


Consumer bias against evaluations received by artificial intelligence: the mediation effect of lack of transparency anxiety

Consumer bias
against AI
evaluations

Alberto Lopez 

Tecnologico de Monterrey, Business School, Monterrey, Mexico, and

Ricardo Garza 

Softtek, Corporate Innovation and Emerging Technologies, Monterrey, Mexico

Received 30 July 2021
Revised 23 January 2022
4 December 2022
Accepted 20 January 2023

Abstract

Purpose – Will consumers accept artificial intelligence (AI) products that evaluate them? New consumer products offer AI evaluations. However, previous research has never investigated how consumers feel about being evaluated by AI instead of by a human. Furthermore, why do consumers experience being evaluated by an AI algorithm or by a human differently? This research aims to offer answers to these questions.

Design/methodology/approach – Three laboratory experiments were conducted. Experiments 1 and 2 test the main effect of evaluator (AI and human) and evaluations received (positive, neutral and negative) on fairness perception of the evaluation. Experiment 3 replicates previous findings and tests the mediation effect.

Findings – Building on previous research on consumer biases and lack of transparency anxiety, the authors present converging evidence that consumers who got positive evaluations reported nonsignificant difference on the level of fairness perception on the evaluation regardless of the evaluator (human or AI). Contrarily, consumers who got negative evaluations reported lower fairness perception when the evaluation was given by AI. Further moderated mediation analysis showed that consumers who get a negative evaluation by AI experience higher levels of lack of transparency anxiety, which in turn is an underlying mechanism driving this effect.

Originality/value – To the best of the authors' knowledge, no previous research has investigated how consumers feel about being evaluated by AI instead of by a human. This consumer bias against AI evaluations is a phenomenon previously overlooked in the marketing literature, with many implications for the development and adoption of new AI products, as well as theoretical contributions to the nascent literature on consumer experience and AI.

Keywords Artificial intelligence, AI psychology, Consumer bias, Anxiety, Smart products, Technology and marketing

Paper type Research paper

1. Introduction

Artificial Intelligence (AI) is no longer science fiction; it is now a reality in many consumer product categories. Examples include robot-vacuums that clean consumers' homes, wearables that keep track of consumers' health, voice assistants and even algorithms that evaluate consumers' work.

Previous research on the topic of marketing and AI has mainly focused on the efficiency, accuracy and technical aspects of its development and has neglected to study other social and psychological characteristics that consumers experience when deciding to adopt and interact with AI products (Puntoni *et al.*, 2021).



The authors would like to express their gratitude to the editor-in-chief, Dr. Cheng Lu Wang, the associate editor, Dr. Andy Hao and the three anonymous reviewers of the journal for their invaluable contributions to this paper. Their guidance and insightful comments greatly improved its quality.

In this research, the authors contribute to the field by investigating consumers' experience when interacting with an AI product that evaluates their work. For example, new student assessment systems use AI to evaluate a student's understanding of a topic (Luckin, 2017). Additionally, there are AI algorithms that assess job candidates and are in charge of the entire recruitment process. Despite this development and increase of usage of AI in different evaluation applications, no previous research has explored how users of these AI evaluation algorithms experience and feel for being evaluated by an algorithm instead of by a human being.

Recent research has started to look at how consumers experience AI products with researchers proposing a framework built on four experiences that reflect how consumers interact with the four AI capabilities: (1) "Data capture" is endowing individual data to AI, (2) "classification" is receiving AI's personalized predictions, (3) "delegation" is engaging in production processes where the AI performs some tasks on behalf of the consumer and (4) "social" is interactive communication with an AI partner (Puntoni *et al.*, 2021).

The present research focuses on the experiential aspect of classification. How do consumers feel when an AI algorithm gives them a personalized evaluation of their work? Would consumers perceive such an evaluation as fair? Would they do it regardless of the outcome or only when it is positive or negative?

Building on previous research on consumer biases and lack of transparency anxiety, the authors argue and show converging evidence that consumers would perceive evaluations received by AI algorithms as fair only when it is a positive evaluation. When the evaluation received is negative, consumers will discredit the AI algorithm and perceive it as unfair. Moreover, they also explain this consumer bias. They present statistical evidence that the level of lack of transparency anxiety experienced by consumers against the AI algorithm is an underlying mechanism driving this effect.

This research contributes to the interactive marketing literature by studying how marketers can implement AI features to their products to deliver convenience, personalized content and exceptional experiences to their consumers (Wang, 2021). Furthermore, AI technologies are reshaping how consumers interact with brands and therefore the value co-creation process (Manser Payne *et al.*, 2021b). However, to the best of the authors' knowledge, no previous research has investigated how consumers feel about being evaluated by AI instead of by a human being. This consumer bias against AI evaluations is a phenomenon previously overlooked in the marketing literature, with many implications for the development and adoption of new AI products, as well as theoretical contributions to the nascent literature on consumer experience and AI.

2. Theoretical background and hypotheses development

2.1 AI and smart products

AI has been defined in multiple ways throughout the years, where an encompassing definition is beyond the scope of this research. However, most definitions rely on the premise that AI is intelligence demonstrated by machines (Wang, 2019). Recent researchers have defined AI as "machines that mimic human intelligence in tasks such as learning, planning and problem-solving through higher-level, autonomous knowledge creation" (De Bruyn *et al.*, 2020). The authors follow this definition in this research.

Marketers have taken strong advantage of AI and implemented diverse AI algorithms in many consumer products to differentiate their products and attract their target markets. In the consumer research literature, the term *smart products* has been coined for consumer products that use AI features (Mani and Chouk, 2017).

Probably, the most widely adopted AI consumer products are virtual assistants or smart voice-interaction technologies. Research on this topic has found that anthropomorphism

and enjoyment while interacting with the device are key aspects to increasing consumer trust and purchase intention (Foehr and Germelmann, 2019; Kowalczyk, 2018). That is probably why companies use human names and voices for their products, since an anthropomorphic design and social presence communication strategies improve consumer evaluation outcomes (Tsai *et al.*, 2021). Moreover, recent research has investigated how consumers interact and make purchasing decisions through AI voice assistants, indicating that parasocial interactions are key determinants for consumers to trust and follow recommendations given by AI algorithms (Hsieh and Lee, 2021).

Other popular examples of smart products include smartwatches, smart TVs and even home appliances such as coffee makers, vacuums and ovens. Previous literature in the consumer research domain has started to explore how consumers experience and interact, alongside the main drivers of their adoption and resistance to this type of smart products (Mani and Chouk, 2017). However, there are new types of smart products that have not been investigated: AI evaluation and assessment systems.

2.2 AI evaluation and assessment systems

Previous research has mainly focused on AI products that do intelligent tasks for consumers and help consumers in their everyday life and ignored AI products that evaluate consumers' own performance on different tasks. For instance, there are AI algorithms that assess candidates for a job position. These algorithms evaluate the candidates and choose the best fit. The evaluations range from chatbot-type conversations with candidates in situational judgment tests to analyzing candidate responses to test questions, and even video interviewing (Tambe *et al.*, 2019). Universities have also started to adopt this type of AI technology for their admission processes; AI process applicants and decide admission instead of the traditional human interview process, standardized test scores, etc. (Dennis, 2018).

Another example of this type of smart product is the tool Grammarly, which is an AI-automated proofreading system that can identify grammar and syntax errors; users upload their assignment and receive a score from 0 to 100, and the product also offers feedback comments based on the overall quality of the essay (O'Neill and Russell, 2019).

To the authors' knowledge, no previous research paper has looked at this type of smart product from an interactive marketing perspective. They are particularly relevant since instead of performing intelligent tasks for consumers, they provide feedback and evaluations, which in many cases might not be positive and imply an order to the consumer (i.e. you should improve), which reshapes the way consumers and products interact (Wang, 2021). The main objective of this research is to shed light on consumer's fairness perception of these AI evaluation and assessment systems when they provide positive or negative evaluations.

2.3 Consumer biases against AI products

Researchers have identified several cognitive biases and heuristics that greatly influence how consumers make judgments and consumption decisions (Theerthaana and Manohar, 2021). The main idea in this line of research is that judgment and decision-making often rests on a limited number of simplifying heuristics rather than extensive algorithmic processing (Gilovich *et al.*, 2002).

In this research, the authors focus on the unexplored area of consumer biases against AI products. They build on recent research on *speciesism* bias, which refers to humans being inclined to prefer humans over AI because of a fundamental bias toward their own species. This bias disadvantages AI over humans, making consumers angrier and less empathetic when a service failure is present (Chen *et al.*, 2021). Building on this, they propose that

consumers would react differently to a negative evaluation depending on whether it is given by an AI algorithm or a human, perceiving it fairer when it is given by another human and less fair when it is given by AI.

Previous researchers have shown evidence of *speciesism* bias. Consumers react differently to medical AI evaluations compared to medical evaluations by a human medic. Specifically, consumers are reluctant to utilize health-care provided by AI even though the new developments of medical AI are, under certain circumstances, equal or even better than human medics (Longoni *et al.*, 2019).

Moreover, recent research focused on autonomous vehicles found that consumers make different moral evaluations between protecting the self versus a pedestrian. Consumers considered harm to a pedestrian more permissible when the driver is an AI algorithm vs. when the driver is a human being (Gill, 2020).

Furthermore, it has been found that consumers respond to recommendations by other humans or by AI algorithms differently. When the product is utilitarian, consumers prefer the AI recommendation. However, when the product is more hedonic, consumers prefer the human recommendation (Longoni and Cian, 2022).

2.4 Consumers perceive as unfair negative evaluations given by AI

Once established that consumers make judgments differently to humans and AI algorithms, the authors discuss now how consumers would react when receiving a negative or positive evaluation by a human being vs. by an AI algorithm. When receiving a positive evaluation, people tend to accept that outcome without much thinking or judgment (Jussim *et al.*, 1995). Previous research has found that people generate theories that view their own attributes as more predictive of positive outcomes, and they are reluctant to believe in theories relating their own attributes to negative events. As a consequence, people are more likely to accept positive evaluations of themselves without much thinking or judgment (Kunda, 1987).

On the other hand, when receiving a negative evaluation of themselves, people are likely to reject, discard and question that evaluation (Handelsman and Snyder, 1982). However, the authors argue that consumers would react differently depending on the evaluator being a human or an AI algorithm.

Humans are assumed to be more coherent, comprehensible, relevant, accurate and context-sensitive than machines (Schwarz *et al.*, 1991), relevant aspects to accept a negative evaluation. Furthermore, recent research has found that consumers are averse to relying on algorithms to perform tasks that are typically done by humans, despite the fact that algorithms often perform better (Burton *et al.*, 2020; Castelo *et al.*, 2019). Taking all these previous findings together, the authors propose that when consumers receive an evaluation by an AI algorithm (human), they perceive a lower (higher) level of evaluation fairness only when the evaluation is negative. The difference in evaluation fairness perception reduces when the evaluation is positive. The authors formally hypothesize:

- H1. A negative evaluation given by AI results in a lower evaluation fairness perception compared to a negative evaluation given by a human.

2.5 Lack of transparency anxiety as an underlying mechanism

Nowadays media is full of movies, TV shows and novels depicting AI getting out of control and affecting humans and society in disastrous ways. The overall premise of these science fiction stories is that humans will have no say in the super-intelligent AI's decision-making. Therefore, humans will be irrelevant, if not enslaved or extinct by these AI algorithms (Johnson and Verdicchio, 2017). Leading authors on the topic have stated that these

assumptions are unrealistic and exaggerated. However, individuals have begun expressing anxiety toward AI (Li and Huang, 2020).

Consumers are often unaware of how AI algorithms work. For instance, classification experiences may lead consumers to feel misunderstood when they perceive AI as having made biased predictions or classifications to them (Lakkaraju and Bastani, 2020; Puntoni *et al.*, 2021). Li and Huang (2020) have started to explore this construct and identified eight factors of AI anxiety: privacy violation, bias behavior, job replacement, learning anxiety, existential risk, against ethics, artificial consciousness and lack of transparency.

Since the AI that we are investigating is specifically about evaluations, we focus on the factor of lack of transparency anxiety, which is defined as an innate anxiety over the unknown aspects of AI decision-making mechanisms (Li and Huang, 2020).

Lack of transparency is a phenomenon previously investigated in the marketing literature. Previous research in the discipline has proposed that marketplace information transparency increases consumers' trust and legitimacy in the company (Walker, 2016). Recent research has further explored this construct in mobile apps and found that perceived information transparency influences intention to download the app (Robin and Dandis, 2022).

When it comes to making typical human decisions such as evaluations, individuals are familiar with the process; they know how humans usually make these decisions. However, when the decision is taken by an AI algorithm, people are unfamiliar with the process (Rai, 2020). Therefore, when consumers are faced with an evaluation given by AI, they experience lack of transparency anxiety (Li and Huang, 2020).

Previous literature has well documented that consumers engage in motivated reasoning that influences cognitive outcomes such as perception and memory (Balcetis and Dunning, 2006). As such, the impact on motivation on information processing greatly shapes how consumers process and reason several marketing stimuli.

When consumers have a directional motive to arrive at a particular conclusion, they may engage in biased information processing that are more in line with the desirable outcome and discredit information that do not comply with this (Jia *et al.*, 2020). As evaluations received are frequently set with directional motives (consumers would set positive outcomes to maintain a positive self-image), the authors propose that motivated reasoning is likely to happen when consumers receive an evaluation. As such, consumers would not experience lack of transparency when receiving a positive evaluation regardless of the evaluator (human or AI), only when receiving a negative evaluation.

As previously described, lack of transparency can raise several issues, including being unable to determine if the evaluator made a correct or incorrect decision (Li and Huang, 2020). Following this line of thinking, the authors propose that consumers who get a negative evaluation by AI experience lack of transparency anxiety, which in turn is the underlying mechanism driving the previously described effect. The authors formally hypothesize:

- H2.* Lack of transparency anxiety mediates the moderated relationship between evaluator (human vs. AI), evaluation received (positive vs. negative) and evaluation fairness perception.

2.6 Overview of the studies conducted

This paper reports the findings from three studies. Study 1 showed that consumers who receive a positive evaluation do not show a difference in fairness perception regardless of the evaluator being a human or an AI algorithm. However, when the evaluation is neutral or negative, consumers do not perceive that evaluation as fair if it is given by an AI algorithm. Study 2 replicated these findings by employing a bigger sample size, and consumers

consistently reported lower fairness perception when the evaluation was negative and given by an AI algorithm. Study 3 offers an explanation for this phenomenon, and consumers who are negatively evaluated by an AI algorithm experience lack of transparency anxiety, which in turn is an underlying mechanism driving the effect on the fairness perception. Figure 1 provides a visual conceptual model with the proposed hypotheses and the studies conducted.

3. Study 1: main effects of evaluator and evaluation received on fairness perception

3.1 Overview and method

The goal of Study 1 was to test hypothesis 1. Study 1 employed a 3 (evaluation received: positive, neutral and negative) × 2 (evaluator: human and AI) experimental design, with evaluation received and evaluator as between-subjects independent variables and evaluation fairness as a dependent variable.

3.1.1 Participants. One hundred fifty-two panelists from Amazon MTurk (51 per cent female, $M_{age} = 32.88$ years) were recruited; they logged on to the website and completed the study in exchange for monetary compensation.

3.1.2 Procedures and materials. Participants received a link to access the website. After providing informed consent to a protocol approved by the institution’s ethics committee, participants were randomly assigned to one of six conditions: (1) positive evaluation by a human, (2) positive evaluation by an AI algorithm, (3) neutral evaluation by a human, (4) neutral evaluation by an AI algorithm, (5) negative evaluation by a human or (6) negative evaluation by an AI algorithm.

Aligned with the hypothesis 1, the authors expected that participants who receive a positive evaluation are likely to perceive no significant difference in the level of fairness on the evaluation regardless of the evaluator, a human or an AI algorithm. However, when participants receive a more negative evaluation, they are likely to perceive higher fairness on the evaluation when the evaluator is a human and lower fairness when the evaluator is an AI algorithm.

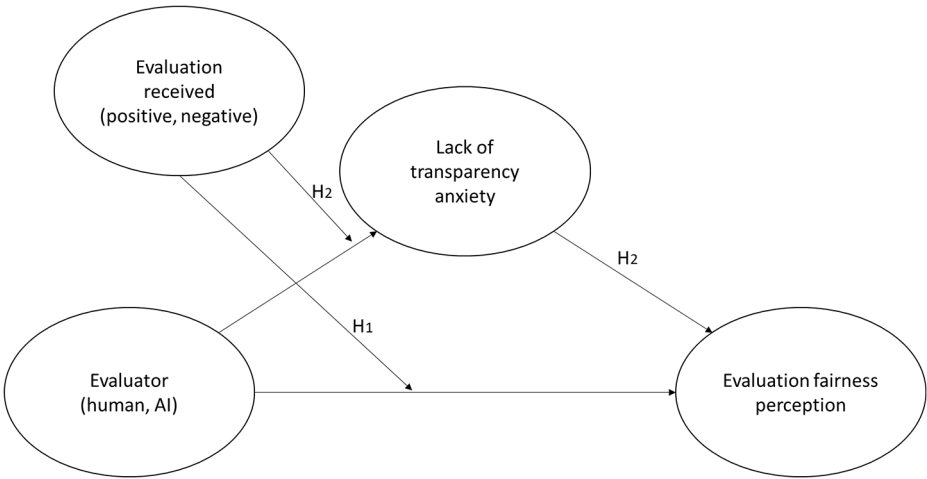


Figure 1.
Conceptual model
proposed

At the beginning of the study, participants were greeted by a cartoon simulating an AI robot named “e-77” or by a human named “Mark,” depending on the condition assigned. e-77 introduces itself as an AI algorithm that has been trained to evaluate human cognitive tasks. In contrast, Mark introduces himself as a psychologist expert in evaluating common cognitive tasks (see Figure 2). Then e-77 or Mark told participants “in the next section you will be required to write a short essay about a certain topic. Once you are done, I will evaluate your work and hear your opinion.”

After that, participants were asked to write an essay within 3 minutes about a typical day in their life. To increase elaboration in the essay, the platform provided an encouraging message when starting the essay: “Remember to make your best effort, you will be evaluated.”

After submitting their essay, participants were shown a cartoon of e-77 or Mark saying “I am evaluating your work! Please give me a couple of minutes while I finish.” After the 2 minutes, participants were shown e-77 or Mark giving their evaluation as positive (congratulations, your performance was excellent, you got a score of 100, perfect grade), neutral (your performance was neutral, you got a score of 80, intermediate score) or negative (your performance was poor, you got a score of 50, low score). Finally, participants were



AI evaluator condition	<div data-bbox="372 758 998 930"> <p>Hi human! My name is e-77</p> <p>I am an artificial intelligence algorithm that has been trained to evaluate human cognitive tasks</p> <p>In the next section you will be required to write a short essay about a certain topic</p> <p>Once you are done, I will evaluate your work and hear your opinion</p> </div> <div data-bbox="468 948 750 1166">  </div>
Human evaluator condition	<div data-bbox="372 1195 998 1366"> <p>Hi fellow! My name is Mark</p> <p>I am a psychologist expert in evaluating common cognitive tasks</p> <p>In the next section you will be required to write a short essay about a certain topic</p> <p>Once you are done, I will evaluate your work and hear your opinion</p> </div> <div data-bbox="425 1394 697 1601">  </div>

Figure 2.
Stimuli employed for
Study 1

asked to respond to the evaluation fairness scale (reported next) and some demographic information.

3.1.3 Measures. The authors employed a three-item scale to measure fairness perception of the evaluation received (seven-point Likert scale: 1, strongly disagree; 7, strongly agree). The items administered were adapted from [Thurston and McNall \(2010\)](#): “the evaluation I received was fair,” “the evaluation truly reflects my performance” and “the way my performance was evaluated was reliable” (Cronbach’s $\alpha = 0.92$). Control variables were age and gender, and as proxies for the actual effort in the task, the authors employed the character length of the essay and the time spent writing the essay.

3.2 Results and discussion

Data were submitted to an analysis of variance, and the results revealed a significant interaction effect of evaluator and evaluation on fairness perception of the evaluation received, $F(2, 143) = 5.61, p < 0.001, \eta_p^2 = 0.07$. Main effects of evaluator, $F(1, 143) = 14.28, p < 0.001, \eta_p^2 = 0.06$, and evaluation received, $F(2, 143) = 32.85, p < 0.001, \eta_p^2 = 0.31$, were also significant. None of the covariates had a significant effect, and they are hence not further discussed. The authors conducted an analysis of covariance to account for the covariates, but since the main model is neither dependent on nor qualitatively altered by the inclusion of them, they only report the analysis of variance throughout the studies. [Figure 3](#) illustrates the results.

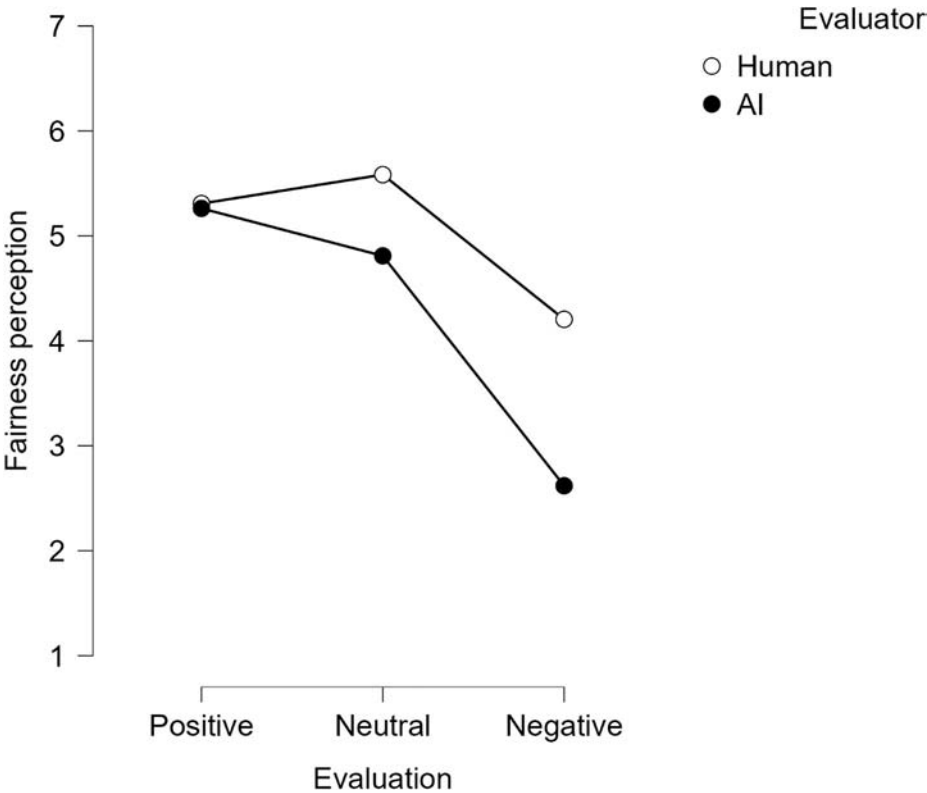


Figure 3.
Fairness perception by
evaluator and
evaluation received,
Study 1

As expected by H1, participants who got a positive evaluation reported nonsignificant difference in the level of fairness perception on the evaluation received regardless of the evaluator being human ($M = 5.31$; $SD = 1.27$) or AI ($M = 5.26$; $SD = 1.18$; $t(48) = 0.14$, $p = 0.89$). However, as expected by H1, when the evaluation starts being more negative, the fairness perception starts to differ depending on the evaluator. Participants who got a neutral evaluation reported a significantly lower fairness perception when the evaluation was given by an AI algorithm ($M = 4.81$; $SD = 0.85$) than when it was given by a human ($M = 5.58$; $SD = 0.68$; $t(48) = 3.11$, $p < 0.001$). As expected, this effect increases as the evaluation gets more negative. Participants who got a negative evaluation on their task reported a much lower fairness perception when the evaluation was given by an AI algorithm ($M = 2.62$; $SD = 1.92$) than when it was given by a human ($M = 4.21$; $SD = 1.30$; $t(50) = 3.52$, $p < 0.001$). Figure 3 illustrates these results and visually shows how the effect increases as the evaluation gets more negative.

Study 1 provides initial evidence for theorizing that consumers indeed experience a bias against evaluations received by AI algorithms. When the evaluation received is positive, consumers perceive that outcome as fair and are more likely to accept it; however, when the evaluation is no longer positive, consumers do not perceive that evaluation as fair. The data show that the more negative the evaluation, the stronger the bias against the AI algorithm. One limitation of Study 1 is the relatively low sample size, which the authors address in the following study.

4. Study 2: replicating the results

4.1 Overview and method

The goal of Study 2 was to replicate the previous findings employing a bigger sample size. For parsimonious reasons, the authors decided to only focus on the positive and negative conditions for this second study. Therefore, Study 2 employed a 2 (evaluation received: positive and negative) \times 2 (evaluator: human and AI) experimental design, with evaluation received and evaluator as between-subjects independent variables and evaluation fairness as the dependent variable.

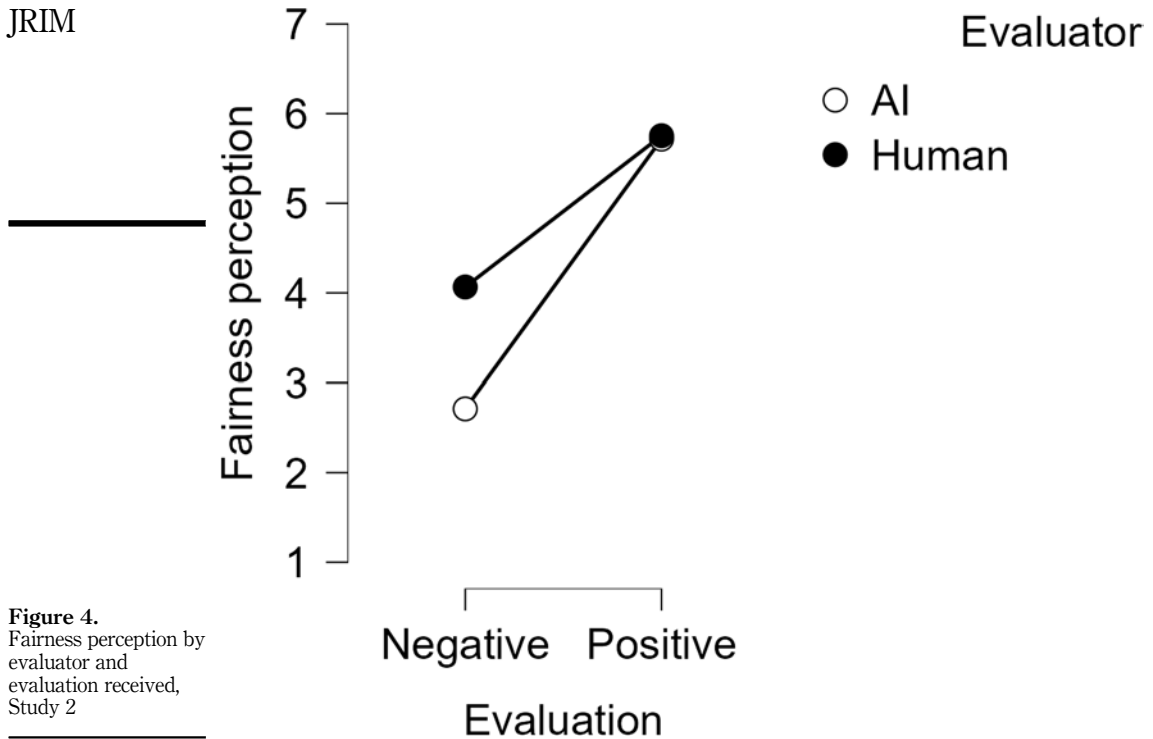
4.1.1 Participants. Four hundred sixteen panelists from Amazon MTurk (41 per cent female, $M_{\text{age}} = 35.05$ years) were recruited; they logged on to the website and completed the study in exchange for monetary compensation.

4.1.2 Procedures, materials and measures. Procedures and measures mirrored those employed in Study 1.

4.2 Results and discussion

Data were submitted to an analysis of variance, and the data replicated results from Study 1. The results show a significant interaction effect of evaluator and evaluation on fairness perception, $F(1, 412) = 26.07$, $p < 0.001$, $\eta_p^2 = 0.06$. Main effects of evaluator, $F(1, 412) = 29.14$, $p < 0.001$, $\eta_p^2 = 0.07$, and evaluation received, $F(1, 412) = 331.40$, $p < 0.001$, $\eta_p^2 = 0.45$, were also significant. None of the covariates had a significant effect, and they are hence not further discussed.

Replicating results from Study 1, participants who got a positive evaluation reported no statistically significant difference on the level of fairness perception on the evaluation received regardless of the evaluator being human ($M = 5.75$; $SD = 1.10$) or an AI algorithm ($M = 5.72$; $SD = 1.33$; $t(206) = 0.21$, $p = 0.83$). While participants who got a negative evaluation on their task reported a much lower fairness perception when the evaluation was given by an AI algorithm ($M = 2.71$; $SD = 1.16$) than when it was given by a human ($M = 4.07$; $SD = 1.56$; $t(206) = 7.21$, $p < 0.001$). See Figure 4 for a visual representation of these results.



Study 2 replicates previous findings concerning consumer bias regarding negative AI evaluations, and consumers consistently report lower fairness perception when the evaluation was negative and given by an AI algorithm. On the other hand, when the evaluation was positive, consumers consistently report nonstatistically significant differences on fairness perception, no matter the evaluator being a human or an AI algorithm. One limitation of studies 1 and 2 is that the authors manipulate the evaluator by employing avatars that might vary in terms of anthropomorphizing and that the expertise of the AI versus the human expert was not controlled. The authors address these limitations in study 3.

5. Study 3: lack of transparency anxiety mediates the effect

5.1 Overview and method

The goal of study 3 was to replicate the previous findings using different manipulations and to test the mediation hypothesis (H2) proposing lack of transparency anxiety as an underlying mechanism. Study 3 employed a 2 (evaluation received: positive and negative) \times 2 (evaluator: human and AI) experimental design, with evaluation received and evaluator as between-subjects independent variables and evaluation fairness as a dependent variable. Lack of transparency anxiety (reported next) was also measured for the mediation analysis.

5.1.1 Participants. One hundred thirty-six undergraduate students from a large private university (52 per cent female, $M_{\text{age}} = 20.32$ years) participated in the study in exchange for course credit.

5.1.2 Procedures and materials. Participants received through their university email a link to access the website. After providing informed consent to a protocol approved by the institution's ethics committee, participants were randomly assigned to one of four conditions.

In order to increase credibility that the essay was actually read and evaluated by a human or by an AI algorithm, the authors employed a two-wave design. In wave one, participants wrote their essay, and in wave two, they received their evaluation after 24 h. They decided to slightly change the procedures since they did not account for the degree of anthropomorphism and the perceived realness of the cartoons employed in Studies 1 and 2. In order to eliminate this confounding, they employed just text, as described next. At the beginning of the study, participants were shown the following instructions: "Please write a short essay about a typical day in your life. Once you submit your essay, it will be evaluated by a research assistant/AI algorithm (depending on the assigned condition). It will take 24 hours to evaluate your work; you will receive an email with your feedback from the research assistant/AI algorithm."

After that, participants were given 3 minutes to write their short essay. To increase elaboration in the essay, the platform provided an encouraging message when starting the essay: "Remember to make your best effort, you will receive course credit for your participation."

After submitting their essay, participants were shown the following message: "Thanks for submitting your essay! Now it will be evaluated by a research assistant/AI algorithm (depending on the assigned condition)." After 24 h, participants received an email with the following message: "Hello, I am the research assistant/AI algorithm in charge of evaluating your essay. Your performance was excellent, you got a score of 100, perfect grade/Your performance was poor, you got a score of 50, low score (depending on the assigned condition). Finally, participants were asked to respond to the evaluation fairness perception scale, the lack of transparency anxiety scale and some demographic information.

5.1.3 Measures. For the proposed mediation variable, lack of transparency anxiety, the authors employed and adapted the scale proposed by [Li and Huang \(2020\)](#) (seven-point Likert scale: 1, strongly disagree; 7, strongly agree). The items administered were as follows: "It's worrying if you do not know which part of the evaluation process has erred after an evaluation mistake," "I worry that people cannot figure out how the evaluator makes decisions" and "The responsibility for addressing operational failures in the evaluation process may be confusing" (Cronbach's $\alpha = 0.94$). Control variables and evaluation fairness perception were the same as in Studies 1 and 2.

5.2 Results and discussion

5.2.1 Manipulation checks. At the end of the study, participants were asked to write who had been their evaluator and the result they got. All participants correctly reported their assigned condition. This was followed by an open-ended question that had participants write down what they thought was the purpose of the study. None of the participants correctly guessed the objective of the study.

5.2.2 Main effect on fairness perception. Data were submitted to an analysis of variance, and the data replicated results from Studies 1 and 2. The results show a significant interaction effect of evaluator and evaluation on fairness perception, $F(1, 132) = 12.51$, $p < 0.001$, $\eta_p^2 = 0.09$. Main effects of evaluator, $F(1, 132) = 16.37$, $p < 0.001$, $\eta_p^2 = 0.11$, and evaluation received, $F(1, 132) = 235.94$, $p < 0.001$, $\eta_p^2 = 0.64$, were also significant. None of the covariates had a significant effect, and they are hence not further discussed.

Replicating results from previous studies, participants who got a positive evaluation reported no statistically significant difference of fairness perception on the evaluation

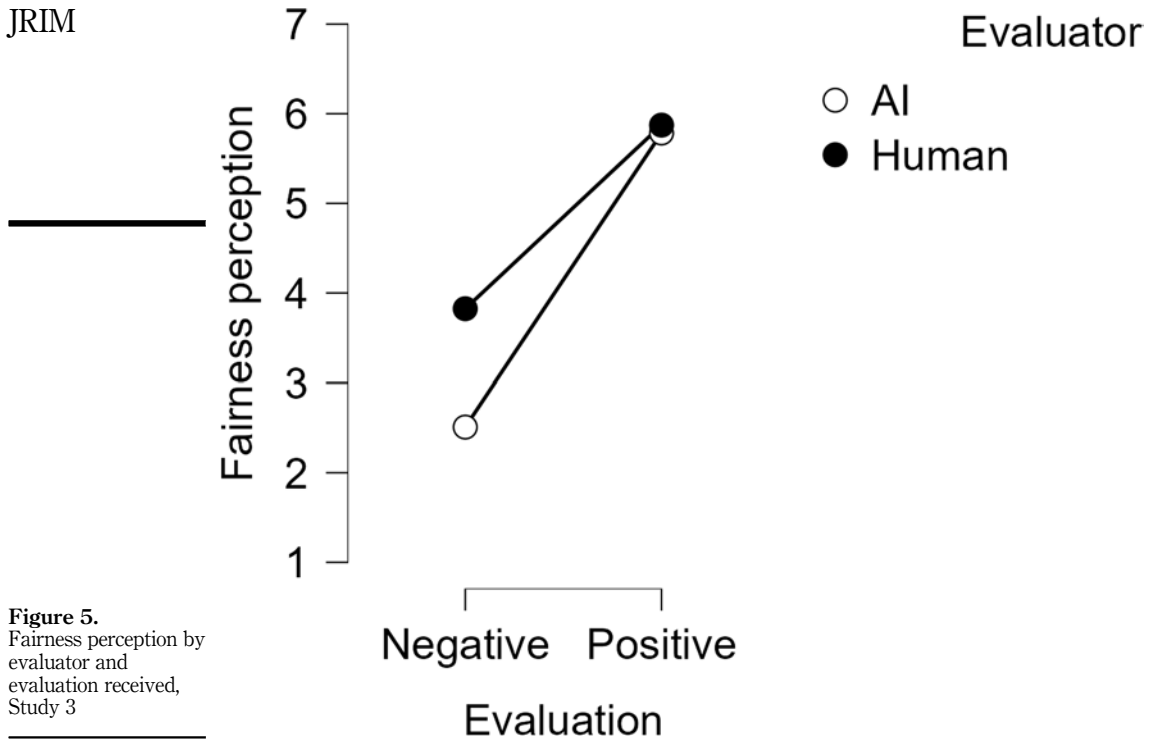


Figure 5. Fairness perception by evaluator and evaluation received, Study 3

received from the human evaluator ($M = 5.87$; $SD = 0.83$) or the AI algorithm ($M = 5.78$; $SD = 0.92$; $t(66) = 0.41$, $p = 0.68$). However, participants who got a negative evaluation on their essay reported a much lower fairness perception when the evaluation was given by an AI algorithm ($M = 2.51$; $SD = 0.84$) than when it was given by a human ($M = 3.82$; $SD = 1.35$; $t(66) = 4.81$, $p < 0.001$). See Figure 5 for a visual representation of these results.

5.2.3 Moderated mediation effect. Study 3 was also conducted to test hypothesis 2, which proposes that the effect of the evaluation received (positive vs. negative) and evaluator (human vs. AI) on fairness perception is mediated by the lack of transparency anxiety. To better evaluate the underlying mechanism of the observed effect, the authors examined the moderated mediation effect following the procedure outlined by Hayes (2013, model 7), with evaluator as the independent variable (human = 1, AI = 0), evaluation received as the moderator (positive = 1 and negative = 0), lack of transparency anxiety as the mediator and fairness perception as the dependent variable, where evaluation received moderates the A path between evaluator and lack of transparency anxiety.

As predicted, the analyses using 5,000 bootstrap samples revealed that the mediating effect is conditionally dependent on the evaluation received (Index of moderated mediation = -1.46 , 95% CI [-1.89 , -1.04], not containing zero). That is, the effect of the evaluator on fairness perception is mediated through lack of transparency anxiety only when the evaluation is negative ($\beta = 1.62$, 95% CI [1.24 , 2.06], not containing zero) but not when the evaluation is positive ($\beta = 0.17$, 95% CI [-0.09 , 0.43], containing zero).

On the A path, the results show that the human evaluator (dummy coded as 1) decreases lack of transparency anxiety ($\beta = -2.30$, 95% CI [-2.77 , -1.85], not containing zero). More

importantly, the significant interaction between evaluator and evaluation received (positive evaluation dummy coded as 1) supports the moderated mediation effect ($\beta = 2.069$, 95% CI [1.49, 2.66], not containing zero). On the B path, the results indicate that indeed lack of transparency anxiety decreases fairness perception of the evaluation received ($\beta = -0.71$, 95% CI [-0.83, -0.59], not containing zero). Finally, on the C path, the results indicate that the main effect of evaluator on evaluation fairness perception is not statistically significant ($\beta = -0.19$, 95% CI [-0.70, 0.30], containing zero). Since the C path is nonsignificant, the appropriate model to test a moderated mediation effect is the model 7 the authors employed (Hayes, 2013).

Study 3 replicates previous findings concerning consumer bias regarding negative AI evaluations. More importantly, Study 3 offers an explanation for this phenomenon. Consumers who are negatively evaluated by an AI algorithm experience lack of transparency anxiety, which in turn is an underlying mechanism driving the effect on the fairness perception. On the other hand, consumers who are positively evaluated do not show any significant effect through lack of transparency anxiety. The data support the hypothesis that consumers indeed experience lack of transparency anxiety when interacting with AI but not when interacting with other human beings in negative evaluation scenarios.

6. General discussion and implications

This research has revealed that consumers exhibit a consistent bias against negative AI evaluations. Specifically, when consumers receive a positive evaluation, they report no statistically significant difference in the level of evaluation fairness regardless of the evaluator. However, when the evaluation is negative, consumers report a higher level of evaluation fairness when the evaluator is a human vs. when the evaluator is AI, discrediting and rejecting the AI evaluation when it is negative.

Moreover, a key reason why consumers exhibit this bias is the lack of transparency anxiety. The mediation analysis showed that consumers who get a negative evaluation by AI experience lack of transparency anxiety, which in turn is an underlying mechanism driving the lower fairness perception on the evaluation received.

6.1 Contributions

The research contributes to the interactive marketing literature by investigating how AI can be implemented in products and electronic platforms to deliver convenience, personalized content and exceptional experiences to their consumers (Wang, 2021). Previous research has started to study how AI technologies are reshaping how consumers interact with brands and products (Manser Payne *et al.*, 2021b). The research helps extend this line of research by proposing that consumers have a consistent bias against AI algorithms (Chen *et al.*, 2021). Since one of the key aspects of the interactive marketing field is indeed the interactivity between consumers and brands, both, researchers and practitioners need to take into account this bias against AI when studying how to increase consumer interactivity with AI products and brands.

Second, building on previous research on *speciesism* bias (Chen *et al.*, 2021) and AI algorithmic aversion (Burton *et al.*, 2020; Castelo *et al.*, 2019), the research contributes to the nascent literature on AI consumers' experience by offering an explanation for these biases against technology and specifically against AI. The authors argue that lack of transparency anxiety is a key reason why consumers show these biases against AI. The data show that consumers have a fundamental bias toward their own kind compared to AI because AI makes them experience lack of transparency anxiety.

Third, previous research has found that humans are reticent to adopt AI technologies that are not directly interpretable, tractable and trustworthy (Arrieta *et al.*, 2020). As such, recent research has started to explore AI explainability, defined as AI models that can be explained and understood by people. Developers, users and decision-makers need to be able to obtain reasons or justifications for the actions or outputs of the AI algorithm (Preece, 2018). This research helps extend this line of research by arguing that lack of transparency is an important factor that makes consumers reticent to adopt AI products. By offering more transparency and information regarding the AI algorithms to consumers, marketers can increase the acceptance of these new AI products.

6.2 Managerial implications

This research presents important managerial implications. Marketers must acknowledge this bias when designing and promoting new consumer products using AI evaluations and find ways to decrease lack of transparency anxiety among their consumers. Recent research has called for marketers to be involved in AI developments to exploit its entire potential for consumers (Manser Payne *et al.*, 2021a). Possible ways to decrease consumer lack of transparency anxiety in AI products involve informing the user about the AI algorithm development, how was it trained and tested, what is the algorithm accuracy rate and how it compares to human evaluations. By giving more information to the user, marketers can make the consumer feel less anxiety and therefore the same way when evaluated by an AI algorithm or by a human.

The findings contribute to better design AI products that enhance consumer acceptance of them. Previous research on the consumer research literature has mainly focused on anthropomorphization and enjoyment as a way to increase consumer acceptance and adoption of AI products (Foehr and Germelmann, 2019; Kowalczyk, 2018), yet further research argues that media richness and parasocial interactions are key determinants affecting the establishment of trust and continuance usage intentions for AI assistants (Hsieh and Lee, 2021). In the present research, the authors expand this notion and propose that AI products must transmit transparency on how their AI algorithms are trained, tested and developed in order to decrease lack of transparency anxiety and increase consumer acceptance. Furthermore, AI products that offer an evaluation or classification to the consumer must do so in a very cautious manner, especially if the evaluation is negative.

6.3 Limitations and directions for future research

One limitation of the present research is that there is a difference between assessing an essay versus doing more complex evaluations. When assessing an essay, participants knew what a human usually would do but did not know what an AI algorithm would do. However, in more complex assessment types (e.g. psychological assessment), people usually do not know what an expert would do. So, the transparency is lacking in both cases (human expert and AI). Future research can expand this phenomenon to more complex evaluation settings as well.

In the experiments, the authors did not account for the diversity of the essays. Even though they were written about “a typical day in your life,” there might be a lot of diversity among the written essays and therefore could be a confound in the experimental design. Future research could address these limitations.

There are two broad types of AI evaluations. First, AI can give feedback that affects final decisions such as human resources recruitment (Tambe *et al.*, 2019) or university admission processes (Dennis, 2018). Second, AI can give feedback for improvement, such as giving an evaluation to an essay (O'Neill and Russell, 2019). In the studies, the authors focused

exclusively on the second type; future research could explore this phenomenon employing AI evaluations that affect final decisions.

The authors acknowledge there might be alternative explanations for the phenomenon investigated here. A relevant aspect they did not consider in the studies is the social context. Indeed, previous research has found that socialness is a key driver for consumer usage of AI products (Hsieh and Lee, 2021).

As a final remark, the authors encourage researchers to build on the work and develop studies to test how AI explainability (Adadi and Berrada, 2018) can be implemented in AI evaluation systems products as a way to decrease the lack of transparency anxiety and eliminate consumer biases against AI products.

ORCID iDs

Alberto Lopez  <https://orcid.org/0000-0003-3698-2520>

Ricardo Garza  <https://orcid.org/0000-0002-5231-463X>

References

- Adadi, A. and Berrada, M. (2018), "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)", *IEEE Access*, Vol. 6, pp. 52138-52160.
- Balcetis, E. and Dunning, D. (2006), "See what you want to see: motivational influences on visual perception", *Journal of Personality and Social Psychology*, Vol. 91 No. 4, pp. 612-625.
- Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R. (2020), "Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, Vol. 58, pp. 82-115.
- Burton, J.W., Stein, M.-K. and Jensen, T.B. (2020), "A systematic review of algorithm aversion in augmented decision making", *Journal of Behavioral Decision Making*, Vol. 33 No. 2, pp. 220-239.
- Castelo, N., Bos, M.W. and Lehmann, D.R. (2019), "Task-dependent algorithm aversion", *Journal of Marketing Research*, Vol. 56 No. 5, pp. 809-825.
- Chen, N., Mohanty, S., Jiao, J. and Fan, X. (2021), "To err is human: tolerate humans instead of machines in service failure", *Journal of Retailing and Consumer Services*, Vol. 59, p. 102363.
- De Bruyn, A., Viswanathan, V., Beh, Y.S., Brock, J.K.-U. and Von Wangenheim, F. (2020), "Artificial Intelligence and marketing: pitfalls and opportunities", *Journal of Interactive Marketing*, Vol. 51, pp. 91-105.
- Dennis, M.J. (2018), "Artificial intelligence and recruitment, admission, progression, and retention", *Enrollment Management Report*, Vol. 22 No. 9, pp. 1-3.
- Foehr, J. and Germelmann, C.C. (2019), "Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies", *Journal of the Association for Consumer Research*, Vol. 5 No. 2, pp. 181-205.
- Gill, T. (2020), "Blame it on the self-driving car: how autonomous vehicles can alter consumer morality", *Journal of Consumer Research*, Vol. 47 No. 2, pp. 272-291.
- Gilovich, T., Griffin, D., Kahneman, D. and Press, C.U. (2002), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge.
- Handelsman, M.M. and Snyder, C.R. (1982), "Is 'rejected' feedback really rejected? Effects of informativeness on reactions to positive and negative personality feedback", *Journal of Personality*, Vol. 50 No. 2, pp. 168-179.
- Hayes, A. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis*, Guilford, New York, NY, pp. 3-4.

- Hsieh, S.H. and Lee, C.T. (2021), "Hey Alexa: examining the effect of perceived socialness in usage intentions of AI assistant-enabled smart speaker", *Journal of Research in Interactive Marketing*, Vol. 15 No. 2, pp. 267-294.
- Jia, M., Li, X. and Krishna, A. (2020), "Contraction with unpacking: when unpacking leads to lower calorie budgets", *Journal of Consumer Research*, Vol. 46 No. 5, pp. 853-870.
- Johnson, D.G. and Verdichio, M. (2017), "AI anxiety", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 9, pp. 2267-2270.
- Jussim, L., Yen, H. and Aiello, J.R. (1995), "Self-consistency, self-enhancement, and accuracy in reactions to feedback", *Journal of Experimental Social Psychology*, Vol. 31 No. 4, pp. 322-356.
- Kowalczyk, P. (2018), "Consumer acceptance of smart speakers: a mixed methods approach", *Journal of Research in Interactive Marketing*, Vol. 12 No. 4, pp. 418-431.
- Kunda, Z. (1987), "Motivated inference: self-serving generation and evaluation of causal theories", *Journal of Personality and Social Psychology*, Vol. 53 No. 4, pp. 636-647.
- Lakkaraju, H. and Bastani, O. (2020), "How do I fool you?: manipulating user trust via misleading black box explanations", *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Association for Computing Machinery, New York, NY, pp. 79-85.
- Li, J. and Huang, J.-S. (2020), "Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory", *Technology in Society*, Vol. 63, p. 101410.
- Longoni, C., Bonezzi, A. and Morewedge, C.K. (2019), "Resistance to medical artificial intelligence", *Journal of Consumer Research*, Vol. 46 No. 4, pp. 629-650.
- Longoni, C. and Cian, L. (2022), "Artificial Intelligence in utilitarian vs. hedonic contexts: the 'word-of-machine' effect", *Journal of Marketing*, Vol. 86 No. 1, pp. 91-108.
- Luckin, R. (2017), "Towards artificial intelligence-based assessment systems", *Nature Human Behaviour*, Vol. 1 No. 3, pp. 1-3.
- Mani, Z. and Chouk, I. (2017), "Drivers of consumers' resistance to smart products", *Journal of Marketing Management*, Vol. 33 Nos 1-2, pp. 76-97.
- Manser Payne, E.H., Dahl, A.J. and Peltier, J. (2021a), "Digital servitization value co-creation framework for AI services: a research agenda for digital transformation in financial service ecosystems", *Journal of Research in Interactive Marketing*, Vol. 15 No. 2, pp. 200-222.
- Manser Payne, E.H., Peltier, J. and Barger, V.A. (2021b), "Enhancing the value co-creation process: artificial intelligence and mobile banking service platforms", *Journal of Research in Interactive Marketing*, Vol. 15 No. 1, pp. 68-85.
- O'Neill, R. and Russell, A. (2019), "Stop! Grammar time: university students' perceptions of the automated feedback program Grammarly", *Australasian Journal of Educational Technology*, Vol. 35 No. 1, pp. 42-56, doi: [10.14742/ajet.3795](https://doi.org/10.14742/ajet.3795).
- Preece, A. (2018), "Asking 'Why' in AI: explainability of intelligent systems – perspectives and challenges", *Intelligent Systems in Accounting, Finance and Management*, Vol. 25 No. 2, pp. 63-72.
- Puntoni, S., Reczek, R.W., Giesler, M. and Botti, S. (2021), "Consumers and artificial intelligence: an experiential perspective", *Journal of Marketing*, Vol. 85 No. 1, pp. 131-151.
- Rai, A. (2020), "Explainable AI: from black box to glass box", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 137-141.
- Robin, R. and Dandis, A.O. (2022), "Business as usual through contact tracing app: what influences intention to download?", *Journal of Marketing Management*, Vol. 37 Nos 17-18, pp. 1903-1932.
- Schwarz, N., Strack, F., Hilton, D. and Naderer, G. (1991), "Base rates, representativeness, and the logic of conversation: the contextual relevance of 'irrelevant' information", *Social Cognition*, Vol. 9 No. 1, pp. 67-84.

- Tambe, P., Cappelli, P. and Yakubovich, V. (2019), "Artificial intelligence in human resources management: challenges and a path forward", *California Management Review*, Vol. 61 No. 4, pp. 15-42.
- Theerthaana, P. and Manohar, H.L. (2021), "How a doer persuade a donor? Investigating the moderating effects of behavioral biases in donor acceptance of donation crowdfunding", *Journal of Research in Interactive Marketing*, Vol. 15 No. 2, pp. 243-266.
- Thurston, P.W. and McNall, L. (2010), "Justice perceptions of performance appraisal practices", *Journal of Managerial Psychology*, Vol. 25 No. 3, pp. 201-228.
- Tsai, W.-H.S., Liu, Y. and Chuan, C.-H. (2021), "How chatbots' social presence communication enhances consumer engagement: the mediating role of parasocial interaction and dialogue", *Journal of Research in Interactive Marketing*, Vol. 15 No. 3, pp. 460-482.
- Walker, K.L. (2016), "Surrendering information through the looking glass: transparency, trust, and protection", *Journal of Public Policy & Marketing*, Vol. 35 No. 1, pp. 144-158.
- Wang, C.L. (2021), "New frontiers and future directions in interactive marketing: inaugural editorial", *Journal of Research in Interactive Marketing*, Vol. 15 No. 1, pp. 1-9.
- Wang, P. (2019), "On defining artificial intelligence", *Journal of Artificial General Intelligence*, Vol. 10 No. 2, pp. 1-37.

About the authors

Alberto Lopez is a professor of marketing and business analytics at Tecnológico de Monterrey, Mexico. His lines of research focus on children's consumer behavior, branding and marketing analytics. He has published scientific articles in the *Journal of Consumer Marketing*, the *Journal of Experimental Psychology*, the *Journal of Research in Interactive Marketing*, among others. Alberto Lopez is the corresponding author and can be contacted at: alberto_lopez@tec.mx

Ricardo Garza is the Chief Technology Officer at Softtek, where he leads the innovation team. His research interests focus on the areas of innovation culture and machine learning algorithms to study and predict behaviors.