Journal of Community & Applied Social Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

When the Source Is Artificial: Negative Emotions, Intergroup Anxiety, and Threat Responses to AI-Generated Video Content

Gabbiadini Alessandro | Manfredi Anna | Serrao Fabrizio | Puzella Giulio

Department of Psychology, University of Milano Bicocca, Milano, Italy

Correspondence: Gabbiadini Alessandro (alessandro.gabbiadini@unimib.it)

Received: 2 November 2025 | **Revised:** 18 March 2026 | **Accepted:** 28 March 2026

Keywords: artificial intelligence | intentions to use | intergroup anxiety | negative emotions | threat perception

ABSTRACT

Generative artificial intelligence increasingly produces realistic video, raising questions about how people emotionally and socially respond to AI outputs. We extended prior work on text, image, and audio generation by focusing on AI-generated video and on affective and intergroup mechanisms linked to intentions to use AI. In a 2(source: AI vs. human)×2(quality: flawless vs. error-containing) between-subjects design, participants ($N = 138$) watched a short video and reported negative emotions, identity and realistic threat, intergroup anxiety, and intentions to use AI tools. We also explored whether the presence of AI errors (hallucinations) altered affective reactions. Source attribution (AI vs. human), but not output quality or their interaction, increased negative emotions, indicating that discomfort was driven more by the AI label than by the presence of hallucinations. We then estimated a staged mediation model with source as predictor, negative emotions as a first-stage mediator, and threat perceptions and intergroup anxiety entered in parallel as downstream mediators. The only reliable indirect effect linked AI source to lower intentions via negative emotions and intergroup anxiety. These findings highlight the relevance of intergroup frameworks for understanding public responses to generative video and for guiding human–AI interface design. Adoption, trust, and communication implications are discussed.

1 | Introduction

Artificial intelligence (AI) has rapidly evolved into one of the most transformative technological forces of the 21st century, profoundly reshaping the ways in which individuals, organizations, and societies function. From optimizing service delivery in sectors such as healthcare, education, and hospitality to automating tasks and generating content across creative industries, AI technologies now permeate numerous aspects of everyday life. Among these advancements, generative AI models represent a particularly significant breakthrough. Unlike earlier systems designed primarily for rule-based problem solving, generative AI produces outputs that resemble human-created text, images, music, and even video. Capable of simulating creative and cognitive processes once considered uniquely human (Siemens et al. 2022), these systems move capabilities once confined to science fiction into the realm of ordinary experience.

The promise of AI lies not only in its ability to enhance efficiency and innovation but also in its potential to augment human performance (Mirbabaie et al. 2022). Nevertheless, this expansion is accompanied by widespread ambivalence. Alongside enthusiasm for its potential benefits, AI adoption has triggered apprehension, skepticism and fear (Xu et al. 2024). As scholars have argued, the social and psychological consequences of AI extend far beyond technical issues

of performance and accuracy; they touch upon fundamental questions of identity, agency, and intergroup relations (Gabbiadini, Ognibene, et al. 2024; Gabbiadini, Durante, et al. 2024; Kieslich et al. 2021).

Previous research has mainly focused on *what* AI can do, but much less is known about *how well* it does it and how this shapes people's reactions. Generative AI models routinely produce errors and "hallucinations" (i.e., confident but factually wrong or impossible outputs produced by generative AI, such as invented facts, inconsistent details, or events that could not occur in reality) that might fuel algorithm aversion (for a review, see Burton et al. 2020; Gaube et al. 2021; Jussupow et al. 2020; Mahmud et al. 2022) and avoidance. Yet current research has not clarified whether such errors dampen threat and negative emotions, or instead amplify distrust and resistance. More broadly, the literature has identified several candidate mechanisms—negative emotions, realistic and identity threat (Gabbiadini, Ognibene, et al. 2024), intergroup anxiety (Zhou et al. 2022)—but it is still unclear how these processes relate to each other, which of them is most closely linked to lower AI adoption, and whether output quality influence these pathways.

The current study aimed to replicate and extend recent findings (Gabbiadini, Ognibene, et al. 2024) by examining emotional and intentional responses to generative AI, with a specific focus on how an initial negative emotional reaction may translate into downstream psychological mechanisms and future intentions to engage with AI technologies. We also sought to examine whether typical AI "hallucinations" in video outputs would shape these reactions by contrasting accurate versus error-containing AI-generated content. Accordingly, we adopted a staged process perspective and planned, first, to assess how exposure to AI-versus human-generated video content (and their respective quality: accurate vs. error-containing) relates to general negative emotions, perceived realistic and identity threat, intergroup anxiety, and intentions to use AI; and second, to test a process model in which negative emotions are treated as an early affective signal potentially associated with three conceptually distinct, more proximal pathways: identity threat, realistic threat, and intergroup anxiety, offering a more comprehensive picture of the mechanisms that may shape public reactions to generative AI.

Moreover, whereas previous work tested responses to text, images, and audio, the present research investigates reactions to AI-generated video content, a novel generative AI modality—video production, which was not publicly available at the time of the original study—thereby broadening the ecological scope of this line of research. In this regard, the field of AI-generated video content is rapidly expanding across both social media and the entertainment industry. The global market for artificial intelligence in media and entertainment is projected to grow substantially between 2024 and 2033, reflecting the widespread adoption of AI for content creation (Markets and Markets 2024). As of late 2023, a large share of U.S. movie, TV, and animation companies reported using or planning to use generative AI in production (Statista 2024), while recent releases of video-generation tools have attracted wide public attention and uptake (The Verge 2025). This extension also allowed for the investigation of whether the patterns observed with text, images, and music generalize to a more complex AI format.

2 | Perceiving Generative AI as an Outgroup: Threat, Negative Emotions, and Intergroup Anxiety

Recent work suggests that Generative AI is not seen only as a tool but can also be construed in group-like terms, producing human-like outputs (text, music, images; McKee et al. 2023). Research on the "computers as social actors" paradigm shows that individuals can spontaneously apply social scripts and norms to non-human systems when these display human-like cues or agency, responding to them as if they were interaction partners rather than neutral tools (Nass and Moon 2000). Related work on mind perception and anthropomorphism further indicates that, when a technology appears intentional, expressive, or capable of autonomy, people are more likely to treat it as a social entity (Epley et al. 2007; Waytz et al. 2010). In this sense, AI can become a *social* target, especially when it is framed as an agentic "source" that produces human-like outputs and may evaluate, replace, or compete with humans (Złotowski et al. 2017; Gabbiadini, Ognibene, et al. 2024). Under these conditions, it becomes meaningful to extend intergroup frameworks to human–AI relations, framing AI as a potential outgroup.

According to the Stereotype Content Model (SCM; Fiske et al. 2002), groups judged high in competence but uncertain or low in warmth (i.e., skillful but of unclear intent) trigger mixed reactions: admiration for what they can do, and concern about what they might do (Fiske 2018; Cuddy et al. 2008). In line with this, people seem to rate AI as highly competent, while warmth ratings vary (McKee et al. 2023). Although McKee et al. did not measure threat directly, SCM theory suggests that such a profile (high competence + uncertain warmth) fits the "envy" cluster, which is linked to heightened attention, careful evaluation, and sometimes defensive responses when status or control could be at risk (Fiske 2018; Gabbiadini, Ognibene, et al. 2024). This competence–warmth profile is theoretically relevant because it implies an ambivalent stance that may be accompanied by an initial negative affective response and, in a second step, more differentiated threat-related reactions.

Indeed, a growing body of work shows that interactions with autonomous technologies can follow intergroup patterns: people often feel threatened by highly competent outgroups, even when those outgroups are robots or software, and show ingroup bias favouring humans over AI (Maddux et al. 2008; Yogeeswaran et al. 2016; Morewedge 2022). This is especially clear for professional identity: working with AI can be seen as a threat to one's status and role, which increases resistance to adoption (Jussupow et al. 2022; Mirbabaie et al. 2022).

Encounters with generative AI, especially when the system is framed as an agentic "source" rather than a neutral tool, are unlikely to trigger one single reaction. Recent works (Gabbiadini, Durante, et al. 2024; Brauner et al. 2025) has highlighted that people often judge AI by weighing perceived risks against perceived benefits. Indeed, AI technologies seem

to be evaluated along two broad and partly opposing dimensions: their social value (what they can bring to society) and their social risk (what harm or costs they may create). In this perspective, risk-based theories help clarify how an initial global affective reaction can emerge, while intergroup frameworks help specify which types of meaning (e.g., identity-, resource-, and interaction-related concerns) that reaction may become linked to.

In this regard, the risk-as-feelings hypothesis suggests that responses to risk are not based solely on deliberate reasoning about probabilities and outcomes. Accordingly, people often rely on a fast, intuitive affective reaction that can shape judgments and choices at the moment they evaluate a target (Loewenstein et al. 2001). Such affective responses are particularly influential when the target is complex, uncertain, or difficult to evaluate, conditions that often characterize novel technologies such as generative AI, whose mechanisms are opaque and whose outputs are hard to anticipate. Later work also argues that perceived risk often reflects two interacting modes of processing: a fast affective mode and a slower analytic mode (Slovic 2004). When stimuli are complex and outcomes are hard to predict, as is often the case for generative AI, people may rely more on initial affective impressions, which can then bias subsequent judgments of risks and benefits (Slovic et al. 2005, 2007). A plausible implication is that general negative emotions may function as an early affective signal: an initial "alarm" response indicating that an encounter may involve risk, loss, or instability. Consistent with this view, prior work indicates that exposure to generative AI's human-like capabilities is linked to stronger negative emotions than comparable human outputs, and that negative emotions are associated with heightened threat perceptions (Gabbiadini, Ognibene, et al. 2024). This suggests that an initial negative reaction may potentially inform more differentiated, action-relevant responses. In this context, three downstream responses appear especially relevant because they are conceptually close to disengagement and avoidance: identity (symbolic) threat, realistic threat, and intergroup anxiety. Two of these responses—identity and realistic threat—map directly onto Intergroup Threat Theory (Stephan and Stephan 1985), which provides a well-established framework for specifying what is perceived to be at stake in intergroup encounters.

The Intergroup Threat Theory distinguishes realistic threats (to safety, employment, and material resources) from symbolic threats (to values, beliefs, and identity). Applied beyond human–human settings, the same logic helps explain human–AI relations: AI can function as an ambiguous "other" whose capabilities unsettle existing social categories and human distinctiveness (Branscombe et al. 1999; Stephan et al. 1999; Złotowski et al. 2017; Gabbiadini, Ognibene, et al. 2024). In this way, AI technologies can pose symbolic threats by undermining a sense of human identity and uniqueness, and realistic threats by jeopardizing jobs, expertise, and control over outcomes (Stephan and Stephan 1985; Stephan et al. 2008, 2016). Generative AI intensifies the symbolic side of this equation by entering domains historically seen as uniquely human—creative and cultural production—thereby heightening concerns about status, autonomy, and expertise. These appraisals are often accompanied by negative emotions (e.g., anxiety, frustration, moral indignation) that can translate into defensive motivation and resistance to adoption (Gabbiadini, Ognibene, et al. 2024).

Importantly, Intergroup Threat Theory specifies what is perceived to be at stake (identity and resources), but it does not fully capture how people anticipate the interpersonal costs of engaging with a powerful "other". For this reason, intergroup anxiety provides a complementary mechanism focused on anticipated interaction. Intergroup anxiety refers to the apprehension experienced when anticipating interaction with an outgroup, driven by uncertainty and expected negative outcomes (Stephan and Stephan 1985). It can arise even before any actual interaction, resulting in heightened vigilance. While emotions like fear, anger, and sadness often accompany threat appraisals in human–AI relations, they are not functionally equivalent. Most threat-related emotions are reactive, triggered by specific risks (e.g., fear of job loss, anger over injustice). In contrast, intergroup anxiety is anticipatory and self-focused, emerging from uncertainty about how one will be evaluated or treated by a perceived outgroup (Stephan and Stephan 1985; Mackie and Smith 2018). Unlike discrete emotions triggered by concrete threats, intergroup anxiety stems from concerns about social judgement and possible loss of status. Thus, intergroup anxiety can contribute to avoidance intentions as a distinct, interaction-related pathway, rather than as merely another expression of general negative affect. Applied to human–AI interaction, intergroup anxiety is especially relevant because AI can plausibly be construed as a high-status outgroup (McKee et al. 2023; Zhou et al. 2022). Generative AI's advanced capabilities challenge human uniqueness, eliciting both realistic threats (to jobs and resources) and symbolic threats (to identity and status; Stephan et al. 1999). In this context, intergroup anxiety can capture social-evaluative concerns—such as the fear that AI could assess, outperform, or diminish the individual's standing. More generally, intergroup anxiety may operate as a proximal pathway to avoidance-oriented responses (Zhou et al. 2022), potentially overlapping with yet conceptually distinct from threat appraisals. Empirical evidence supports the idea that anxiety influences technology adoption and human–robot or human–virtual agent interactions, shaping avoidance behaviour and resistance over and above general negative affect (Gabbiadini, Ognibene, et al. 2024; Sapru 2025).

Taken together, these perspectives clarify *why* generative AI can elicit threat-related appraisals and intergroup anxiety, and how these responses may translate into avoidance-oriented intentions. Yet, the intensity and direction of these responses likely depend not only on who the "source" is (AI vs. human) but also on how the source performs, namely, the perceived quality and accuracy of its outputs.

2.1 | Current Research Gap and Aims of the Present Study

Although threat perceptions are often explained by what AI can do, it is equally important to consider how well it does it. Generative models are not infallible; they frequently produce errors, inconsistencies, or factually incorrect content—so-called "hallucinations." These inaccuracies create a paradox. On the one hand, they undermine trust and fuel skepticism

about reliability, reinforcing algorithm aversion (Burton et al. 2020; Gaube et al. 2021; Jussupow et al. 2020; Mahmud et al. 2022). On the other hand, according to the Stereotype Content Model (SCM), visibly incorrect or implausible outputs could lower perceived competence, potentially shifting AI away from the high-competence/ambiguous-warmth ("envy") quadrant most associated with threat. In other words, obvious mistakes can undermine the impression that AI matches human skill, thereby reducing the perceived threat to human uniqueness and dampening negative emotions (Gabbiadini, Ognibene, et al. 2024).

This dual effect underscores a critical gap in the literature. While much attention has been devoted to concerns about bias, transparency, and reliability (Daneshjou et al. 2021; Ipsos Mori 2017), less is known about how the *quality of outputs* shapes emotional responses and threat perception. Do errors mitigate perceptions of AI as a threatening outgroup, or do they amplify distrust in ways that reinforce resistance? This unresolved question has important implications for both theory and practice, highlighting the need to disentangle the psychological consequences of AI's capabilities from those of its limitations.

Responses to AI are not driven by performance alone (Dietvorst et al. 2015): AI may be construed as an outgroup-like target, eliciting threat appraisals tied to concerns about human distinctiveness, status, and control and prompting initial negative emotions that shape subsequent meaning-making (Gabbiadini, Ognibene, et al. 2024; Stephan et al. 2016; Smith and Ellsworth 1985; Cottrell and Neuberg 2005).

In the context of generative AI, this points to a set of plausible, partly overlapping pathways linking exposure to downstream responses such as realistic threat, identity threat, and intergroup anxiety—responses that are conceptually relevant to reduced willingness to adopt AI. However, the literature does not yet provide a clear consensus on how these mechanisms relate to one another or which of them is most closely associated with AI adoption intentions. Moreover, it remains unclear whether output quality contributes meaningfully to these reactions: do errors primarily affect negative emotions, do they shape threat-related appraisals and anxiety, or might they attenuate perceived threat by undermining perceived competence?

The present study addresses these questions by examining responses to AI-generated video content, thereby extending prior work beyond text, images, and audio (see Gabbiadini, Ognibene, et al. 2024). We focus on how exposure to an AI (vs. human) source relates to negative emotional reactions and to downstream responses that have been highlighted as relevant in prior theory—namely identity threat, realistic threat (Gabbiadini, Ognibene, et al. 2024), and intergroup anxiety (Zhou et al. 2022)—and how these responses, in turn, relate to intentions to engage with AI. Given that these mechanisms can be theorized as interrelated, we do not assume a definitive causal ordering among them. Instead, we evaluate them jointly to clarify which pathways are most consistent with lower adoption intentions in our data. In addition, we explore whether output quality—particularly the presence of hallucinations—influences these associations, helping to disentangle the psychological consequences of AI's capabilities from those of its limitations. By doing so, the study contributes to a more nuanced understanding of the affective and intergroup correlates of public reactions to generative AI.

2.2 | Overview and Ethical Statement

The present study aimed to replicate and extend previous findings (see Gabbiadini, Ognibene, et al. 2024) by examining a set of affective and cognitive responses to generative AI and how these responses relate to future behavioural intentions. We adopt a staged conceptual organization, distinguishing an initial broad affective reaction from more differentiated responses that are commonly discussed in intergroup and technology-adoption literatures, while acknowledging that our post-exposure measures were assessed at the same time point and therefore do not establish temporal ordering among mediators.

H1. *The first hypothesis concerns negative emotional reactions to AI-generated content. Prior research has often relied on high-quality AI stimuli (see* Gabbiadini, Ognibene, et al. 2024*), leaving open whether typical errors or distortions intensify negative reactions or whether reactions are primarily driven by source attribution (AI vs. human). Accordingly, we hypothesized that negative emotions would differ as a function of content source (AI vs. human). Given the limited and mixed evidence on how errors in generative outputs shape evaluations, we did not formulate a directional hypothesis regarding output quality or its interaction with source.*

H2. *The second hypothesis concerns how negative emotional reactions relate to more differentiated responses that are theoretically relevant to disengagement from AI. Drawing on intergroup approaches, reactions to AI can involve distinct threat appraisals (*Gabbiadini, Ognibene, et al. 2024*) as well as intergroup anxiety (Zhou et al. 2022), intended as a negative social comparison when engaging with an outgroup (*Stephan and Stephan 1985*). Consistently, we hypothesized that higher negative emotions would be associated with higher realistic threat, identity threat, and intergroup anxiety. This hypothesis is not intended as evidence of strict temporal sequencing, but as a theory-guided expectation about how these constructs tend to cohere in responses to an ambiguous and potentially consequential target, such as a generative AI system.*

H3. *The third hypothesis concerns whether these downstream responses are linked to lower willingness to engage with AI tools in the future. Building on prior evidence that negative reactions to AI co-occur with heightened defensiveness and resistance* (e.g., Gabbiadini, Ognibene, et al. 2024; Xu et al. 2024; Zhou et al. 2022), *we expected that realistic threat, identity threat, and intergroup anxiety would be negatively associated with intentions to use AI. We also examined whether these associations hold when the downstream responses are considered simultaneously, thereby clarifying which pathways are most consistent with lower intentions in our data.*

Overall, these hypotheses extend previous work by integrating behavioural intentions with multiple theoretically grounded

mechanisms (threat appraisals and intergroup anxiety) and by examining the role of output quality alongside source attribution in a novel modality (AI-generated video). The study was approved by the local Department of Psychology's minimal risk research committee. All procedures were in accordance with APA ethical guidelines and the Declaration of Helsinki. Informed consent was obtained prior to participation.

3 | Materials and Methods

3.1 | Participants

To determine the minimum sample size required to achieve sufficient statistical power (≥ 0.80) for the proposed mediation model, we conducted an a priori Monte Carlo simulation using R (R Core Team 2022) with the lavaan package (Rosseel 2012). Our hypothesized model comprised one independent variable (X), a first-stage mediator (M1), three second-stage parallel mediators (M2a, M2b, M2c), and a dependent variable (Y). Data were simulated under the assumption of normally distributed variables. Considering previous research (see Gabbiadini, Ognibene, et al. 2024), we assumed non-standardized path coefficients of 0.52 for the effect of X on M1, and a medium effect of 0.25 for the causal effect of M1 on each of M2a, M2b, and M2c, as well as 0.25 for the effects of each mediator on Y. For each simulation, we generated datasets across a range of sample sizes (from 100 to 200, in increments of 1) and conducted 1000 replications per sample size. In each replication, the model was estimated using maximum likelihood estimation, and the significance of each path was determined using an alpha level of 0.05. The empirical power for each path was computed as the proportion of replications in which the corresponding effect was statistically significant. Based on these parameters derived from previous work (see Gabbiadini, Ognibene, et al. 2024), we determined that a minimum sample size of 135 participants was required to achieve 80% power across all key paths in the model. This simulation-based a priori power analysis ensured that our study design was adequately powered to detect the hypothesized mediation effects.

After obtaining participants' informed consent, participants were presented with a set of items to measure sociodemographic variables (age, gender, level of education). A total of 328 participants accessed the survey. Of these, 153 either did not provide informed consent or completed less than 30% of the questionnaire and were therefore excluded. Among the remaining 175 participants, 37 failed the manipulation check (i.e., "The video you just saw was created by an AI/was filmed by a human operator") and were removed from subsequent analyses. Therefore, the final sample consisted of 138 participants. Of these, 29 (21.0%) identified as male and 109 (79.0%) as female. The vast majority reported Italian nationality (98.6%), while one participant identified as Argentine (0.7%) and one as Brazilian (0.7%). Regarding education, 2.9% held a lower secondary school diploma, 2.2% a professional diploma, and 24.6% a high school diploma. Furthermore, 17.4% held a bachelor's degree, 40.6% a master's or single-cycle degree, 9.4% a postgraduate master's, and 2.9% a PhD.

3.2 | Procedure and Measures

Participants were told that the study would take approximately 15 min. Before the experimental manipulation, participants were asked to complete a set of descriptive measures. First, they reported their general familiarity with digital technologies ("How familiar are you, in general, with new digital technologies?"; 1 = no familiarity at all, 7 = complete familiarity; M = 4.64, SD = 1.49). They then assessed their level of experience with AI technologies ("How would you rate your level of experience/familiarity with AI technologies?"; 1 = no experience, 2 = beginner, 3 = competent, 4 = expert, 5 = very experienced; M = 1.99, SD = 0.82) and their general knowledge of AI ("How would you rate your knowledge of AI technologies?"; 1 = no knowledge, 2 = limited, 3 = average, 4 = good, 5 = very good; M = 2.36, SD = 0.88).

Participants were also asked about their use of AI-based generative tools for media creation ("Have you ever used applications for AI-based generation of video, audio, or images?"; 1 = never used, 5 = very frequent use; M = 1.78, SD = 0.96). In addition, they reported their frequency of ChatGPT use ("How often do you use ChatGPT or similar AI applications?"; 1 = never used, 5 = very frequent use; M = 2.14, SD = 1.21), which was included as an indicator of participants' general engagement with generative AI technologies, given the widespread diffusion and accessibility of ChatGPT as a representative example of such tools.

Consistent with previous research examining psychological responses to emerging technologies (Schepman and Rodway 2020), participants' attitudes toward AI were also assessed using the *General Attitudes Toward Artificial Intelligence Scale*. Sample items included "I think Artificial Intelligence is dangerous" and "I am impressed by what Artificial Intelligence can do," anchored on a 7-point scale (1 = completely disagree; 7 = completely agree).

Following the preliminary measures, participants were randomly assigned to one of four experimental conditions in a 2 (source: AI vs. human video maker) × 2 (accuracy: presence of errors vs. professional) between-subjects design. In each condition, participants were instructed to watch a short neutral video of equal length (i.e., 9 s) depicting a basketball entering a hoop. This type of video was selected because it was commonly used to demonstrate the capabilities of generative AI systems and, at the time of the study, represented a domain where such algorithms often produced *hallucinations* or unrealistic outputs (see the additional materials for the videos used in the experimental manipulation).

In the error conditions, participants viewed either an AI-generated video in which the ball passed unrealistically through the rim of the basket (AI, hallucination condition), or a video produced by a professional video maker that was intentionally pixelated and shaky (human, low accuracy). The rationale behind this choice was that, while in the AI condition the manipulation of "low accuracy" could naturally rely on the presence of a typical AI hallucination, the only way to balance this manipulation for the human condition was to present a technically flawed video (pixelated and unstable) in order to represent poor quality production.

In the professional conditions, the same high-quality video of a basketball cleanly entering the hoop was presented; however, participants were informed that this identical video had either been generated by an AI system (SORA; OpenAI 2024) or created by a professional video maker, depending on the condition. After watching the video corresponding to their assigned condition, participants were asked to complete a series of scales measuring the dependent variables. Since research has shown that the capabilities of generative AI can produce strong emotional reactions (Gabbiadini, Ognibene, et al. 2024), participants were asked to report the emotions elicited by AI. To ensure that responses referred specifically to the *source* of the video rather than to its visual content, participants were instructed as follows: "What emotions does knowing that the video you have just seen was created by a [professional video maker/AI] evoke in you? Please note that we are not asking about the emotions you felt while watching the content of the video itself, but about the emotions generated by knowing who produced it." Negative emotions (anger, fear, disgust, anxiety, sadness, desire, happiness, and joy; positive items were reverse-scored so that higher scores reflect greater overall negative affect) were then assessed using items adapted from Harmon-Jones et al. (2016), rated on a 7-point scale (1 = not at all; 7 = very much).

Afterwards, realistic and symbolic threats posed by AI were assessed by adapting the scale proposed by Yogeeswaran et al. (2016). Five items measured the extent to which AI was perceived as a realistic threat to humans (e.g., "The increased use of AI in our everyday life is causing more job losses for humans"), while another five assessed symbolic threats to human identity and distinctiveness (e.g., "Technological advancements in the area of AI are threatening human uniqueness"). All items were rated on a 7-point scale (1 = strongly disagree; 7 = strongly agree). Moreover, intergroup anxiety was measured with the scale developed by Stephan and Stephan (1985), adapted to the context of human–AI interaction. Participants were asked to indicate the degree of discomfort they would feel in interactions with AI technologies, using a 7-point scale (1 = not at all anxious; 7 = very anxious). To control for potential order effects, the presentation of the two measures (threat and anxiety) was randomized across participants.

Finally, participants' intentions to use AI tools were assessed with a set of ad hoc items. Examples included: "I intend to use an AI-based software in the near future," "I believe I will use Artificial Intelligence to perform some of my tasks in the near future," and "I intend to use an AI-based software to generate videos in the near future." All items were anchored on a 7-point scale (1 = completely disagree; 7 = completely agree).

At the end of the survey, participants were thanked for their participation, provided with a written debriefing explaining the aims of the study, and given the researchers' contact information for any further questions.

4 | Results

4.1 | Preliminary Analyses

All the following statistical analyses were performed using SPSS software (v.29) and RStudio (v.2024.12.1 + 563). Means, standard deviations, and reliability for all measures are reported in Table 1.

Participants reported a relatively high level of general familiarity with digital technologies (M = 4.64, SD = 1.49). In contrast, their experience with AI technologies was more limited. Specifically, self-reported experience with AI tools averaged 1.99 (SD = 0.82), while general knowledge of AI technologies was slightly higher (M = 2.36, SD = 0.88). Use of generative AI tools (e.g., for creating video, audio, or images) was low (M = 1.78, SD = 0.96), and the frequency of using ChatGPT or similar AI services was also limited (M = 2.14, SD = 1.22). Despite this, participants expressed generally positive attitudes toward AI, with a mean score of 4.16 (SD = 0.76).

Correlation analyses (see Table 1) revealed that negative emotions were significantly associated with all three mediators of interest, namely intergroup anxiety, realistic threat, and symbolic threat. In turn, intergroup anxiety, realistic threat, and symbolic threat were negatively associated with intentions to use AI tools.

4.2 | Analysis of Variance

Before testing the hypothesized mediation model, we conducted a series of ANCOVAs to examine whether the effects of our experimental manipulation varied as a function of the source (AI vs. Human), the quality of the output (Good vs. Faulty), or their interaction. This preliminary step was intended to disentangle whether responses to AI were driven primarily by the generative source of the content, by the quality of the material produced, or by a combination of the two. All analyses controlled for age, gender, and study level.

A set of 2 (source: AI vs. human) × 2 (quality: good vs. faulty) ANCOVAs was conducted for negative emotions, identity threat, realistic threat, intergroup anxiety, and intentions, controlling for age, gender, and education.

For negative emotions, the main effect of source was significant, $F(1, 131) = 25.03$, $p < 0.001$, $\eta p^2 = 0.160$: participants in the AI condition reported higher negative emotions (M = 2.65, SD = 1.08) than those in the human condition (M = 1.88, SD = 0.52). Neither the main effect of quality, $F(1, 131) = 0.03$, $p = 0.871$, $\eta p^2 < 0.001$, nor the source × quality interaction, $F(1, 131) = 3.45$, $p = 0.065$, $\eta p^2 = 0.026$, reached significance. Gender emerged as a significant covariate, $F(1, 131) = 5.77$, $p = 0.018$, $\eta p^2 = 0.042$, whereas age and education were not ($ps \geq 0.168$).

For identity threat, the main effect of source was significant, $F(1, 131) = 6.48$, $p = 0.012$, $\eta p^2 = 0.047$, with higher identity threat in the AI condition (M = 4.20, SD = 1.45) than in the human condition (M = 3.53, SD = 1.37). The effect of quality and the source × quality interaction were not significant (both $ps \geq 0.597$). Education was a significant covariate, $F(1, 131) = 8.05$, $p = 0.005$, $\eta p^2 = 0.058$. Age and gender were not significantly associated with the dependent variable ($ps \geq 0.281$). Instead, when considering realistic threat, neither source, $F(1, 131) = 1.29$, $p = 0.259$, $\eta p^2 = 0.010$, nor quality, $F(1, 131) = 1.71$, $p = 0.194$, $\eta p^2 = 0.013$, nor their interaction, $F(1, 131) = 0.51$, $p = 0.475$, $\eta p^2 = 0.004$, were significant predictors. Age, $F(1, 131) = 5.32$, $p = 0.023$, $\eta p^2 = 0.039$, and education, $F(1, 131) = 12.04$, $p < 0.001$, $\eta p^2 = 0.084$, emerged as significant covariates. When considering intergroup anxiety as an outcome, no significant effects were observed for source, $F(1, 129) = 2.43$, $p = 0.122$, $\eta p^2 = 0.018$, quality, $F(1, 129) = 0.64$,

TABLE 1 | Correlations among study variables.

	α	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Age	—	29.42	12.76	—													
2. Gender	—	1.79	0.41	−0.123	—												
3. Education	—	4.30	1.25	−0.004	0.156	—											
4. Digital familiarity	—	4.64	1.49	−0.100	−0.284***	−0.137	—										
5. AI experience	—	1.99	0.82	−0.205*	−0.190*	−0.055	0.539***	—									
6. AI knowledge	—	2.36	0.88	−0.130	−0.293***	−0.146	0.579***	0.740***	—								
7. AI use for media creation	—	1.78	0.96	−0.164	−0.254***	−0.217*	0.474***	0.645***	0.581***	—							
8. ChatGPT use	—	2.14	1.22	−0.253**	−0.247**	−0.081	0.463***	0.686***	0.576***	0.722***	—						
9. Attitude	0.83	4.16	0.76	−0.097	−0.129	0.115	0.246**	0.524***	0.361***	0.473***	0.564***	—					
10. Negative emotions	0.92	1.59	1.09	0.037	−0.198*	−0.212*	0.081	0.031	0.205*	0.062	−0.009	−0.221**	—				
11. Positive emotions	0.84	2.10	1.40	−0.152	0.008	−0.127	0.094	0.123	0.112	0.214*	0.113	0.071	−0.076	—			
12. Anxiety	0.92	3.70	1.22	0.215*	0.011	−0.156	−0.125	−0.377***	−0.165	−0.275**	−0.359***	−0.690***	0.395***	−0.101	—		
13. Realistic threat	0.82	3.73	1.28	0.203*	−0.137	−0.303***	0.007	−0.179*	0.021	−0.113	−0.171*	−0.498***	0.374***	−0.130	0.657***	—	
14. Identity threat	0.87	3.90	1.45	0.112	−0.121	−0.262**	−0.080	−0.223**	−0.045	−0.154	−0.243**	−0.488***	0.452***	−0.023	0.546***	0.802***	—
15. Intentions	0.89	3.55	1.73	−0.100	−0.088	0.075	0.280***	0.403***	0.217*	0.401***	0.451***	0.626***	−0.204*	0.081	−0.548***	−0.271**	−0.222**

Note: $N = 138$. Gender was coded as 0 = male, 1 = female. Correlations involving gender represent Spearman's rho coefficients. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

$p = 0.426$, $\eta p^2 = 0.005$, or their interaction, $F(1, 129) = 1.04$, $p = 0.310$, $\eta p^2 = 0.008$. Age was a significant covariate, $F(1, 129) = 6.98$, $p = 0.009$, $\eta p^2 = 0.051$, whereas gender and education were not (both $ps \geq 0.063$).

Finally, for intentions, there were no significant effects of source, $F(1, 127) = 0.40$, $p = 0.529$, $\eta p^2 = 0.003$, quality, $F(1, 127) = 0.12$, $p = 0.729$, $\eta p^2 = 0.001$, or their interaction, $F(1, 127) = 0.01$, $p = 0.941$, $\eta p^2 < 0.001$. No covariate effects were observed (all $ps \geq 0.176$).

Across the outcome measures, the results showed a selective pattern. Participants in the AI-source condition reported significantly higher negative emotions than those in the human-source condition, whereas quality (good vs. faulty) did not yield significant main effects and did not reliably interact with source. For the more differentiated responses and behavioural intentions, direct effects of the experimental factors were weaker: identity threat showed a main effect of source (higher under AI than human), whereas realistic threat, intergroup anxiety, and intentions did not differ significantly as a function of source, quality, or their interaction.

Given that the ANCOVAs showed significant effects only for source attribution, these preliminary results indicate that the AI vs. human source was the primary driver of broad negative affect. This empirically supports our decision to focus subsequent analyses on the process through which this initial affective response relates to more specific downstream appraisals and, ultimately, to intentions. Accordingly, we proceeded to test a staged process model using PROCESS Model 81, with source (AI vs. human) as the predictor, negative emotions as the first-stage mediator, and identity threat, realistic threat, and intergroup anxiety specified as second-stage mediators operating in parallel and conceptually closer to intentions. This approach allows us to assess whether the impact of source attribution on intentions is better captured by indirect pathways—via an initial negative emotional reaction and subsequent threat- and interaction-related responses—rather than by direct mean differences on each downstream outcome in isolation.

4.3 | Mediational Analyses: Direct and Indirect Effects

A staged mediation model (PROCESS, Model 81; Hayes 2018) was estimated to examine whether the effect of the experimental condition (X) on participants' intentions to use AI (Y) was mediated by negative emotion (M1) in a first stage, and by anxiety (M2a), realistic threat (M2b), and identity threat (M2c) entered in parallel as downstream mediators. The model included age, gender, and educational level as covariates. All model estimates, including direct and indirect effects, are reported in Table 2 (See also the Supporting Information for additional robustness checks, including alternative model specifications, a model re-estimated with Quality and the Source × Quality interaction entered as covariates, and a further model re-estimated after removing the anxiety item from the negative-emotions composite to minimize conceptual overlap with intergroup anxiety; in all cases, the overall pattern of results remained substantively unchanged).

The experimental manipulation significantly increased negative emotions, which were in turn positively associated with realistic threat, identity threat, and intergroup anxiety. By contrast, the manipulation had no direct effect on these downstream responses. When predicting intentions, intergroup anxiety emerged as the only significant predictor, whereas negative emotions, realistic threat, and identity threat were not significant. The direct effect of the experimental manipulation on intentions was not significant, and the total effect was also not significant (see Figure 1).

Pairwise bootstrap contrasts revealed that the only significant indirect effect, linking the experimental condition to reduced intentions via negative emotions and intergroup anxiety, was also significantly stronger than all alternative pathways tested. Specifically, it was larger in magnitude than the indirect path through negative emotions and realistic threat ($\Delta = -0.08$, SE = 0.038, 95% CI [−0.1704, −0.0182]), the path through negative emotions and identity threat ($\Delta = -0.11$, SE = 0.05, 95% CI [−0.2290, −0.0288]), and the path through negative emotions alone ($\Delta = -0.08$, SE = 0.04, 95% CI [−0.1677, −0.0229]). All other contrasts were non-significant, with confidence intervals including zero. This pattern supports the conclusion that the dominant mechanism linking exposure to AI-generated content and reduced intentions to use AI operates through a sequential process of affective reactivity and intergroup anxiety, rather than through threat-based appraisals alone.

5 | Discussion

The present study provides a comprehensive investigation into how exposure to AI-generated content elicits emotional and cognitive responses, and how these responses relate to behavioural intentions.

One of the primary questions concerned the role of AI output quality versus source attribution in shaping early affective reactions (H1). While the present study did not experimentally manipulate performance quality, the inclusion of AI-generated videos containing typical hallucinations allowed for an indirect assessment of this factor. The results suggest that negative emotional responses are driven primarily by the source of the content (i.e., AI versus human) rather than by the presence of errors or distortions in the output (e.g., AI-hallucinations). Importantly, participants responded with negative emotions even when the AI outputs were technically flawless. This suggests that it is not the quality of the product that determines affective reactions, but the meaning attributed to the source. Knowing that a content was created by AI appears sufficient to generate emotional discomfort independently of accuracy. These results align with theories of AI aversion, which posit that the label "AI" activates cognitive representations of non-humanness, artificiality, and lack of authenticity, translating into negative emotional responses (Burton et al. 2020; Gaube et al. 2021; Jussupow et al. 2020; Mahmud et al. 2022). Although flawed outputs may signal fallibility and potentially attenuate threat appraisals, even imperfect AI content elicited stronger negative affect than human-generated material.

Yet, several other interpretations can be proposed. For example, exposure to AI's ability to generate seemingly realistic scenes,

TABLE 2 | Unstandardized and standardized coefficients for the mediation model (PROCESS Model 81).

Predictor	Outcome	β	b	SE	*t*	*p*	95% CI	R^2
Experimental condition	Negative emotions	0.33	0.29	0.07	4.14	< 0.001	[0.15, 0.43]	0.19
Age		0.04	0.00	0.01	0.50	0.615	[−0.01, 0.02]	
Gender		−0.18	−0.41	0.19	−2.13	0.035	[−0.79, −0.03]	
Education		−0.12	−0.09	0.06	−1.43	0.155	[−0.21, 0.03]	
Experimental condition	Realistic threat	0.02	0.03	0.10	0.26	0.794	[−0.17, 0.22]	0.24
Negative emotions		0.32	0.43	0.11	3.76	< 0.001	[0.20, 0.65]	
Age		0.19	0.02	0.01	2.44	0.016	[0.00, 0.03]	
Gender		0.01	0.03	0.25	0.12	0.902	[−0.47, 0.53]	
Education		−0.25	−0.26	0.08	−3.14	0.002	[−0.42, −0.09]	
Experimental condition	Identity threat	0.08	0.11	0.11	1.03	0.307	[−0.10, 0.33]	0.24
Negative emotions		0.36	0.54	0.13	4.16	< 0.001	[0.28, 0.79]	
Age		0.09	0.01	0.01	1.16	0.246	[−0.01, 0.03]	
Gender		0.02	0.05	0.29	0.19	0.850	[−0.51, 0.62]	
Education		−0.21	−0.24	0.09	−2.63	0.010	[−0.43, −0.06]	
Experimental condition	Intergroup anxiety	−0.07	−0.08	0.09	−0.81	0.418	[−0.26, 0.11]	0.23
Negative emotions		0.42	0.54	0.11	4.92	< 0.001	[0.32, 0.75]	
Age		0.20	0.02	0.01	2.56	0.011	[0.00, 0.03]	
Gender		0.14	0.41	0.24	1.67	0.097	[−0.07, 0.89]	
Education		−0.12	−0.12	0.08	−1.52	0.131	[−0.28, 0.04]	
Experimental condition	Intentions to use AI	0.11	0.18	0.12	1.42	0.160	[−0.07, 0.42]	0.33
Negative emotions		−0.08	−0.14	0.16	−0.85	0.396	[−0.46, 0.18]	
Realistic threat		0.13	0.17	0.19	0.91	0.365	[−0.21, 0.55]	
Identity threat		0.02	0.02	0.15	0.15	0.878	[−0.28, 0.33]	
Intergroup anxiety		−0.62	−0.87	0.14	−6.03	< 0.001	[−1.16, −0.59]	
Age		−0.01	0.00	0.01	−0.14	0.892	[−0.02, 0.02]	
Gender		−0.08	−0.35	0.33	−1.06	0.291	[−1.00, 0.30]	
Education		0.03	0.04	0.11	0.41	0.685	[−0.17, 0.26]	

Indirect effects	b	β	SE	95% CI
Experimental condition → negative emotions → intentions	−0.04	−0.03	0.05	[−0.13, 0.05]
Experimental condition → realistic threat → intentions	0.00	0.00	0.03	[−0.05, 0.07]
Experimental condition → identity threat → intentions	0.00	0.00	0.03	[−0.05, 0.06]
Experimental condition → intergroup anxiety → intentions	0.07	0.04	0.08	[−0.09, 0.22]
Experimental condition → negative emotions → realistic threat → intentions	0.02	0.01	0.03	[−0.03, 0.08]
Experimental condition → negative emotions → identity threat → intentions	0.00	0.00	0.03	[−0.05, 0.06]
Experimental condition → negative emotions → intergroup anxiety → intentions	−0.14	−0.09	0.04	[−0.23, −0.07]
Total effect	0.10	0.06	0.14	[−0.17, 0.36]

Note: Analyses are based on $N = 134$ due to missing responses on certain scales.

even when these outputs contain occasional hallucinations, may activate cognitions about how such capabilities could be used to create convincing *deepfakes* (i.e., AI-generated media in which a real person's appearance or voice is realistically fabricated or manipulated without their involvement). Another interpretation could be that the quality manipulations (e.g.,

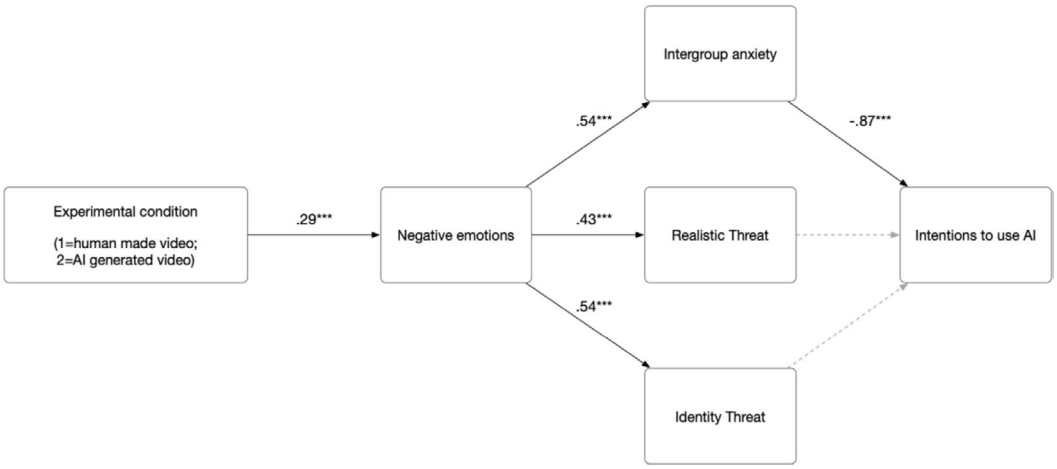

FIGURE 1 | Simplified mediation model (PROCESS Model 81). Effects of the experimental condition on intentions to use AI through elicited negative emotions and their impact on perceived realistic threat, symbolic threat, and intergroup anxiety. Unstandardized regression coefficients are reported (see Table 2 for standardized coefficients). ***$p < 0.001$.

pixelation and shakiness for human videos, physical impossibility for AI) were not perceived as psychologically equivalent in terms of error salience or significance. Alternatively, the experimental context may have led participants to focus primarily on the provided label ("AI" vs. "human"), with source attribution overshadowing any impact of output quality—possibly amplified by the explicitness of the instructions. Finally, it is possible that the brief exposure and limited personal relevance of the stimuli reduced the sensitivity of the measures to detect subtler quality effects.

Taken together, these considerations suggest that source attribution may operate as a particularly strong trigger of early negative affect. Future studies might benefit from more immersive or ecologically valid manipulations of quality, as well as the inclusion of manipulation checks to clarify how participants interpret and evaluate different types of errors.

Apart from the effect of source attribution and quality, the study examined how these initial negative emotional reactions relate to more differentiated downstream responses (H2). In this study, even in the absence of direct interaction, participants reacted with heightened intergroup anxiety when exposed to AI-generated videos, suggesting that intergroup anxiety may function as a key anticipatory, social-evaluative mechanism linking early affective discomfort to behavioural intentions. In this regard, recent findings indicate that, when confronted with AI, individuals' resistance and avoidance are driven less by abstract collective concerns and more by perceived individual-level implications—such as anxiety over personal evaluation, fears of being treated as interchangeable, and doubts about the recognition of one's unique characteristics in human–AI interactions (Zhou et al. 2022). These individual-level appraisals may foster intergroup anxiety, making the psychological boundary between "self" and "machine" particularly salient in shaping responses to generative AI.

At the same time, negative emotions were also associated with stronger perceptions of both realistic threat (e.g., concerns about jobs, resources, and control) and identity threat (e.g., challenges to human distinctiveness, status, and autonomy), consistent with the idea that an early global affective signal is quickly tied to more specific meanings about what is at stake. However, in the process model, these threat appraisals did not uniquely predict lower intentions to use AI once intergroup anxiety was taken into account, suggesting that they may operate as part of a broader threat "landscape" that co-occurs with anxiety rather than as independent pathways. Consistent with a staged process, generalized negative emotions appear to capture an early affective signal, whereas intergroup anxiety reflects a more proximal, interaction-focused response that is conceptually closer to avoidance.

Finally, the study investigated whether downstream responses predict future intentions to use AI systems (H3). The results indicate that intergroup anxiety plays a particularly central role in translating exposure to AI into reduced willingness to engage with AI technologies.

Taken together, these findings support the view that generative AI may be perceived as a socially relevant outgroup capable of challenging human uniqueness and competence, consistent with prior evidence showing that AI performing human-like creative tasks evokes heightened negative appraisal (Gabbiadini, Ognibene, et al. 2024; Xu et al. 2024). From the perspective of Intergroup Threat Theory (Stephan et al. 2008), symbolic threats (e.g., challenges to identity, creativity, and social status) and realistic threats (e.g., employability, autonomy, and control) may contribute to early negative affect, but it is the anticipatory anxiety surrounding potential interaction with AI that most directly informs avoidance-oriented intentions. It is also possible that negative emotions are largely channelled through intergroup anxiety, which then directly shapes behavioural intentions, or that the threat constructs measured here share substantial variance and therefore do not emerge as independent mediators once modelled simultaneously. These findings underscore the need for further theoretical refinement—clarifying the unique versus shared contributions of different threat-related and affective constructs—and for studies using larger samples or alternative analytic strategies (e.g., latent variable modelling, longitudinal designs) to better disentangle the causal architecture of emotional responses to AI.

Integrating these findings with technology adoption models such as TAM (Davis 1989) and UTAUT (Venkatesh et al. 2003) clarifies why technically valid outputs do not automatically increase adoption intentions: participants' affective and social-evaluative responses appear to override perceived usefulness or ease of use. Specifically, intergroup anxiety may reduce the predictability of interacting with AI, raising anticipated cognitive and emotional effort. This supports the idea that discomfort associated with the AI source can outweigh technical quality or output accuracy when it comes to adoption intentions.

These results can also be interpreted in light of research on algorithm aversion. While there is evidence of algorithm appreciation—the tendency for individuals to trust algorithmic recommendations more than human advice in specific scenarios (Logg et al. 2019)—the more pervasive pattern observed among lay populations is that of algorithm aversion. In this context, people generally display a preference for human guidance over AI-generated recommendations and remain reluctant to trust or adopt AI systems, even when these algorithms have been shown to perform at levels comparable to or exceeding human experts (Burton et al. 2020; Gaube et al. 2021; Jussupow et al. 2020; Mahmud et al. 2022).

Several studies illustrate this phenomenon across various domains. For instance, Riedl et al. (2024) found that patients consistently prefer interacting with a human doctor above all other options, followed by a human doctor who is supported by AI. The least preferred scenario, by a significant margin, was interacting with an AI system alone. This hierarchy was evident not only in general preference, but also across a range of patient outcomes: trust in the provider, willingness to disclose personal health information, adherence to treatment, and satisfaction were all highest when the provider was human. Conversely, distrust and concerns about privacy invasion peaked in the AI-alone condition. Notably, these differences were most pronounced in psychiatric contexts, where empathy, trust, and the sense of human connection are particularly salient, thus making exclusive AI involvement especially problematic. These findings are in line with the pattern observed in our study, where social-evaluative discomfort—not system performance—emerged as the most robust predictor of reduced behavioural intentions.

A similar dynamic is observed in the field of human resources. Research shows that recruiters typically report greater trust in their own professional judgement than in algorithmic recommendations, even as they recognize the potential for AI to enhance recruitment processes, efficiency, and fairness (Ore and Sposato 2022). The tension between acknowledging the objective advantages of algorithmic decision-making and the reluctance to delegate final authority to AI systems exemplifies the psychological complexity of algorithm aversion. It underscores that factors such as perceived empathy, accountability, and the anticipation of being evaluated or misrecognized by a non-human agent remain crucial to human acceptance and integration of AI across domains.

Nevertheless, while these findings highlight how AI can elicit patterns of aversion and distrust similar to those experienced toward human outgroups, it is crucial to recognize that AI agents fundamentally differ from traditional social groups. Unlike human social groups, AI lacks shared history, intentionality, agency, and a sense of collective identity, which may render the boundaries between groups more ambiguous. While highly competent AI systems may indeed trigger both symbolic threats (e.g., challenging human uniqueness, creativity, or autonomy) and realistic threats (e.g., concerns over professional security or job displacement), the core driver of defensive reactions may lie in the affective discomfort and anticipatory anxiety related to imagined interaction, rather than in zero-sum intergroup conflict.

Noteworthy, a growing literature suggests that, because AI does not possess genuine agency, consciousness, or a shared social trajectory, individuals may not spontaneously perceive it as a social group in the sense outlined by Intergroup Threat Theory (Złotowski et al. 2017; Bartneck et al. 2009). Instead, threat responses to AI are likely to be more diffuse, less rooted in established intergroup schemas, and shaped by alternative psychological processes. These include mind perception (the tendency to attribute mental states to non-human agents), anthropomorphism (assigning human-like qualities to machines), and technological essentialism (perceiving technology as inherently different from biological entities; Waytz et al. 2010). This suggests that intergroup anxiety in response to AI may operate drawing on the conventional social-cognitive mechanisms of group categorization, acting more as a response to ambiguous agency or uncanny intentionality than to clear outgroup membership. Such distinctions highlight the need for future research to deepen the conditions and the individual differences leading people to categorize AI as a "social outgroup" capable of eliciting intergroup-like threat and anxiety. This also calls for a more integrative theoretical approach, combining ITT with complementary models of social cognition, mind perception, and human–machine interaction.

Our results are consistent with this account. Participants exposed to AI-generated content—even when technically flawless—reported lower future intentions to engage with AI systems, mirroring the mechanisms of algorithm aversion. In this study, negative affect and intergroup anxiety appear to function as key psychological intermediaries through which algorithmic competence is translated into avoidance behaviour. In other words, the mere knowledge that the content was produced by an AI system triggers early negative affect, followed by anticipatory anxiety about future engagement, which in turn reduces willingness to adopt or interact with AI. These results underscore that algorithm aversion is not merely about errors or performance quality, but about the perceived social and evaluative implications of AI—particularly when AI is framed as a capable, autonomous, and human-like agent.

5.1 | Limitations and Future Directions

Despite the theoretical and empirical contributions of this study, several important limitations should be considered. First, all mediators and outcomes were assessed at the same post-exposure time point. As a result, the staged process that we propose, where negative emotions precede threat appraisals and intergroup anxiety, should be interpreted as a theoretically guided model of associations, not as definitive evidence of temporal

or causal ordering. Future research should therefore test these pathways using designs that allow temporal precedence to be established more rigorously, for example by measuring mediators and outcomes at multiple time points or by experimentally manipulating specific mediators (e.g., anxiety, threat salience). Such approaches would provide a stronger empirical basis for inferring causal mechanisms in emotional and intergroup responses to generative AI.

Second, the design involved brief, laboratory-based exposure to videos labelled as either AI- or human-generated. While this approach ensured a high degree of experimental control, it does not reflect the ongoing and evolving nature of real-world human–AI interactions. In daily life, repeated exposure to AI tools, gradual habituation, and the accumulation of experiences can lead individuals to recalibrate their expectations and possibly develop more nuanced or even positive attitudes toward AI over time. The short, decontextualized exposure used here may therefore capture only initial, surface-level reactions rather than the dynamic shifts that occur with increased familiarity. Future research would benefit from longitudinal or repeated-exposure designs, enabling a more realistic exploration of how threat perceptions, anxiety, and acceptance of AI evolve with sustained interaction.

Third, the study relied exclusively on self-report measures to assess threat perception and anxiety. While these constructs are inherently subjective, this methodological choice limits the interpretability of the findings and may not fully capture the complexity of participants' responses. Self-reports are susceptible to social desirability bias and may fail to detect subtle or implicit affective reactions. Future studies could enhance validity by integrating behavioural tasks that assess actual engagement with or avoidance of AI tools.

Fourth, the generalizability of the results is limited by the specific sample and sociocultural context. Perceptions of AI are shaped not only by individual experiences, but also by broader cultural narratives, policy discourse, and labour market conditions. In societies where AI is promoted as a symbol of progress and national achievement, people may perceive AI as an opportunity rather than a threat. Conversely, in contexts marked by economic uncertainty or job insecurity, the same technologies may intensify feelings of vulnerability. Furthermore, demographic factors such as age, occupation, and technological literacy are likely to moderate these responses. To address these complexities, future research should prioritize cross-cultural comparisons and consider socio-demographic moderators to better capture the diversity of human–AI relations.

Finally, the ecological validity of the experimental manipulation warrants consideration. The use of brief, explicitly labelled videos may not be sufficient to activate the deeper social-cognitive processes posited by theories of intergroup threat. In real-world settings, interactions with AI typically unfold over time, are embedded in complex social and organizational environments, and involve richer feedback and interpersonal cues. The minimal and artificial nature of the experimental stimuli may have prompted only transient or superficial reactions, potentially inflating the salience of the "AI" label due to explicit instructions. Future studies should strive for greater ecological validity by employing longer, interactive tasks, naturalistic scenarios, or even field studies to determine whether the mechanisms observed in the lab are robust and generalizable to real-world encounters with AI.

With regard to ecological validity, another limit is represented by how we operationalized output quality. In our design, low-quality AI videos involved a semantic/physical violation (a visually plausible but physically impossible shot), whereas low-quality human videos were degraded via pixelation and camera shakiness. This asymmetry reflects a broader methodological challenge: typical generative AI "errors" in video are hallucinations (violations of what happens in the scene), whereas human errors are more often technical (how the scene is recorded). For ecological reasons, we chose stimuli that were representative of these different error types, as this was arguably the most realistic way to contrast "typical" AI versus human imperfections. However, the two low-quality conditions were therefore not fully equivalent; therefore, results should be interpreted with caution. Future studies should aim for more closely matched quality manipulations across sources, ideally combined with manipulation checks of perceived error type and severity.

6 | Conclusion

Overall, the present findings underscore that generative AI is not perceived solely as a neutral technological tool, but as a socially meaningful outgroup that activates cognitive and affective mechanisms typically reserved for human social categories. Exposure to AI-generated content—even when technically flawless—can elicit both symbolic and realistic threats, negative emotions, and anticipatory intergroup anxiety. This tendency aligns with broader patterns of algorithm aversion, where individuals prefer human input over AI guidance not necessarily due to error rates, but because of the perceived social and evaluative significance of interacting with a competent, autonomous agent. These affective responses play a pivotal role in shaping behavioural intentions, influencing whether people embrace or avoid AI systems.

Given AI's growing presence in everyday life—including communication, information-seeking, and professional decision-making—these psychological mechanisms have important implications for the design and deployment of AI technologies. The findings suggest that it is insufficient to optimize AI solely for technical performance or efficiency; human affective and cognitive responses, such as perceived threat and social-evaluative apprehension, must be actively addressed to ensure user acceptance and positive engagement. Designing AI as a supportive and complementary partner, rather than as a replacement for human abilities, may help mitigate negative emotions, reduce anticipatory intergroup anxiety, and foster greater willingness to engage.

This approach is consistent with the principles of Augmented Intelligence, which emphasize the enhancement of human capacities through AI collaboration rather than substitution. By recognizing and addressing the psychological reactions AI provokes—including concerns about autonomy, identity, and social status—developers and policymakers can maximize the

benefits of both human and artificial intelligence while minimizing resistance and emotional barriers. Successful AI integration, therefore, hinges as much on understanding and managing human psychological responses as on technological innovation itself, pointing toward a future where collaborative, human-centred AI applications can support rather than threaten human functioning.

The implications of these findings extend directly to AI communication strategies and design practices. Even minimal cues of AI involvement can trigger negative emotions and anticipatory anxiety, regardless of the objective quality of AI outputs. This highlights the need for thoughtful framing and attribution of AI in user-facing contexts. Presenting AI as an assistive or collaborative tool—rather than as a competitor or replacement—can help reduce symbolic threat and intergroup anxiety. Transparent communication about AI's limitations, its reliance on human oversight, and its specific role within broader socio-technical systems can reduce ambiguity and foster trust.

Furthermore, organizational policies and training programs that normalize positive human–AI collaboration, offer gradual exposure, and explicitly address concerns about loss of uniqueness or status may be key in reducing resistance. Interface designs that humanize AI, foreground shared goals, and support user agency can further buffer against threat perceptions, facilitating more adaptive and positive integration of AI technologies.

Acknowledgements

The authors have nothing to report. Generative AI (ChatGPT-5, OpenAI) was used solely to refine the English language of this manuscript, and all AI-edited text was subsequently reviewed and approved by the authors. Open access publishing facilitated by Universita degli Studi di Milano-Bicocca, as part of the Wiley - CRUI-CARE agreement.

Funding

The authors have nothing to report.

Consent

Participants provided informed written consent prior to study enrollment.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on the Open Science Framework web platform https://osf.io/wjsyr/overview?view_only=84332deee472412daac67893fc0adcd7.

References

Bartneck, C., T. Kanda, O. Mubin, and A. Al Mahmud. 2009. "Does the Design of a Robot Influence Its Animacy and Perceived Intelligence?" *International Journal of Social Robotics* 1, no. 2: 195–204. https://doi.org/10.1007/s12369-009-0013-7.

Branscombe, N. R., N. Ellemers, R. Spears, and B. Doosje. 1999. "The Context and Content of Social Identity Threat." In *Social Identity: Context, Commitment, Content*, edited by N. Ellemers, R. Spears, and B. Doosje, 35–58. Blackwell.

Brauner, P., F. Glawe, G. L. Liehner, L. Vervier, and M. Ziefle. 2025. "Mapping Public Perception of Artificial Intelligence: Expectations, Risk–Benefit Tradeoffs, and Value as Determinants for Societal Acceptance." *Technological Forecasting and Social Change* 220: 124304. https://doi.org/10.1016/j.techfore.2025.124304.

Burton, J. W., M. K. Stein, and T. B. Jensen. 2020. "A Systematic Review of Algorithm Aversion in Augmented Decision Making." *Journal of Behavioral Decision Making* 33, no. 2: 220–239. https://doi.org/10.1002/bdm.2155.

Cottrell, C. A., and S. L. Neuberg. 2005. "Different Emotional Reactions to Different Groups: A Sociofunctional Threat-Based Approach to 'Prejudice'." *Journal of Personality and Social Psychology* 88, no. 5: 770–789. https://doi.org/10.1037/0022-3514.88.5.770.

Cuddy, A. J., S. T. Fiske, and P. Glick. 2008. "Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map." *Advances in Experimental Social Psychology* 40: 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0.

Daneshjou, R., M. P. Smith, M. D. Sun, V. Rotemberg, and J. Zou. 2021. "Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review." *JAMA Dermatology* 157, no. 11: 1362–1369. https://doi.org/10.1001/jamadermatol.2021.3129.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* 13, no. 3: 319–340. https://doi.org/10.2307/249008.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." *Journal of Experimental Psychology. General* 144, no. 1: 114–126. https://doi.org/10.1037/xge0000033.

Epley, N., A. Waytz, and J. T. Cacioppo. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review* 114, no. 4: 864–886. https://doi.org/10.1037/0033-295X.114.4.864.

Fiske, S. T. 2018. *Social Beings: Core Motives in Social Psychology*. John Wiley & Sons.

Fiske, S. T., A. J. C. Cuddy, P. Glick, and J. Xu. 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, no. 6: 878–902. https://doi.org/10.1037/0022-3514.82.6.878.

Gabbiadini, A., F. Durante, C. Baldissarri, and L. Andrighetto. 2024. "Artificial Intelligence in the Eyes of Society: Assessing Social Risk and Social Value Perception in a Novel Classification." *Human Behavior and Emerging Technologies*: 7008056. https://doi.org/10.1155/2024/7008056.

Gabbiadini, A., D. Ognibene, C. Baldissarri, and A. Manfredi. 2024. "The Emotional Impact of Generative AI: Negative Emotions and Perception of Threat." *Behaviour & Information Technology* 44, no. 4: 676–693. https://doi.org/10.1080/0144929X.2024.2333933.

Gaube, S., H. Suresh, M. Raue, et al. 2021. "Do as AI Say: Susceptibility in Deployment of Clinical Decision-Aids." *npj Digital Medicine* 4, no. 1: 31. https://doi.org/10.1038/s41746-021-00385-9.

Harmon-Jones, C., B. Bastian, and E. Harmon-Jones. 2016. "The Discrete Emotions Questionnaire: A New Tool for Measuring State Self-Reported Emotions." *PLoS One* 11, no. 8: e0159915. https://doi.org/10.1371/journal.pone.0159915.

Hayes, A. F. 2018. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (Methodology in the Social Sciences)*. 2nd ed. Guilford Press.

Ipsos Mori. 2017. *Public Views of Machine Learning. Findings From Public Research Engagement Conducted on Behalf of the Royal Society*. Ipsos Mori, The Royal Society.

Jussupow, E., I. Benbasat, and A. Heinzl. 2020. "Why Are We Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion. In Proceedings of the 28th European Conference on Information Systems (ECIS)". https://aisel.aisnet.org/ecis2020_rp/168.

Jussupow, E., K. Spohrer, and A. Heinzl. 2022. "Identity Threats as a Reason for Resistance to Artificial Intelligence: Survey Study With Medical Students and Professionals." *JMIR Formative Research* 6, no. 3: e28750. https://doi.org/10.2196/28750.

Kieslich, K., M. Lünich, and F. Marcinkowski. 2021. "The Threats of Artificial Intelligence Scale (TAI) Development, Measurement, and Test Over Three Application Domains." *International Journal of Social Robotics* 13, no. 7: 1563–1577. https://doi.org/10.1007/s12369-020-00734-w.

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch. 2001. "Risk as feelings." *Psychological Bulletin* 127, no. 2: 267–286. https://doi.org/10.1037/0033-2909.127.2.267.

Logg, J. M., J. A. Minson, and D. A. Moore. 2019. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment." *Organizational Behavior and Human Decision Processes* 151: 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005.

Mackie, D. M., and E. R. Smith. 2018. "Intergroup Emotions Theory: Production, Regulation, and Modification of Group-Based Emotions." In *Advances in Experimental Social Psychology*, vol. 58, 1–69. Academic Press. https://doi.org/10.1016/bs.aesp.2018.03.001.

Maddux, W. W., A. D. Galinsky, A. J. C. Cuddy, and M. Polifroni. 2008. "When Being a Model Minority Is Good... and Bad: Realistic Threat Explains Negativity Toward Asian Americans." *Personality and Social Psychology Bulletin* 34, no. 1: 74–89. https://doi.org/10.1177/0146167207309195.

Mahmud, H., A. N. Islam, S. I. Ahmed, and K. Smolander. 2022. "What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion." *Technological Forecasting and Social Change* 175: 121390. https://doi.org/10.1016/j.techfore.2021.121390.

Markets and Markets. 2024. "AI in Media and Entertainment Market Size, Share, Growth Analysis and Forecast to 2033". https://artsmart.ai/blog/ai-in-media-and-entertainment-statistics/.

McKee, K. R., X. Bai, and S. T. Fiske. 2023. "Humans Perceive Warmth and Competence in Artificial Intelligence." *iScience* 26, no. 8: 107256. https://doi.org/10.1016/j.isci.2023.107256.

Mirbabaie, M., F. Brünker, N. R. J. Möllmann (Frick), and S. Stieglitz. 2022. "The Rise of Artificial Intelligence—Understanding the AI Identity Threat at the Workplace." *Electronic Markets* 32: 73–99. https://doi.org/10.1007/s12525-021-00496-x.

Morewedge, C. K. 2022. "Preference for Human, Not Algorithm Aversion." *Trends in Cognitive Sciences* 26, no. 10: 824–826. https://doi.org/10.1016/j.tics.2022.07.007.

Nass, C., and Y. Moon. 2000. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56, no. 1: 81–103. https://doi.org/10.1111/0022-4537.00153.

OpenAI. 2024. "*Sora* [Generative AI model]". https://openai.com/sora.

Ore, O., and M. Sposato. 2022. "Opportunities and Risks of Artificial Intelligence in Recruitment and Selection." *International Journal of Organizational Analysis* 30, no. 6: 1771–1782. https://doi.org/10.1108/IJOA-07-2020-2291.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing (Version 2024.12.1) [Computer Software]*. R Foundation for Statistical Computing. https://www.R-project.org/.

Riedl, R., S. A. Hogeterp, and M. Reuter. 2024. "Do Patients Prefer a Human Doctor, Artificial Intelligence, or a Blend, and Is This Preference Dependent on Medical Discipline? Empirical Evidence and Implications for Medical Practice." *Frontiers in Psychology* 15: 1422177. https://doi.org/10.3389/fpsyg.2024.1422177.

Rosseel, Y. 2012. "Lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48, no. 2: 1–36. https://doi.org/10.18637/jss.v048.i02.

Sapru, A. 2025. "Psychological Resistance to AI: How Regulatory Focus Fuels AI Anxiety and Negative Attitudes Toward AI. SSRN 5295832". https://doi.org/10.2139/ssrn.5295832.

Schepman, A., and P. Rodway. 2020. "Initial Validation of the General Attitudes Towards Artificial Intelligence Scale." *Computers in Human Behavior Reports* 1: 100014. https://doi.org/10.1016/j.chbr.2020.100014.

Siemens, G., F. Marmolejo-Ramos, F. Gabriel, et al. 2022. "Human and Artificial Cognition." *Computers and Education: Artificial Intelligence* 3: 100107. https://doi.org/10.1016/j.caeai.2022.100107.

Slovic, P. 2004. "What's Fear Got to Do With It - It's Affect we Need to Worry About." *Missouri Law Review* 69: 5.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2007. "The Affect Heuristic." *European Journal of Operational Research* 177, no. 3: 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006.

Slovic, P., E. Peters, M. L. Finucane, and D. G. MacGregor. 2005. "Affect, Risk, and Decision Making." *Health Psychology* 24, no. 4S: S35–S40. https://doi.org/10.1037/0278-6133.24.4.s35.

Smith, C. A., and P. C. Ellsworth. 1985. "Patterns of Cognitive Appraisal in Emotion." *Journal of Personality and Social Psychology* 48, no. 4: 813–838. https://doi.org/10.1037//0022-3514.48.4.813.

Statista. 2024. "Artificial Intelligence (AI) and Video Streaming in the United States". https://www.statista.com/topics/12669/artificial-intelligence-ai-and-video-streaming-in-the-united-states/.

Stephan, W. G., C. L. Renfro, and M. D. Davis. 2008. "The Role of Threat in Intergroup Relations." In *Improving Intergroup Relations: Building on the Legacy of Thomas F. Pettigrew*, edited by S. Oskamp, 55–72. Wiley. https://doi.org/10.1002/9781444303117.ch5.

Stephan, W. G., and C. W. Stephan. 1985. "Intergroup Anxiety." *Journal of Social Issues* 41, no. 3: 157–175. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x.

Stephan, W. G., O. Ybarra, and G. Bachman. 1999. "Prejudice Toward Immigrants." *Journal of Applied Social Psychology* 29, no. 11: 2221–2237. https://doi.org/10.1111/j.1559-1816.1999.tb00107.x.

Stephan, W. G., O. Ybarra, and K. Rios. 2016. "Intergroup Threat Theory." In *International Encyclopedia of the Social & Behavioral Sciences*, edited by J. D. Wright, 2nd ed., 638–644. Elsevier. https://doi.org/10.1002/9781118783665.ieicc0162.

The Verge. 2025. "OpenAI's Sora Has Already Hit More Than 1 Million Downloads in Fewer Than Five Days". https://www.theverge.com/news/797752/openai-sora-app-1-million-downloads.

Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis. 2003. "Unified Theory of Acceptance and Use of Technology (UTAUT) [Database Record]. APA PsycTests". https://doi.org/10.1037/t57185-000.

Waytz, A., C. K. Morewedge, N. Epley, G. Monteleone, J. H. Gao, and J. T. Cacioppo. 2010. "Making Sense by Making Sentient: Effectance Motivation Increases Anthropomorphism." *Journal of Personality and Social Psychology* 99, no. 3: 410–435. https://doi.org/10.1037/a0020240.

Xu, Y., G. Zhou, R. Cai, and D. Gursoy. 2024. "When Disclosing the Artificial Intelligence (AI) Technology Integration Into Service Delivery Backfires: Roles of Fear of AI, Identity Threat and Existential Threat." *International Journal of Hospitality Management* 122: 103829. https://doi.org/10.1016/j.ijhm.2024.103829.

Yogeeswaran, K., J. Złotowski, M. Livingstone, C. Bartneck, H. Sumioka, and H. Ishiguro. 2016. "The Interactive Effects of Robot Anthropomorphism and Robot Ability on Perceived Threat and Support for Robotics Research." *Journal of Human-Robot Interaction* 5, no. 2: 29–47. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran.

Zhou, Y., Y. Shi, W. Lu, and F. Wan. 2022. "Did Artificial Intelligence Invade Humans? The Study on the Mechanism of Patients' Willingness to Accept Artificial Intelligence Medical Care: From the Perspective of Intergroup Threat Theory." *Frontiers in Psychology* 13: 866124. https://doi.org/10.3389/fpsyg.2022.866124.

Złotowski, J., K. Yogeeswaran, and C. Bartneck. 2017. "Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction." *International Journal of Social Robotics* 9, no. 2: 211–214. https://doi.org/10.1007/s12369-016-0385-9.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information. **Video S1:** AI-generated, low-quality (with hallucinations). Video used in the experimental manipulation for the AI-generated/Low-quality condition (containing typical generative AI hallucinations). **Video S2:** AI-generated, high-quality (no hallucinations). Video used in the experimental manipulation for the AI-generated/High-quality condition (no AI-related hallucinations). **Video S3:** Human-made, low-quality (with errors). Video used in the experimental manipulation for the Human-made/Low-quality condition (with visual distortions). **Video S4:** Human-made, high-quality (no errors). Video used in the experimental manipulation for the Human-made/High-quality condition (no visual distortions).