Artificial Impressions: Trust and Credibility in AI-Enhanced Profile Pictures

Jacob A. Long, Jingyi Xiao, Shamira McCray, Ertan Ağaoğlu, Abdullah M. Alajmi,

Chinwendu Akalonu, Yanzhen Xu

Contact: jacob.long@sc.edu

Abstract

This study examines perceptions of AI-generated images used as profile pictures and subsequent evaluations of trustworthiness and credibility of someone who uses them. We conducted a pre-registered experiment in which participants evaluated profile pictures, some of which were AI-generated and some of which were disclosed as such. The comparison set were professional photos. Results indicate that the quality of AI-generated images is rated highly, yet disclosure of their AI origins leads to lower quality perceptions.

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

Artificial Impressions: Trust and Credibility in AI-Enhanced Profile Pictures

Artificial intelligence (AI) tools for generating realistic images of people have advanced rapidly in recent years. New consumer-facing services and apps allow ordinary users to create fake but convincing photos depicting themselves or others. The quality of these AI-generated images now rivals professional photography, creating new opportunities — and risks — for self-presentation online. Unlike concerns over misinformation driven by AI-generated images of celebrities and politicians (Devlin & Cheetham, 2023) or completely fictionalized personas (Menczer et al., 2023; Ross et al., 2019), under consideration here is the more mundane use of AI for what would likely be considered at worst to be mild deception for self-enhancement. This paper investigates how people perceive AI-generated profile pictures, both when they do and do not know the provenance of such images. In so doing, we seek to link the growing literature on the social perception of AI and its products with the existing work on online self-presentation.

Past research shows that social media users engage in selective self-presentation to appear attractive and gain positive impressions (Ellison et al., 2006; Hollenbaugh, 2021). Profile pictures are a key cue shaping initial judgments of a person's character and credibility. Attractive photos tend to increase perceptions of trustworthiness. However, excessive editing can raise suspicions of deception and reduce trust. Newly-developed generative AI tools provide ways to refine one's online image that may boost attractiveness or alter one's appearance in other ways. Much like photo editing, the use of generative AI introduces questions about authenticity in self-presentation. There is little evidence for whether AI-generated images presented as if they are photographs will enhance users' reputations or undermine them. Furthermore, because of the (potential) realism of AI-generated images, it is unclear whether they are likely to be perceived as faked or edited when not described as such.

To start building that evidence base, we conducted a pre-registered experiment in which participants were asked to evaluate what we described as social media profile pictures. These images were experimentally manipulated such that some were AI-generated and some were labelled as such, allowing for exploration of the separate effects of the image itself and reactions to knowledge that an image is AI-generated. Participants rated the photos' quality and the attractiveness and trustworthiness of the person depicted. Our factorial design allows us to distinguish the influence of actual AI generation from beliefs about AI origins. Comparing AI-generated photos, real professional photos, and disclosure effects tests how this emerging technology could shape online self-presentation and first impressions. Consistent with the apparent quality of AI-generated "photography," we find that such images are perceived positively. On the other hand, while participants rate images they are told are AI-generated as lower quality, they do not seem to think poorly of a person who elects to use an AI-generated photo.

Conceptualizing AI

The past few years have seen significant growth in interest around generative AI technologies like large language models (LLMs) such as ChatGPT and image generators like Midjourney. These tools can write essays, poems, news articles, and even computer code with impressive fluency. They can also generate highly realistic images from text descriptions. News media have focused substantial attention on the capabilities of these systems, often portraying them as possessing near-human intelligence to the extent that even their failures are perceived as threatening rather than evidence of non-intelligence (Tornoe, 2023). However, experts argue that analogizing these technologies to human cognition is inappropriate and the various technologies are so different from one another that to group them under a single term such as "artificial intelligence" may be misleading (Broussard, 2018; Marcus & Davis, 2020). LLMs like ChatGPT have no real understanding of the texts they generate — they simply predict sequences of words that tend to co-occur in their training data. Similarly, image generators have no meaningful concepts of the objects and scenes they depict. Their outputs, while often superficially convincing, lack deeper meaning. In the case of generated images, they often betray serious inconsistency with reality upon close inspection; for instance, up to now these tools have become somewhat notorious for their struggles in producing human hands with the usual number of fingers (Chayka, 2023), although there have been improvements in this area (Verma, 2023).

Furthermore, the umbrella term “AI” is applied to a diverse range of technologies that the lay public may fail to distinguish between. In addition to LLMs and image generators, “AI” also includes technologies like machine learning algorithms that detect credit card fraud, optimize supply chains, perform medical diagnoses, and more. These more narrow applications are quite different from technologies like ChatGPT that aim to mimic human abilities and generate content. However, media coverage and public discourse often group these technologies together under the broad "AI" label without acknowledging their differences, perhaps reasonably so considering the likely low level of general interest in the technical details. For this reason, while acknowledging the imprecision and contested nature of AI, we use this term in our discussion and in the stimulus materials because we accept the cultural discourse as it is. This study is interested in, among other things, the psychological reaction to the idea of AI being responsible for generating an image. We expect that laypeople do not make fine distinctions between the various underlying technologies and conceptualize AI as a unitary entity.

Public Opinion About AI

In the United States, citizens express a mixture of concern, optimism, and apathy towards AI. Data from 2023 (Tyson & Kikuchi, 2023) indicate a majority of Americans (52%) express more concern than excitement about the increased use of AI, while only 10%

are more excited than concerned, representing a growth in concern over the past several years. 90% of adults indicate that they have heard either a lot (33%) or a little (56%) about AI, demonstrating the rapid uptake of these technologies in the cultural discourse. Opinions on AI's impact in specific areas vary. Positive views are observed in cases such as online product and service searches, creating safe vehicles, and aiding health management. However, concerns are prominent regarding AI's impact on personal information privacy, with 53% believing that AI does more harm than good. Demographic differences play a role in shaping attitudes, with higher-educated and higher-income individuals perceiving more positive impacts of AI. In general, public opinion is contingent on the specific applications and contexts of AI and human enhancement technologies (Rainie et al., 2022). Public opinion appears to be influenced by concerns about autonomy, unintended consequences, and potential societal changes. Recent work indicates uncertainty and ambivalence in public opinion, with significant percentages expressing uncertainty about the societal impact of certain technologies (Rainie et al., 2022).

ChatGPT, the generative AI product that is probably most responsible for the surge in social and scientific interest in the role of AI in society, was known to about 74% of Americans as of late summer in 2023 (Park & Gelles-Watnick, 2023). Of those who had heard of it, about a quarter had used it, with young people far more likely to have used it (41% of 18-29 year olds) than older people (5% of those 65 or older). Americans tend to take a dim view of the likelihood that generative AI can outperform humans on social tasks, like writing news articles or provide mental health services, compared to relative optimism about AI’s potential for technical tasks such as genetic engineering and disease detection from medical images (Funk et al., 2023). That being said, there is little data regarding norms on the use of generative AI by people in their everyday lives, such as for communication and impression management.

AI Image Generation

Regarding image generation, artificial intelligence (AI) tools have ushered in a new and sometimes concerning era of AI-generated fake images, as exemplified by recent instances involving former President Donald Trump. AI-generated images, depicting scenarios such as the false arrest of Mr. Trump, often exhibit hyper-realistic qualities reminiscent of staged art rather than authentic photographs. Experts typically note some technological limitations, especially in rendering specific body parts such as hands, where inconsistencies like the number of fingers become apparent upon closer inspection (Devlin & Cheetham, 2023). Nevertheless, publicly-available generative AI tools are capable of generating images that closely resemble real photographs and are able to deceive those who are inattentive or unskilled in distinguishing real from generated images. In some cases, the outputs from generative AI can be so realistic that even a relatively savvy person may struggle to know that they are wholly computer-generated. Generally speaking, these tools can produce the most realistic outputs when asked to create an image that is similar to already-existing images in the model's training data. For instance, StyleGAN2, a generative AI algorithm trained on human headshots, produces images of non-existent people that are hardly distinguishable from real photos (E. J. Miller et al., 2023).

This proficiency at imitation has led to a rise in interest in generating photos for online impression management. AI-powered apps such as Remini, Try It On AI, and AI Suit Up have allowed users to transform casual selfies into photos that often look like professional headshots. This is not mere editing or the use of filters, but rather generating an image from whole cloth that is designed to look like a specific person *if* that person was photographed professionally. These services use a series of user-uploaded photos to give additional training to the underlying generative AI model. The user makes choices about the desired outfit, backdrop, and so on which then prompts the algorithm to generate the likeness of the user under those conditions. Within minutes, the services will provide the user a series of generated pseudo-photos meant to match their specifications (Kudhail, 2023). The technological process aims to emulate the work of expert photographers, presenting users with a selection of images that project a sophisticated and appealing professional image. Although these services often generate likenesses that may improve the user's appearance (e.g., eliminating blemishes, whitening teeth) they are not primarily oriented towards touching up — which can be accomplished via other, simpler tools — but instead the value proposition is in the staging and apparent quality of the pseudo-photo. For instance, the user may upload photos that are poorly lit, wearing informal clothing, and never directly facing the camera. The service will nevertheless generate images with the likeness of the user posed properly in adequate lighting. This is not to say that the process always works perfectly; indeed, sometimes the generated images have obvious signs of inauthenticity or otherwise do not sufficiently resemble the user. However, if the goal is to obtain just one or a small number of seemingly-real photos, a user can simply disregard the outputs that are unsatisfactory.

Although there are surely many motivations for users of these services, among them is that AI-generated headshots are perceived to be a cost-effective alternative to traditional professional photography (Hagy, 2023; Kudhail, 2023). There is no doubt that these services — some of which offer free trials — cost less, but the less obvious question is whether the quality is sufficient for use cases when one's reputation may be at stake. User responses to AI-generated profile pictures vary, with some appreciating the cost-effectiveness and ease of obtaining polished images while others express concerns about authenticity and potential misrepresentation (Kircher & Holtermann, 2022; Weiss, 2023). Indeed, the reputational risk inherent in using an AI-generated image to represent oneself is that it may perceived as deceptive. This is particularly true if the image seems to make the person depicted look better according to the standards of whoever perceives it. Although this phenomenon is too new for established social norms about use to exist, the potential for deception online is far from new.

Trustworthiness and Credibility in Online Self-Presentation

Online self-presentation is the act of sharing aspects of oneself in a way that portrays a particular image (Fullwood et al., 2020; Hsu & Lin, 2020) with the aim of shaping other’s impressions (Dhir et al., 2017). In an era marked by the ubiquitous presence of digital technologies, social media takes on the role of mediating people’s social interactions and their impression management process. Due to its significant role, researchers have been investigating the dynamics of self-presentation on social media platforms. Much of the research in this area focuses on the ways individuals use social media to represent themselves (see Hollenbaugh, 2021). However, there is relatively limited work on how individuals perceive others via their online representation as opposed to investigating the phenomenon from the perspective of the one doing the presenting.

Online platforms provide individuals with the tools to “plan, adopt, and carry out” (Arkin, 1981) their impression management, most often to present themselves in idealized ways. Users utilize a variety of cues, enabled by the technological affordances of the platforms, for their online self-presentation efforts (Osterholz et al., 2023), such as text- or image-based information about themselves (Coduto, 2022) or more implicit cues such as privacy choices, number of followers, etc. (Osterholz et al., 2023). Image-based cues, like profile pictures or avatars, stand out as one of the most important cues as they represent an individual’s physical form in the digital realm (Wu et al., 2015). They also serve as markers of physical attractiveness (van der Zanden et al., 2022), a key factor in shaping initial impression formation (A. G. Miller, 1970). This aspect significantly influences others' overall judgments regarding a person's character and personality. For example, research shows that attractive individuals are perceived as having better social skills and being more intelligent than their less attractive counterparts (Peng et al., 2020). In the social media context, people tend to form positive impressions in the absence of negative information on user profiles. Positive evaluations of attractiveness and personality tend to co-occur in this context (Bacev-Giles & Haji, 2017). First impressions of faces tend to produce highly-correlated evaluations across seemingly-unrelated dimensions such as physical attractiveness and trustworthiness (Todorov et al., 2008).

In summary, evidence suggests that AI-generated profile photos may have similar or better quality than professional photos, depending on the quality of the AI model and the available training data. Furthermore, a quality photo is expected to lead to a person who is perceived as attractive. And research on first impressions suggests that attractive people will also be evaluated as more trustworthy. With this information in mind, we offer the following hypotheses[1]:

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

H2a: AI-generated profile pictures will lead the person in the image to be perceived as more attractive than when the subject is depicted using a real, professional photo.

H3a: AI-generated profile pictures will lead the subject of the image to be perceived as more trustworthy than when the subject is depicted using a real, professional photo.

Perceived Authenticity of Images

The tendency to form impressions based on limited cues underscores the challenges of navigating the digital sphere where technological affordances let users engage in selective self-presentation, bringing with it concerns regarding authenticity and deception. On online platforms, authenticity is problematic due to the mediated and asynchronous nature of these

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

platforms enabling individuals to construct the information about (or, the presentation of) themselves selectively and removing the possibility of unintentional cues that can damage their impression in the eyes of others (Sessions, 2009). Online platforms are often characterized by idealized self-presentation (Harris & Bardey, 2019). As one of the most important parts of online self-presentation, profile pictures are scrutinized for their authenticity. The affordances of social media platforms in combination with photography tools (both hardware and software) provide users with the ability to stage, choose, edit, and adjust their photographs to selectively present themselves (Fox & Vendemia, 2016). When people present themselves in idealized ways, this creates an inconsistency between their online and offline selves which is often seen as deception (Sessions, 2009), that is the “deliberate falsification or omission of information by a communicator, with the intent being to mislead the conversational partner” (McCornack & Levine, 1990, p. 120). This unsurprisingly harms one’s trustworthiness and credibility (Sharabi & Caughlin, 2019).

Social media users over time have become more conscious of the image and personas they portray for their audiences, sometimes turning to photo editing as a way to create a socially desirable appearance online. Artificial intelligence has become the new frontier in editing photos, giving people a potentially easier way to create more attractive versions of themselves for online consumption. These new editing tools are already embedded into some digital graphics editors like Photoshop and social media apps, such as Instagram, that offer photo filtering options for users (Fatima, 2020). People can use several methods to edit a photo, including applying filters to the image, cropping, or using AI enhancements. Research suggests that idealized body norms, including those presented in the mass media, can be detrimental to a person’s psychosocial health and can lead to eating disorders and a multitude of other issues (Fredrickson & Roberts, 1997; Grabe et al., 2008; Naderer et al., 2021).

Disclaimers provide a way to alert viewers that the physical appearance represented in a photo is not necessarily real or attainable. Researchers have studied the effectiveness of such disclaimers, and several have found that they are not usually effective in mitigating the negative effects of comparing one's own body to the idealized representation online (Lewis et al., 2020; Tiggemann & Brown, 2018). One study failing to find an effect of disclaimers (Naderer et al., 2022) indicated that one reason disclaimers might have previously been shown to be ineffective is because they are too unobtrusive and not recognized by recipients. Another noted that many participants report suspicions of edited photos even without a disclaimer telling them so (Lewis et al., 2020), suggesting the disclaimers are redundant for some people.

Recent research indicates that AI-generated social media profiles can be perceived as real by expert users, even in direct comparison to genuine social media profiles (Paradise et al., 2018). Such profiles can even gain a foothold in social networks (Freitas et al., 2016), raising concerns about the potential for abuse. Malicious bots do not need to comprise a large portion of a network to have a significant effect on discourse, so they do not need to be conspicuous to be important (Ross et al., 2019). The use of AI-generated images to accompany profiles intended to deceive can increase the likelihood of detection if there are obvious defects with the image, but even when defects are present many users fail to notice them (Mink et al., 2022). Although the present study is not primarily concerned with entirely fictional personas, the approach is relevant for the case of more significant deception as well.

Effects of AI Disclosure

In general, the disclosure that content is generated by AI tends to decrease user trust across a number of domains. Consistent findings show that telling people something has been done with AI affects their perceptions and, subsequently, their behavior towards the AI or the task it has performed. People seem to prefer tasks performed by humans compared to tasks performed by AI even when the AI performs better on average, at least in high-uncertainty tasks like driving and medical diagnosis (Dietvorst & Bharti, 2020). Many researchers have shown that even when these tasks are executed efficiently, there is still a persistent negative perception or behavior after the disclosure of AI use. Evidence suggests there is a particular aversion to artificial intelligence separate from the actual outputs of artificial intelligence; for instance, even with equivalent interactions between customer and a chatbot interlocutor, sales drop significantly when the chatbot is revealed to be a bot rather than a human (Luo et al., 2019). Similar findings exist in areas like employee recruitment (Keppeler, 2023), customer service (Mozafari et al., 2021), and supervisory feedback (Tong et al., 2021).

On the contrary, one study found that disclosure of whether a visual artist was human or AI did not affect participants' level of appreciation of the artwork (Messingschlager & Appel, 2023). Another found that disclosure of the author of poetry being AI did not harm evaluations of the poetry except in the case of emotional, first-person formats (Raj et al., 2023). When comparing marketing materials produced by human professionals and generative AI, evaluators were more negative towards the AI content when disclosed as such but still preferred it over the human-generated alternatives, suggesting that people may feel some aversion but not always so strongly that it overrides their objective assessment (Zhang & Gosline, 2023). Overall, although current research tends to find some level of hesitancy towards the use of AI, it clearly varies based on individual attitudes and experiences (e.g., Araujo et al., 2023) as well as contextual factors that are not yet fully understood.

Given these patterns in the research findings, we offer the following hypotheses regarding disclosure of the use of generative AI:

H1b: Profile pictures that are disclosed as generated by AI will be perceived as lower quality compared to pictures disclosed as taken by a professional photographer.

H2b: Profile pictures that are disclosed as generated by AI will lead the person in the image to be perceived as less attractive than when the profile picture is disclosed as having been taken by a professional photographer.

H3b: Profile pictures that are disclosed as generated by AI will lead the person in the image to be perceived as less trustworthy than when the profile picture is disclosed as having been taken by a professional photographer.

Methods

Preregistration

This study's hypotheses, design, and analysis were all pre-registered prior to the start of data collection. There are no deviations from the preregistration to report. The preregistration is available in anonymous format for peer review at https://osf.io/b47d6/?view_only=cc1543db86df45bb8aeb9381ba4d98a8.

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 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% men, 50% women, and less 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 <1% as another race. Average age was 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 towards images produced by generative AI, we produced images with generative AI as well as real photographs to which comparisons can be made. 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 — most typically 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 — that is, there are 12 AI-generated photos used as stimuli and 12 real photos used as stimuli.

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

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. A table in a supplementary file[4] included the anonymized online registration contains the images used. 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.

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 summary, the study design is a 2x2 factorial experiment with repeated measures and varying stimuli. The experimental factors are whether the image is AI-generated (as opposed to a real, professional photograph) and whether the image is characterized to the participant as generated by AI (as opposed to crediting a photography agency). Participants are informed that the study is designed to assess how people perceive social media profile pictures. Participants are then shown an image with a note that they will be asked questions

[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] https://osf.io/tmxgy?view_only=0d340362ba774867b89be7bf9b13631f

about it afterwards. Here is where the two experimental factors are manipulated. One of the 12 stimulus variants for their image condition (AI-generated or not) is displayed. The photo is accompanied with a caption that states 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 of the experiment means that sometimes these captions are misleading by either describing a real photo as AI-generated or an AI-generated photo as taken by a photographer, allowing for analyses to distinguish effects that arise from AI generation in contrast to the mere perception of AI generation. After seeing the image and caption, respondents are presented with a series of items assessing their perception of the photo’s quality as well as their perceptions of the photo subject’s attractiveness and apparent trustworthiness/credibility.

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.

Measures

Perceived photo quality

Participants are presented with a series of statements regarding the quality of the photo with response choices ranging from “Strongly disagree” (coded as 1) to “Strongly agree” (7). There are 4 such statements: “this is a good photo”; “this photo is appropriate for professional settings”; “this photo is an accurate representation of the person in it”; and “if the person in the photo was a friend of mine, I would encourage them to use it for their social media profiles.” The responses are combined into a single measure by averaging ($M = 5.27$, $SD = 1.27$). Reliability, as assessed by McDonald’s Omega to account for the repeated measures design (Geldhof et al., 2014), is acceptable at .79.

Social attractiveness

Participants were then asked to assess 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 the respondents. For attractiveness, the responses for the adjectives “attractive,” “likable,” and “warm or sympathetic” are averaged to create the measure used for analysis ($M = 6.34$, $SD = 1.85$). The McDonald’s Omega reliability for this measure is acceptable at .72.

Trustworthiness

Along with the adjectives used to tap into social attractiveness, responses for 5 other adjectives — “competent,” “trustworthy,” “honest,” “dependable,” and “reliable” — are averaged to generate our measure of trustworthiness perceptions ($M = 6.25$, $SD = 1.79$). McDonald’s Omega reliability is measured at .84.

Results

As specified in the preregistration, hypotheses are tested by way of multilevel linear regression models due to the clustering introduced by the repeated measures design and to account for the multiple stimulus variants. There are 3 models, one with each of the key variables as outcome: perceived photo quality, social attractiveness, and trustworthiness. Models have 3 fixed predictors: an indicator for whether the image is AI-generated, an indicator for whether the caption claimed the image was AI-generated, and an interaction term. There are no hypotheses about interactions, but the term is included due to the factorial design and to rule out any unanticipated interactions. Finally, the models include random intercepts for the participant as well as the subject of the photo to account for idiosyncrasies at those two levels. Inferences are made based on the main effect terms; however, the presence of an interaction term complicates the interpretation of the lower order terms so they are coded as -0.5/0.5 (equivalent to contrast and effect coding in this two-condition design) as opposed to the more typical 0/1 dummy coding. This allows the main effect terms to be used for inference since they represent the average effect. 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, 2023). Details of the three models are included in Table 1.

Results indicate that, as predicted by hypothesis H1a, AI generated images are perceived as higher quality (B = 0.21, $p = .005$). Additionally, as predicted by H1b, disclosure that the image is generated by AI reduces the perceived quality of the image (B = -0.15, $p = .041$). H3a predicted that people depicted by AI-generated images would be perceived as more trustworthy, but results do not support the hypothesis (B = 0.12, $p = .306$). Likewise, the data do not support the H3b prediction that disclosure of AI generation would reduce the perceived trustworthiness of the person in the photo (B = -0.05, $p = .655$). The predictions of H2a and H2b are also unsupported for perceived attractiveness; AI-generated images are not associated with significantly greater attractiveness (B = 0.11, p = .313) and AI disclosure is not associated with less attractiveness (B = -0.02, p = .856).

Discussion

Our results suggest that the use of generative AI for online self-presentation may not be particularly risky for the user’s reputation. Even compared against a professional photo, our AI-generated versions were perceived to be of better quality. When told that the image is generated by AI, participants held the image quality in lower esteem, but not to an extent to even cancel out the gain in perceived quality from using the AI-generated images. This pattern of results — AI disclosure harms perceptions, but not enough to undo the positive evaluation of its quality relative to a human-made alternative — is similar to a recent study in which participants evaluated marketing materials (Zhang & Gosline, 2023). The potential negative judgments about one’s trustworthiness and attractiveness that could plausibly flow downstream from the more negative perception of photo quality are not apparent in our study. Of course, we do not see significant benefits on those personal judgments from the better-looking AI-generated images either.

It is likely that the wide availability of generative AI tools will lead people to test the limits of both the technology and social norms about self-presentation online. The technology itself has proved a rapidly moving target in recent years; this study’s design was not possible using consumer-facing tools even a year before it commenced. The quality, quantity, and capabilities of these tools are likely to continue to improve. The applicable social norms are more uncertain, although past research indicates that social media users expect some level of embellishment on profiles (Ellison et al., 2006) and photos are the area with the most apparent flexibility in engaging in light deception (Hancock & Toma, 2009). On a philosophical level, it is not clear that the use of generative AI *must* be deceptive; indeed, it seems one way to conceptualize a high-quality output from such a tool is that it should resemble a (perhaps hypothetical) real photo as closely as possible. In this study, the generative AI tool in some cases produced very realistic representations of the subjects that are not at all easily distinguished from real photos with ordinary post-processing.

A positive of this study's design is that by using professional photos as a point of comparison, the AI-generated photos are contrasted with an idealized alternative. This is also a limitation, however, since it appears the primary appeal to real-world users of these generative AI tools is to substitute for professional photography. Although some users may truly pick between generative AI and high-quality photography, many others likely feel that quality photos are out of their reach and they must instead decide between less-professional photos and AI-generated images styled as professional shots. Our reliance on stock photo models means we are not able to compare perceptions of "selfies" and the like with AI-generated headshots. We applied a more difficult test, but one that is less reflective of the alternatives available to many consumers. Follow-up research should consider incorporating lower-quality photos as another form of experimental control.

Another limitation of this design involves the disclosure. Today's social media platforms provide no obvious way to disclose the provenance of the user's profile picture, presumably since the platforms were created with the assumption that such images were genuine, obviously not-genuine, or that users had no interest in using a platform feature to self-disclose arguably deceptive behavior. We felt this was a necessary tradeoff to understand the extent to which reactions to AI are driven by aspects of the images themselves as opposed to participants' beliefs that the images were AI-generated. In real-world settings, users undoubtedly encounter some AI-generated images but will not frequently receive any verification one way or another. Our control condition of having a caption disclosing a photographer is justified from a pure experimental design standpoint, but such disclosures are not the norm online (albeit not unheard of). It is possible that drawing attention to the creator of the image, whether technology or photographer, could change perceptions compared to the way people would naturally encounter profiles in their daily lives. Beyond this, the timing and manner of disclosure of the use of AI likely matters for how it is received. The way the disclosure was made in this study is perhaps the least deceptive possibility; photo and disclosure are given simultaneously, so perceivers should not feel that they were deceived when learning of the provenance of the image. Real-world disclosures are likelier to be both less certain (i.e., take the form of speculation by the perceiver based on qualities of the image or an allegation by another party) and occur after the first impression has been formed. Future work should explore how to experimentally manipulate disclosure in ways that may be more ecologically valid.

References

Araujo, T., Brosius, A., Goldberg, A. C., Möller, J., & Vreese, C. de. (2023). Humans vs. AI: The role of trust, political attitudes, and individual characteristics on perceptions about automated decision making across Europe. *International Journal of Communication*, *17*, 6222–6249.

Arkin, R. M. (1981). Self-presentational styles. In J. T. Tedeschi (Ed.), *Impression management theory and social psychological research* (pp. 311–333). Academic Press.

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). 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

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.

Chayka, K. (2023, March 10). The uncanny failure of A.I.-generated hands. *The New Yorker*. https://www.newyorker.com/culture/rabbit-holes/the-uncanny-failures-of-ai-generated-hands

Coduto, K. (2022). Online performances of fandom: Selective self-presentation, perceived affordances, and parasocial interactions on social media. In *Routledge Handbook of Sport Fans and Fandom*. Routledge.

Devlin, K., & Cheetham, J. (2023, March 24). Fake Trump arrest photos: How to spot an AI-generated image. *BBC News*. https://www.bbc.com/news/world-us-canada-65069316

Dhir, A., Torsheim, T., Pallesen, S., & Andreassen, C. S. (2017). Do online privacy concerns predict selfie behavior among adolescents, young adults and adults? *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00815

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

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

Fatima, N. (2020). AI in photography: Scrutinizing implementation of super-resolution techniques in photo-editors. *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 1–6. https://doi.org/10.1109/IVCNZ51579.2020.9290737

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

Freitas, C., Benevenuto, F., Veloso, A., & Ghosh, S. (2016). An empirical study of socialbot infiltration strategies in the Twitter social network. *Social Network Analysis and Mining*, *6*(1), 23. https://doi.org/10.1007/s13278-016-0331-3

Fullwood, C., Wesson, C., Chen-Wilson, J., Keep, M., Asbury, T., & Wilsdon, L. (2020). If the mask fits: Psychological correlates with online self-presentation experimentation in adults. *Cyberpsychology, Behavior, and Social Networking*, *23*(11), 737–742. https://doi.org/10.1089/cyber.2020.0154

Funk, C., Tyson, A., & Kennedy, B. (2023, February 22). How Americans view emerging uses of artificial intelligence, including programs to generate text or art. *Pew Research Center*. https://www.pewresearch.org/short-reads/2023/02/22/how-americans-view-emerging-uses-of-artificial-intelligence-including-programs-to-generate-text-or-art/

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/

Hancock, J. T., & Toma, C. L. (2009). Putting your best face forward: The accuracy of online dating photographs. *Journal of Communication*, *59*(2), 367–386. https://doi.org/10.1111/j.1460-2466.2009.01420.x

Harris, E., & Bardey, A. C. (2019). Do Instagram profiles accurately portray personality? An investigation into idealized online self-presentation. *Frontiers in Psychology*, *10*. 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

Hsu, C.-L., & Lin, J. C.-C. (2020). Antecedents and gains of user participation in social media in Taiwan. *Technology in Society*, *61*, 101243. https://doi.org/10.1016/j.techsoc.2020.101243

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*, muad009. https://doi.org/10.1093/jopart/muad009

Kircher, M. M., & Holtermann, C. (2022, December 8). How is everyone making those A.I. selfies? *The New York Times*. https://www.nytimes.com/2022/12/07/style/lensa-ai-selfies.html

Kudhail, P. (2023, October 11). 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). 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. (2023). *<pre>jtools</pre>: Analysis and presentation of social scientific data* (2.2.3) [Computer software]. https://CRAN.R-project.org/package=jtools

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

Marcus, G., & Davis, E. (2020, August 22). *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*. MIT Technology Review. https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/

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

Menczer, F., Crandall, D., Ahn, Y.-Y., & Kapadia, A. (2023). Addressing the harms of AI-generated inauthentic content. *Nature Machine Intelligence*, *5*(7), Article 7. https://doi.org/10.1038/s42256-023-00690-w

Messingschlager, T. V., & Appel, M. (2023). Mind ascribed to AI and the appreciation of AI-generated art. *New Media & Society*, 14614448231200248. https://doi.org/10.1177/14614448231200248

Miller, A. G. (1970). Role of physical attractiveness in impression formation. *Psychonomic Science*, *19*(4), 241–243. https://doi.org/10.3758/BF03328797

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

Mink, J., Luo, L., Barbosa, N. M., Figueira, O., Wang, Y., & Wang, G. (2022). *{DeepPhish}: Understanding user trust towards artificially generated profiles in online social networks*. 1669–1686. https://www.usenix.org/conference/usenixsecurity22/presentation/mink

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 R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4,* 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Osterholz, S., Mosel, E. I., & Egloff, B. (2023). #Insta personality: Personality expression in Instagram accounts, impression formation, and accuracy of personality judgments at zero acquaintance. *Journal of Personality*, *91*(3), 566–582. https://doi.org/10.1111/jopy.12756

Paradise, A., Cohen, D., Shabtai, A., & Puzis, R. (2018). ProfileGen: Generation of automatic and realistic artificial profiles. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 581–583. https://doi.org/10.1109/ASONAM.2018.8508242

Park, E., & Gelles-Watnick, R. (2023, August 28). Most Americans haven't used ChatGPT; few think it will have a major impact on their job. *Pew Research Center*. https://www.pewresearch.org/short-reads/2023/08/28/most-americans-havent-used-chatgpt-few-think-it-will-have-a-major-impact-on-their-job/

Peng, L., Cui, G., Chung, Y., & Zheng, W. (2020). The faces of success: Beauty and ugliness premiums in e-commerce platforms. *Journal of Marketing*, *84*(4), 67–85. https://doi.org/10.1177/0022242920914861

Rainie, L., Funk, C., Anderson, M., & Tyson, A. (2022, March 17). AI and human enhancement: Americans' openness is tempered by a range of concerns. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2022/03/17/ai-and-human-enhancement-americans-openness-is-tempered-by-a-range-of-concerns/

Raj, M., Berg, J., & Seamans, R. (2023). *Art-ificial intelligence: The effect of AI disclosure on evaluations of creative content* (arXiv:2303.06217). arXiv. https://doi.org/10.48550/arXiv.2303.06217

Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., & Stieglitz, S. (2019). Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. *European Journal of Information Systems*, *28*(4), 394–412. https://doi.org/10.1080/0960085X.2018.1560920

Sessions, L. F. (2009). "You Looked Better on MySpace": Deception and authenticity on the Web 2.0. *First Monday*. https://doi.org/10.5210/fm.v14i7.2539

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

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., 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

Tornoe, R. (2023). Media coverage largely focuses on AI alarmism. *Editor & Publisher*, *156*(7/8), 18–19.

Tyson, A., & Kikuchi, E. (2023, August 28). Growing public concern about the role of artificial intelligence in daily life. *Pew Research Center*. https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/

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

Verma, P. (2023, March 26). AI can draw hands now. That's bad news for deep-fakes. *Washington Post*. https://www.washingtonpost.com/technology/2023/03/26/ai-generated-hands-midjourney/

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

Wu, Y.-C. J., Chang, W.-H., & Yuan, C.-H. (2015). Do Facebook profile pictures reflect user's personality? *Computers in Human Behavior*, *51*, 880–889. https://doi.org/10.1016/j.chb.2014.11.014

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* (SSRN Scholarly Paper 4453958). SSRN. https://doi.org/10.2139/ssrn.4453958

Table 1

Multilevel regression results for each dependent variable

	Quality	Attractiveness	Trustworthiness
Fixed effects			
AI-generated image	0.21 (0.07)*	0.11 (0.11)	0.12 (0.11)
AI disclosure	-0.15 (0.07)*	-0.02 (0.11)	-0.05 (0.11)
Disclosure x image	0.11 (0.15)	0.34 (0.22)	0.43 (0.23)
Intercept	5.27 (0.08)*	6.34 (0.11)*	6.25 (0.08)*
Random effects			
S.D. Respondent	0.91	1.46	1.54
S.D. Photo subject	0.26	0.32	0.17
S.D. 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 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 and include the fixed and random effects.

* $p < .05$