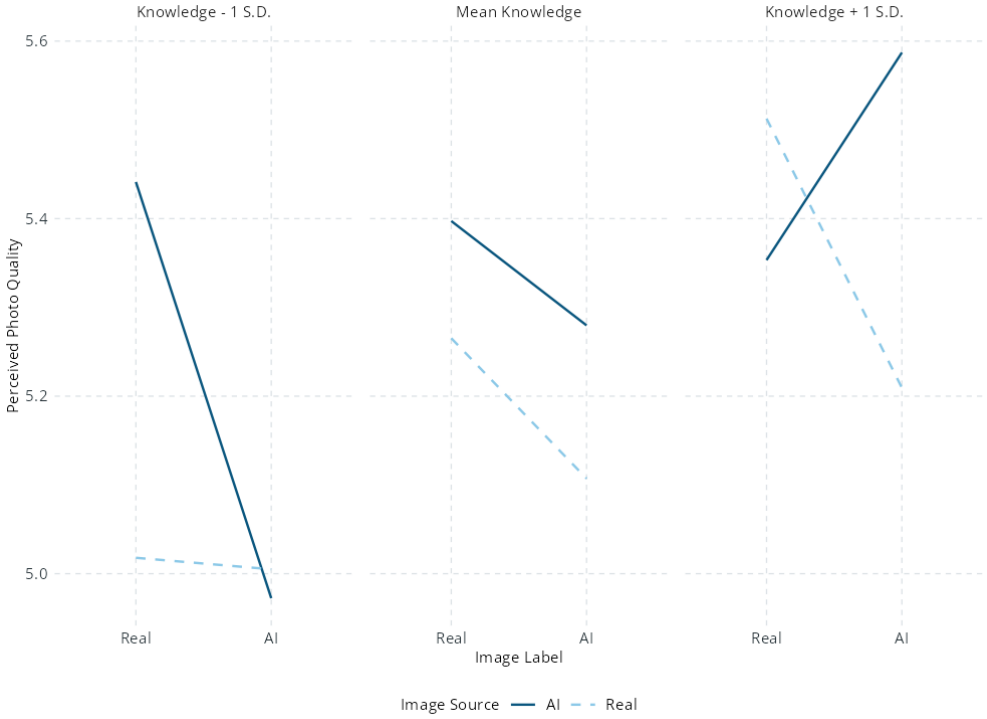


Figure C2. *Visualization of Moderation of Effects on Perceived Attractiveness by AI Knowledge.*

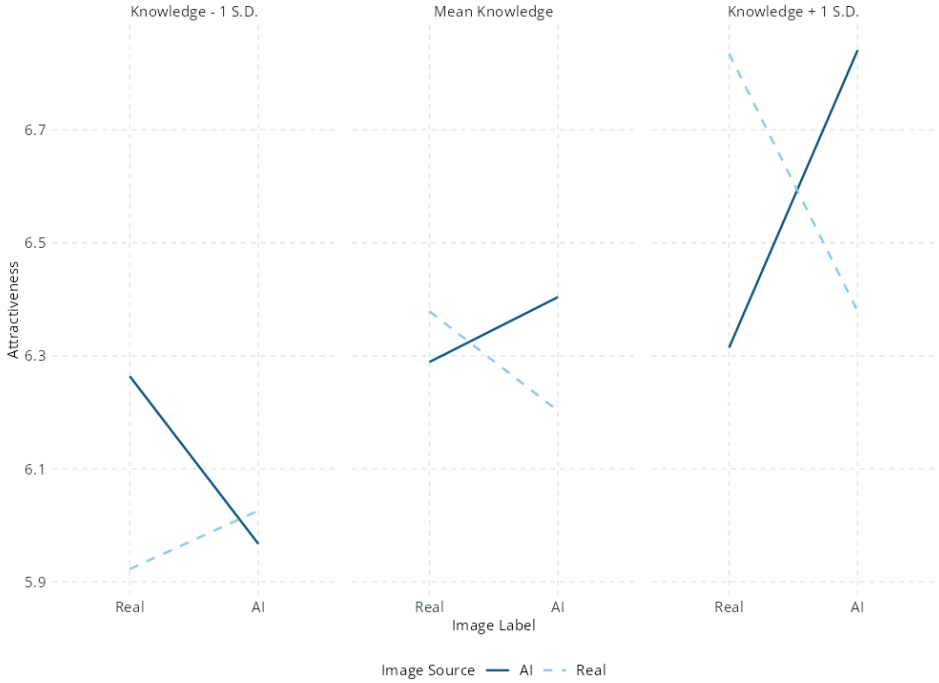
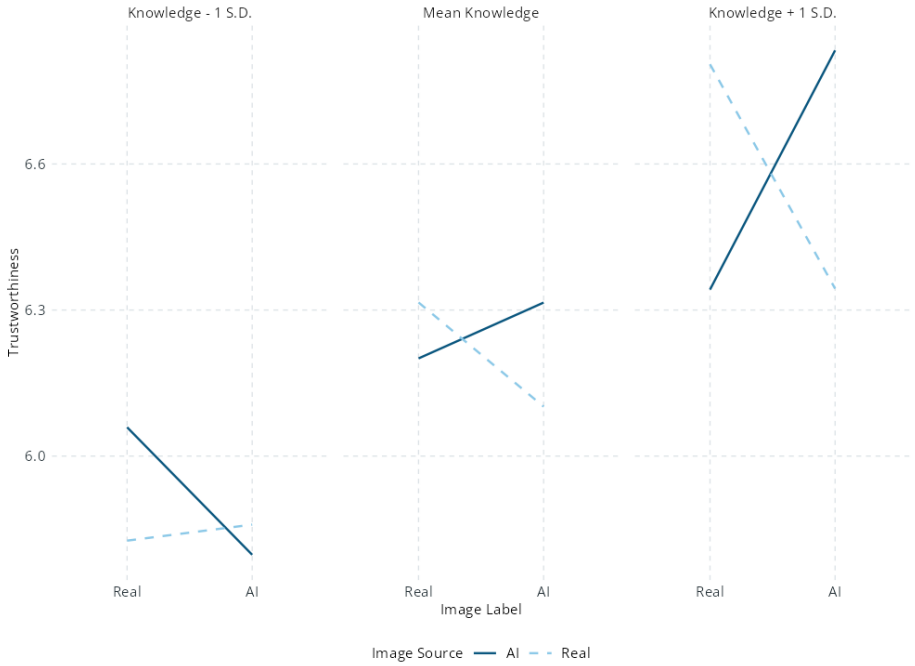


Figure C3. *Visualization of Moderation of Effects on Perceived Trustworthiness by AI Knowledge.*



Appendix D. Power Analysis

Overview and Rationale

We conducted a simulation-based design sensitivity analysis to quantify the smallest effects that our design could detect with high probability, given the observed variance structure. Rather than reporting "observed post hoc power," which is not informative, we use the fitted mixed-effects models and their variance components to simulate new datasets under prespecified effect sizes and estimate power as the proportion of simulations in which the effect is detected at α = .05. This approach answers the practical question: for effects of a given magnitude, what is the probability that a study with this design and variance structure would detect them?

Models and Parameters

For each dependent variable (photo quality, attractiveness, trustworthiness), we began with the preregistered linear mixed-effects model including fixed effects for AI generation and caption disclosure and their interaction, with crossed random intercepts for respondents and photo subjects. Predictors were effect-coded (−0.5, 0.5). The model was fit to the full dataset to obtain variance components and residual variance. We then used these fitted models as data-generating mechanisms in the simulations. The number of respondents, number of photo subjects, and the repeated-measures structure (three trials per participant) were preserved. All simulations were implemented in R using the simr package (Green & MacLeod, 2016), which extends fitted lmer/lmerTest models to power calculations by simulation.

Simulation Procedure

For each outcome, we evaluated power for the main effect of AI generation and the main effect of caption disclosure across a grid of unstandardized effect sizes expressed in the original scale of the outcome variable. For photo quality (7-point scale) as well as attractiveness and trustworthiness (9-point scale), we examined effects of several sizes: 0.1, 0.2, 0.3, 0.4 scale points. For each effect-size condition, we altered the corresponding fixed-effect coefficient in the fitted model and simulated 300 datasets, each time re-estimating the model and recording whether the target effect was significant at α = .05 (Satterthwaite degrees of freedom). Power was computed as the proportion of significant results across simulations. As a robustness check, we repeated the simulations with the trial-order covariate included; results were materially unchanged.

Results

- **Photo quality.** With the observed variance components and design, the design had approximately 30%, 75%, 98%, and 99% power to detect effects of size 0.1, 0.2, 0.3, and 0.4, respectively.
- **Attractiveness.** With the observed variance components and design, the design had approximately 15%, 39%, 73%, and 94% power to detect effects of size 0.1, 0.2, 0.3, and 0.4, respectively.
- **Trustworthiness.** With the observed variance components and design, the design had approximately 15%, 39%, 76%, and 93% power to detect effects of size 0.1, 0.2, 0.3, and 0.4, respectively.

Interpretation and Limitations

These simulations indicate that the present design is well-powered to detect substantively small main effects on perceived photo quality and requires somewhat larger but still not substantively large effects to detect changes in attractiveness or trustworthiness, consistent with the different residual and respondent-level variance for person judgments. We emphasize that these results are design sensitivity benchmarks conditional on the observed variance structure and the model specification. They are not "observed power" and should not be used to reinterpret non-significant results. Simulations assume that the fitted variance components generalize to new samples drawn under the same design; changes in stimuli, outcome distributions, or measurement reliability would change the detectable-effect profiles. Finally, the simulations focus on unstandardized effects in the original scales, which preserves interpretability but implies that detectability is a function of both variance and scale range.

Software and Reproducibility

All simulations were conducted in R, using lme4/lmerTest for model estimation and simr for power calculations (Green & MacLeod, 2016).

Reference

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Appendix E. Measurement Modeling and Alternate Specifications

For the five-item trustworthiness composite, a two-factor model that separates competence-related adjectives (competent, dependable, reliable) from integrity-related adjectives (trustworthy, honest) fit slightly better than a one-factor model (one factor: CFI = .994, TLI = .988, RMSEA = .077, SRMR = .008; two factors: CFI = .999, TLI = .998, RMSEA = .029, SRMR = .003), but the two latent factors were extremely highly correlated (r = .97). In a broader CFA of all eight person-perception adjectives, a three-factor model (social attractiveness, competence, integrity) fit slightly better than the two-factor preregistered structure (three factors: CFI = .995, TLI = .993, RMSEA = .047, SRMR = .009; two factors: CFI = .992, TLI = .988, RMSEA = .059, SRMR = .011), but the latent correlations again indicated limited discriminant separation among the constructs (r = .94–.97). Consistent with these results, robustness checks using separate competence and integrity composites yield the same substantive conclusion as the preregistered person-level outcomes: neither AI generation nor disclosure meaningfully shifted evaluations of the person depicted (competence: b = 0.110, SE = 0.114, p = .337; disclosure: b = −0.055, SE = 0.114, p = .627; integrity: b = 0.129, SE = 0.118, p = .274; disclosure: b = −0.051, SE = 0.118, p = .667).

Table E1. *Robustness Model Summaries Using Separate Competence and Integrity Composites.*

Outcome	Predictor	*b*	*SE*	*p*
Competence	AI-generated image	0.110	0.114	.337
Competence	AI disclosure	−0.055	0.114	.627
Competence	Disclosure × image	0.445	0.228	.051
Integrity	AI-generated image	0.129	0.118	.274
Integrity	AI disclosure	−0.051	0.118	.667
Integrity	Disclosure × image	0.399	0.237	.092

Note. Coefficients are unstandardized.

Table E2. *Robustness Model Summary Using a Broader Photo-Quality Composite.*

Predictor	*b*	*SE*	*p*
AI-generated image	0.210	0.064	.001
AI disclosure	−0.179	0.064	.005
Disclosure × image	0.118	0.128	.358

Note. The preregistered photo-quality model was re-estimated using a six-item composite that included the “skilled photographer” and “looks edited” items. Coefficients are unstandardized.

About Authors

Jacob A. Long, Ph.D. is an assistant professor in the School of Journalism and Mass Communications at the University of South Carolina. He studies political communication, communication technology, and research methodology.

https://orcid.org/0000-0002-1582-6214

Jingyi Xiao, M.A. is a Ph.D. student in the School of Journalism and Mass Communications at the University of South Carolina.

Shamira S. McCray, M.A. is a professional journalist and Ph.D. candidate in the School of Journalism and Mass Communications at the University of South Carolina. Her research explores news framing and how media shape public perceptions, particularly of marginalized communities.

https://orcid.org/0009-0005-1102-4037

Ertan Ağaoğlu, M.A. is a Ph.D. candidate in the School of Journalism and Mass Communications at the University of South Carolina. His research examines the role of artificial intelligence in health communication and society. He focuses on how AI-powered technologies can promote positive health behaviors, particularly among susceptible populations. He also critically investigates the societal implications of AI, with particular attention to age and gender disparities.

https://orcid.org/0000-0002-4220-2431

Abdullah M. Alajmi, M.A. is a Ph.D. candidate in the School of Journalism and Mass Communications at the University of South Carolina. His research interests include public relations, digital divide, digital governance, digital media, cyberpsychology, technology-mediated communication, and artificial intelligence. His dissertation examines how digital skills and government communication efforts influence the use of the Sahel App, Kuwait's integrated e-government platform.

https://orcid.org/0009-0008-2609-3598

Chinwendu P. Akalonu, M.A. is a Ph.D. candidate in the School of Journalism and Mass Communications at the University of South Carolina. Her research explores the intersections of communication, technology, and society. Her work centers on the adoption and societal impact of digital innovations and emerging technologies such as artificial intelligence.

https://orcid.org/0009-0009-8470-1966

Yanzhen Xu, M.A. is a Ph.D. student in the Bauer College of Business at the University of Houston.

https://orcid.org/0009-0009-2269-2574

Correspondence to
Jacob A. Long, School of Journalism and Mass Communications, University of South Carolina, 800 Sumter St, 29208, Columbia, SC, United States of America, jacob.long@sc.edu