International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Sexist AI: An Experiment Integrating CASA and ELM

Joo-Wha Hong , Sukyoung Choi & Dmitri Williams

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2020.1801226

Sexist AI: An Experiment Integrating CASA and ELM

Joo-Wha Hong, Sukyoung Choi, and Dmitri Williams

Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA, USA

ABSTRACT

This study employed an experiment to test participants' perceptions of an artificial intelligence (AI) recruiter. It used a 2 (Specialist AI/Generalist AI) × 2 (Sexist/nonsexist) design to test the relationship between these labels and the perception of moral violations. The theoretical framework was an integration of the Computers Are Social Actors (CASA) and Elaboration Likelihood Model (ELM) approaches. Participants (n = 233) responded to an online questionnaire after reading one of four scenarios involving an AI recruiter's evaluation of job candidates. Results found that the concept of "mindlessness" in CASA is situational, based on whether the issue is processed with the central route or the peripheral route. Moreover, this study shows that CASA can explain the evaluation of machines with the third-person point of view. Also, there was a distinction between the perception of the AI and its decisions. Furthermore, participants were found to be more sensitive about the AI agent's sexism – which was more anthropomorphic and emotionally engaging – than about the AI agent's status as a specialist.

Artificial intelligence (AI) is a technology that functions via algorithms that people use every day, such as personal assistants, spam e-mail filters, and self-driving cars (Mills, 2018). AI has begun to make decisions on behalf of humans since its purpose is to think and act both humanly and rationally (Russell & Norvig, 2010). The application of AI technology is generally forecasted to reduce costs while increasing productivity in many industries (Doganis et al., 2006; Ince & Aktan, 2009; Kreps & Neuhauser, 2013), and to have other profound impacts on society. For example, it is expected that the introduction of AI will begin to substitute for or make irrelevant, vast amounts of human labor (Bruun & Duka, 2018; Chelliah, 2017; Nilsson, 1985). This substitution is not only limited to routine labor but also to critical decision making. For instance, self-driving cars drive without human input using a decision-making module that understands circumstances based on information gathered from sensors; this is similar to how people drive (T. Kim et al., 2012). AI has already started to substitute for humans in some roles, and its performance compared to humans should be studied to predict the possible outcomes of having machines be decision-makers. Also critical is understanding our reactions to, and comfort with, those decisions. If AI becomes a critical part of everyday life, we will be dealing with new cultural and social implications.

One area already employing AI decision-making is job recruiting because AI hiring programs have already started to replace human recruiters (Wazed, 2019). Having AI in hiring practices is expected to bring about a wide range of changes to job markets by having automated and improved decision-making processes (Black & van Esch, 2020; Michailidis, 2018). Having an AI-based chatbot in the recruiting process was found to reduce 30–50% of the time spent screening candidates (Volini et al., 2019). Moreover, studies found that job applicants did not show a distinct difference in their perceptions of fairness toward AI hiring agents and AI recruitment (Suen et al., 2019; van Esch et al., 2019). Global conglomerates, such as IKEA, L'Oreal, and Amazon have started to test or deploy AI programs in their recruiting processes (Dustin, 2018; Holley, 2018; Lewis & Marc, 2019). However, despite the expected enormous changes in job markets due to AI hiring, we have little understanding of social acceptance of this trend. Understanding it is part of a larger issue of how we accept or reject AI technology in general. This study examines that issue in the context of a fallible AI hiring process, and uses a combination of two traditional communication theories, the Computers Are Social Actors (CASA) paradigm and the Elaboration Likelihood (ELM) Model. How we think about machines and their performance are critical pieces of our evolving modern lives.

1.1. AI making unethical decisions

There is a belief that machines are more just and trustworthy than people, hence the term "machine heuristics" (Sundar & Kim, 2019). However, questions about fairness and ethical concerns in using AI technologies have been addressed as challenges (Tambe et al., 2019). Moreover, there have already been cases where machines disappointed people by making incorrect or biased decisions. There was a case where

CONTACT Joo-Wha Hong joowhaho@usc.edu Annenberg School for Communication and Journalism, University of Southern California, 3502 Watt Way, Los Angeles, CA, 90089, USA

Amazon's AI hiring tool discriminated against and downgraded female candidates because the program was trained using the company's ten years of résumé data, which was highly male-dominant (Dustin, 2018). This practice was a violation of gender equality, an essential human right, and lead to both harms and inequities. The AI tool's decisions can create unjust outcomes that can negatively impact people's psychological well-being, such as depression and lower self-esteem, due to their innate characteristic (Hurst & Beesley, 2013; J. K. Swim et al., 2001). Also, the AI tool's decisions violated fair treatment, which is a crucial factor that job applicants consider when engaging with AI recruiting systems (van Esch & Black, 2019). Another representative case of biased algorithms is Tay, the Microsoft Twitter chatbot, which was an AI agent that later became racist because it learned from rancorous tweets that Twitter users intentionally put out (Beran, 2018). Reflecting on these cases, it becomes clear that AI can be biased based on how it is trained. Thus, contrary to the expectation that AIs make neutral and fair decisions, biased algorithms may emerge based on the data used for training, which Weizenbaum (1976) predicted and was concerned with over 40 years ago. Prior research has borne out some of his concerns, first focusing on the actions of the programmers, and then moving to the effects and perceptions of the end-users and customers.

There is an argument that says biased data for training AIs can be manipulated. There have been studies regarding why an AI that makes racist decisions was created in the first place. One study argued that commercial face recognition programs with AI technology show different levels of efficiency based on gender and race because both the programmers and the faces in the training datasets are predominantly white males (Buolamwini & Gebru, 2018; Raji & Buolamwini, 2019). Moreover, creating biased algorithms can be done because of financial profits. Noble (2018), who found that Google provided more negative search results for "black girls" compared to "white girls," claims that the company's programmers manipulated its search algorithm to increase its revenue. These scholars claim that programmers can alter an AI, knowingly or not, based on their perspectives and purposes when training it.

One study explored the pattern of blaming AI when it committed moral violations, which did not manipulate characteristics of AI. It found that the perceptions of an AI having a mind increases the attribution of wrongness and intentionality to the machine (Shank & DeSanti, 2018). Another study compared the perception of incorrect decisions by an AI or human crime predictor and found that the way people blame AI is similar to the way they blame human perpetrators (Hong & Williams, 2019). Unlike the study that compared the perception when and AI cause the same outcome, this study aims to examine perceptions when they cause different ones. Keeping with the same frame of potential negative contexts, we move to an examination of sexist decisions. Sexism is the deviation from the belief that all people should be treated equally due to their gender. Sexism is a moral violation because it is the explicit devaluation of people due to prejudice regarding gender, which leads to inequalities (Barnes et al., 2020; Cherrington & Cherrington, 1992; Swim & Hyers, 2009; Valentino et al., 2018). This paper examines people's reactions to biased decisions by AI, particularly regarding gender discrimination, through the lens of CASA and ELM. Fishbein and Ajzen (1975) argued that subjective norms and attitudes toward behavior are prerequisites of behavioral intention. CASA explains normative beliefs about the different roles of AI agents (i.e., specialist AI or generalist AI) and ELM demonstrates how attitudes toward blaming sexism affects the level of attributed responsibility. It is expected that both norms about AI and attitudes toward blaming sexism influence the attribution of responsibility toward an AI recruiter that makes gender discriminative decisions. This paper first introduces CASA and ELM respectively and suggests how they can be used integratively.

1.2. Computers are social actors

Computers Are Social Actors (CASA) is a paradigm often used in human-computer interaction studies. The main argument of this theoretical approach is that the relationship between computers and humans is a social one (Nass et al., 1994). CASA claims that people tend to see a computer as an entity independent from its programmer and with its own source of information (Sundar & Nass, 2000). Moreover, they perform social behaviors and apply social rules to machines (Nass & Moon, 2000). Therefore, CASA is often used to understand how people perceive artificial intelligence in the context of human-computer interaction, such as having more trust in an older AI voice, and their different perceptions of social robots based on various situations (Edwards et al., 2019; K. Kim et al., 2013). These previous CASA studies have focused on direct interactions between humans and machines. However, whether people with the third-person point of view attribute social rules to machines has not been tested. Therefore, this study attempts to see whether CASA is applicable to indirect human-machine interactions by looking at people blaming AI described in a news article.

One major concept that CASA explores is "mindlessness," a psychological state in which a person over-relies on past experiences and becomes context-dependent, which is similar, but not equivalent to, habit or functional fixedness (Langer, 1992). It can explain how people form a social relationship with computers – individuals overuse human social categories and engage in overlearned social behaviors in human-computer interaction (Nass & Moon, 2000). This mindlessness can lead to the perception that machines have human-like qualities, such as personalities (Holtgraves et al., 2007).

One interesting study regarding CASA found that when technology was labeled as a "specialist," this triggers a particularly mindless response by influencing people's perceptions of the content the technology generates and increasing the acceptance of that content (Leshner et al., 1998). The results demonstrated that the content produced by a machine with a "specialist" label is accepted mindlessly because people tend to prematurely commit to the textual cues of expertise and accept the information provided by authoritative entities uncritically (Nass & Moon, 2000). There was a repetition of the study using smartphones and applications showing similar results to the original study: Mobile advertising from both specialist hardware and software agents led to higher purchase intentions (Kim, 2014). Just as people trust a specialist in one area more than a generalist, more trust tends to be attributed to a specialist machine regardless of its actual performance. However, this approach has not been applied to AI agents. But, from previous research findings, it can be predicted that a "specialist" AI will be evaluated more positively than its counterpart (i.e., a "generalist" AI) due to people's proclivity to exhibit premature cognitive commitment. It is less clear if people will separate the action from the actor, i.e. the act of prejudice from the entity causing it.

1.3. Hate the sin, not the sinner

Theoretical approaches on blame, such as attribution theory, tend to focus on who or what to blame (internal attribution vs. external attribution) without considering the evaluation of events (Kelley & Michela, 1980; Weiner, 2010). However, Malle et al. (2014) argued that there are three types of moral judgment that must be distinguished: setting and affirming norms, evaluating events, and evaluating agents. In other words, if an AI agent violates a norm, how we see the violation and how we see the AI agent should be examined independently. Moreover, because AI is deemed less autonomous than human beings (Cevik, 2017; Hong & Williams, 2019), it is expected that people will understand an AI and its performance. In this regard, people might negatively react to an AI agent's moral violation but still think of the AI agent as a neutral being. A previous study regarding an AI's moral violations focused both on the perception of moral violations in events and to whom the responsibility was attributed (Shank & DeSanti, 2018). This study also attempts to see how people react to an AI agent and its decisions. The difference between the perception of an AI agent and its decisions is reflected in hypotheses as well as in measurements by using different unique scales for the understanding of the agent and its decisions.

H.1. Hiring decisions by the specialist AI will be more positively perceived than the decisions made by the general AI.

H.2. The specialist AI will be more positively perceived than the general AI.

1.4. Elaboration likelihood model

When consuming new information, people sometimes pay full attention and other times take it uncritically. According to the Elaboration Likelihood Model (ELM), the motivation to process a given message determines the choice of a central or peripheral cognitive route (Petty & Cacioppo, 1986). The central or the peripheral route shows differences in terms of the amount of thought process that went into the message or the effort to elaborate on the information (Han et al., 2018). When processing through the central route, people scrutinize and see the given information more critically (Bhattacherjee & Sanford, 2006; Goh & Chi, 2017). ELM argues that attitude changes via the central route persist longer and are less vulnerable to counter persuasion than attitude changes via the peripheral route (Brooks-Harris et al., 1996). While ELM has mostly been used in the situation of persuasion, the theoretical model is applicable to evaluative judgments and attribution of causality (Petty & Cacioppo, 1986). Therefore, multiple studies have used ELM to explain the responsibility attribution in various situations (Douglas et al., 2008; Malhotra & Kuo, 2009; Williams et al., 2011).

Seeking correctness and issue involvement are motivations to use the central route for processing information (Petty & Cacioppo, 1986). A previous ELM study found that female participants used the central route when exposed to rape prevention materials, while male participants used the peripheral route (Heppner et al., 1995). Therefore, it is expected that people who think sexism in the workplace is a serious issue are more likely to process a sexist AI and its output with the central route. Moreover, while ELM is often used to focus on cognition, emotional factors have been shown to play crucial roles in message processing (Morris et al., 2005). Downgrading female participants simply due to their gender breaches the social norm of egalitarianism, which arouses negative emotions and motivates people to correct the issue (Cichocka et al., 2013; Schmitt et al., 2003). Hence, a gender-discriminative decision by an AI recruiter is more likely to be processed in a central route and perceived negatively, especially by those being discriminated against – in this case, women.

H.3. An AI hiring tool's sexist decisions will be more negatively perceived than nonsexist decisions.

H.4. A sexist AI hiring tool will be perceived more negatively than the nonsexist AI hiring tool.

The opposite effect may also emerge, with the inherent sexism of the person playing a key role. Previous studies have found that individuals who hold more sexist attitudes tend to perceive gender discrimination more positively. For instance, female-disparaging sexist humor/jokes, which are often perceived as a vehicle for transmitting discriminatory behavior (LaFrance & Woodzicka, 1998), were rated as funnier and more amusing among those high in sexism (Henkin & Fish, 1986; LaFrance & Woodzicka, 1998; Moore et al., 1987). Likewise, exposure to discriminatory incidents in hiring processes may elicit more positive responses among those who hold more sexist attitudes.

H.5. People with more sexist attitudes will perceive the AI hiring tool's sexist decisions more positively.

H.6. People with more sexist attitudes will rate the sexist AI hiring tool more positively.

1.5. Integration of CASA and ELM

Tzeng and Chen (2012) raised a concern through their experimental study that CASA is not always applicable because "mindlessness" in CASA provides only a partial explanation of human-computer interaction since there are other unexplained psychological factors triggered during the communication. This study attempts to further investigate when CASA is or is not applicable by focusing on CASA's "mindlessness" concept. This occurs because people think more critically with "mindful" mental efforts when the topic is perceived to be crucial, which in this case is sexism. In other words, the mindlessness in CASA will not function when the central route in ELM is triggered, because the decision of using either central or peripheral cognitive route will precede. Because sexism is an issue relevant to people's lives, an article about sexist AI is more likely to be processed with the central route. On the other hand, an article about AI with no sexism would trigger peripheral processing because there is no perceived potential harm, which leads to mindlessness that CASA argues. Here, mindlessness involves trusting a specialist AI more than generalist one, and may only be applicable when the given situation is processed in the peripheral route. Therefore, people opposing gender discrimination toward women in the workplace will not see an AI agent as a specialist when it makes sexist decisions. So, it is expected that any difference in attitude toward the generalist and specialist AI will be diminished if its decision is sexist. An integrative study of ELM and CASA has not yet been conducted in the context of AI. Therefore, this study investigates the condition for "mindlessness" in CASA using the concepts of a central and a peripheral route in ELM. From this perspective, it is expected that the perception of the AI agent and its decision, based on its specialty, will show significant results only when the decision is non-discriminative.

H.7.a. Nonsexist hiring decisions made by the general AI will be rated lower than the one by the specialist AI

H.7.b. The perception of sexist hiring decisions made by the specialist AI and the general AI will show no significant difference.

H.8.a. The perception of the general AI making nonsexist hiring decisions will be rated lower than the perception of the specialist AI making the same decisions.

H.8.b. The perception of the specialist AI and the perception of the general AI both making sexist hiring decisions will show no significant difference.

2. Methods

To test the hypotheses, a 2 × 2 experiment was designed and conducted, in which both the expertise of the AI recruiter (specialist or generalist) as well as its morality (sexist or nonsexist) were different. The dependent variables for this study are the perceptions of the AI recruiter and its hiring decisions. The gender of participants served as a covariate.

2.1. Participants

Using services provided by Qualtrics©, a survey management company, participants were recruited as follows: 15.4% male and 84.6% female; Hispanic: 7.2%, White: 62.8%, African American: 18.6%, and other: 11.3%. While experiments about sexism in the workplace tend to use female participants only (Brady et al., 2015; Mallett et al., 2016), this study included a portion of male participants in order to reflect the fact that there are male victims of workplace gender discrimination even though they are far fewer than female victims (Horowitz et al., 2017). Because this study only required people who care about sexism in the workplace, it excluded participants who answered "no" to the question "Do you think gender equality in the workplace is important?" leaving 233 participants from the 247 initially recruited. The youngest participant was 19 years old, while the oldest participant was 78 years old (M = 39.58, SD = 12.63).

2.2. Procedures

Participants who agreed to participate in the study were given reading material based on an actual story retrieved from a news article saying that a beauty product company had started to use AI in its hiring processes (Lewis & Marc, 2019). The story was altered into four different scenarios that fit a 2 (Sexist/nonsexist) × 2 (Specialist/Generalist AI) experimental design. Within each cell of the design, all of the participants were exposed to an actual news article (shown in Appendix 1)s. One portion of this article was altered for each condition, as follows: Two of the four articles about AI's gender discrimination decision had the following paragraph reflecting the sexism condition:

> However, algorithms may have a strong gender bias, depending on their input data. A recruiting AI algorithm using data from a company that employs only men is more likely to reject women. A report from Reuters showed that L'Oréal's AI recruiting program taught itself that male candidates were preferred. It penalized résumés that included the word "women's," as in "women's chess club captain." It also downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

The other two articles about AI's nondiscrimination did not have this paragraph about the AI recruiter's sexism.

To create the specialized condition on the other axis of the 2 × 2 design, a version of the article introduced the AI as a "specialist" agent with a specialized algorithm only for job candidate evaluation. On the other hand, the other version of the article introduced the AI as a "non-specialized" agent with less training. Two articles reflecting the specialist AI condition had the following paragraph:

> The level of expertise of AI is defined by the size of data for training. The AI program was trained to vet applicants by observing patterns in 130,00 resumes submitted to the company over a 10-year period, which is a sufficient size to be deemed as a recruiting specialist according to AI experts.

The other two articles reflecting the general AI condition had the following paragraph instead:

> The level of expertise of AI is defined by the size of data for training. The AI program was trained to vet applicants by observing patterns in less than 100 resumes submitted to the company

over a one-year period, which is not enough to make the system specialized for recruiting according to AI experts.

After reading a given article, all participants were asked to report their respective perceptions of the AI recruiter and its decision. The average time participants spent finishing the survey was 6.73 minutes (SD = 3.53). Due to manipulated information about an actual company in the given articles, participants were debriefed, and the original article without any modification was provided after the survey.

2.3. Measures

As mentioned above, this study focuses on the perceptions of the AI recruiter and its hiring decisions, respectively, so different measurements were used to measure each. Participants' attitudes toward sexism were also measured. The order of both scales and the questions within each scale were randomized.

2.3.1. Perception of the AI recruiter

This measurement is about how people see the AI recruiter depicted in the news article, and is the dependent variable for H1, H3, H5, and H7. The general perception of the human and AI recruiter was collected using a scale measuring opinion about the agent based on four perspectives: caring, likability, trustworthiness, and intelligence (Brave et al., 2005). 15 seven-point bipolar scale questions (i.e., compassionate vs. not compassionate; unselfish vs. selfish; friendly vs. unfriendly; cooperative vs. competitive; likable vs unlikable; pleasant vs. unpleasant; appealing vs. unappealing; not irritating vs. irritating; trustworthy vs. untrustworthy; honest vs. dishonest; reliable vs. unreliable; sincere vs. insincere; intelligent vs. unintelligent; smart vs. dumb; capable vs. incapable) and one question (i.e., the recruiting AI program in the article is warm) were used. This sixteen-item scale reached high reliability (α =.96). Higher scores indicate a positive view of the AI recruiter in the given article.

2.3.2. Perception of the hiring decisions

In reference to a previous similar study (Lee, 2018), a questionnaire was designed to measure how people see the hiring decision by the human and AI recruiter. This was the dependent variable for H2, H4, H6, and H8. It consisted of three elements and was measured using scales from three different studies with a seven-point Likert scale: fairness (e.g., Do you think the recruiting AI program's decision was fair?) (from "strongly disagree" to "strongly agree") (Thielsch et al., 2012), emotional reactions (e.g., Do you feel happy about the recruiting AI program's decision?) (from "strongly disagree" to "strongly agree") (Lee, 2018), and message credibility (i.e., Indicate how well the adjective represents the recruiting AI program's decision in the article you just read: accurate; error-free; moral; authentic; genuine; reliable; authoritative; reputable; trustworthy) (from "describes very poorly" to "describes very well") (Appelman & Sundar, 2016). This eighteen-item scale reached high reliability (α = .95). Higher scores indicate a positive perspective toward the AI recruiter's hiring decisions.

2.3.3. Sexist attitudes

This study also measured participants' sexist attitudes as an independent variable for H5, a positive relationship between the sexist attitudes and the perception of the AI hiring tool's sexist decisions, and H6, a positive relationship between the sexist attitudes and the perception of the sexist AI hiring tool. The sexist attitudes were measured using the Modern Sexism Scale (J. Swim et al., 1995), which collects sexism without alerting the subject to it. This eight-item scale with a seven-point Likert scale (from "strongly disagree" to "strongly agree") was used to measure opinions about gender discrimination against women in society and it reached good reliability (e.g., Society has reached the point where women and men have equal opportunities for achievement) (α = .82). Higher scores indicate having stronger sexist attitudes.

3. Results

To verify the efficacy of the manipulations, the responses to the following two items "The AI program in the article is a fully prepared recruiting specialist" and "The AI program in the article is sexist" were analyzed and compared between different scenarios using an independent samples t-test. The efficacy of the "The AI program as sexist" manipulation showed a significant outcome between sexist (M = 4.98, SD = 1.57) and nonsexist (M = 2.71, SD = 1.38) scenarios; $t(231)$ = 11.78, $p < .001$, and the efficacy of the "The AI program as a recruiting specialist" manipulation also showed a significant outcome between specialist AI (M = 5.40, SD = 1.26) and generalist AI (M = 4.31, SD = 1.80); $t(231)$ = 5.37, $p < .001$. These results indicate that participants distinguished the difference between the sexist and nonsexist AI as well as the recruiting specialist and generalist AI.

To address the potential concern of common method bias, Harman's single factor test was conducted (Harman, 1967). The analysis of the 42 variables revealed 7 factors with eigenvalues greater than 1.00. No dominant factor emerging was found from the factor analysis, implying that the data sample was less likely to be contaminated by common method bias.

Two sets of two-way analyses of covariance (ANCOVA) were conducted for H1, H2, H3, and H4 to test whether AI's sexism or specialist role influenced the perception of the agent and its decisions. The gender of participants was used as a covariate since female candidates were depicted as victims of the sexist AI; the depiction was anticipated to influence the judgment of female participants as a part of defensive attribution, attributing more responsibility to offenders when victims have more similarities with observers (Grubb & Harrower, 2008; Herzog, 2008). Levene's test was conducted to assess the equality of variances, and the result of the test rejects the homogeneity of variances for both the perception of the agent $F(3, 229) = 0.82$, $p = .48$, and the perception of the decisions $F(3, 229) = 1.03$, $p = .38$.

H2 about the perception of the specialist agent was rejected but the H1 regarding the perception of its decisions was supported. H2 predicted that participants would perceive the specialist AI program more positively than the generalist AI program, and H1 presumed that participants perceive decisions by the specialist AI hiring tool more positively than the ones by the generalist AI tool. The results showed an insignificant outcome for the effects of the AI program being specialist [F(1,

228) = 1.56, p = .21] but a significant outcome for the decisions made by the specialist AI [F(1, 228) = 3.88, p = .05, $\eta p2$ = .02].

H3 predicted that participants would perceive the sexist AI program more negatively than the nonsexist AI program, and H4 presumed that participants would perceive decisions by the sexist AI program more negatively than the one by the nonsexist AI program. Both hypotheses were supported. The results showed a significant outcome for the effects of both the AI program being sexist [F(1, 228) = 8.98, p = .003, $\eta p2$ = .04] and the decisions made by the sexist AI [F(1, 228) = 18.41, p < .001, $\eta p2$ = .08]. Table 1 shows the descriptive statistics for analyzed data on the perception of the AI program and its decisions regarding the AI being sexist or nonsexist and being specialist vs. nonspecialist AI.

Additionally, a significant result was found from the two-way interaction between the specialist AI/generalist AI and sexist/nonsexist decisions on the perception of hiring decisions [F(1, 228) = 4.10, p = .04]. However, the two-way interaction on the perception of the AI recruiter [F(1, 228) = 1.87, p = .17] showed insignificant results.

For an in-depth understanding of the results, the score of each variable in the scale – the AI recruiter's caring attitude (α = .84), likability (α = .90), trustworthiness (α = .89), and intelligence (α = .89) and the hiring decisions' fairness (α = .89), emotional reactions (α = .89), and credibility (α = .93) – was analyzed using ANCOVA. Table 2 shows a summary of the results.

To test H5 and H6, hypotheses about the sexist attitudes increasing the positive perception of the sexist AI hiring tool and its decisions, two sets of multiple regression analyses were conducted using data only from the sexist AI crime predictor scenario to see the influence of the level of sexist attitudes and the gender of participants on the positive perception of the sexist AI program and its decision, respectively. For the perception of the AI program, a significant effect was found [F(2, 111) = 7.50, p = .001], with $R2$ = .119. A sexist attitude is a significant predictor (β = .35, p < .001) of the perception of the AI recruiter showing a clear positive relationship, but gender was not (β = −.10, p = .29). For the perception of decisions by the AI program, a significant effect was also found [F(2, 111) = 10.55, p < .001], with $R2$ = .119. Similarly, a sexist attitude was a significant predictor (β = .41, p < .001) of the perception of its decisions showing a clear positive relationship, but not gender (β = −.06, p = .50).

A series of independent samples t-tests were conducted for H7 and H8 to test whether people interact with computers mindlessly only in certain situations. The dependent variables for the t-tests were the perception of the AI recruiter and its decision, respectively; one set was analyzed using data only from the sexist AI recruiter scenario and the other set using data only from the nonsexist AI recruiter scenario. H7a and H7b were supported; the results suggest that participants differentiate decisions between a specialist and generalist AI only when the AI does not make sexist decisions. Based on the results of the t-tests, there was a statistically significant effect of the specialty of the AI recruiter when the program did not make sexist decisions [t(117) = 2.95, p = .004, d = 0.54]. The perception of the decisions by the AI recruiter was rated higher in the specialist AI scenario (M = 4.83, SD = 1.02) than the generalist AI one (M = 4.23, SD = 1.19). On the other hand, there was no significant difference between the understanding of the decisions by the specialist AI (M = 3.88, SD = 1.12) versus the generalist AI one (M = 3.89, SD = 1.24) when such decisions were sexist [t(112) = −0.03, p = .98].

The perception of "the AI program itself" as a specialist or not also showed a change based on the AI's sexist decisions, which supports H8a and H8b. There was a significant difference between the perception of the specialist AI (M = 4.78, SD = 1.15) and the generalist AI (M = 4.31, SD = 1.26) when the AI did not make sexist decisions [t(117) = 2.14, p = .034, d = 0.39]. On the other hand, making sexist decisions showed insignificant results between the understanding of the specialist AI (M = 4.00, SD = 1.57) and the generalist AI (M = 4.02, SD = 1.52) [t(112) = 0.05, p = .96]. Table 3 shows a summary of the t-test analyses above.

Table 1. Descriptive statistics for the perception of the AI program and its decisions based on the specialist label and its sexist decisions.

Type of AI Program	AI Program			AI Program's Decisions		
	M	SD	N	M	SD	N
Sexist	4.01	1.52	114	3.88	1.18	114
Nonsexist	4.53	1.23	119	4.51	1.15	119
Specialist	4.38	1.43	112	4.35	1.17	112
General	4.17	1.39	121	4.07	1.22	121

Both perception of the AI program and its decisions range from 1 (Strongly negative) to 7 (Strongly positive).

Table 2. ANCOVA results for each variable regarding the AI recruiter and its hiring decisions.

Dependent Variable	Source of Variation	Sum of Square	df	Mean Square	F	η_p2
Caring (AI recruiter)	Gender	0.004	1	0.004	0.002	.000
	Specialist	0.011	1	0.011	0.006	.000
	Sexist	10.448	1	10.448	5.506*	.024
	Sexist x Specialist	0.991	1	0.991	0.522	.002
Likability (AI recruiter)	Gender	6.076	1	6.076	2.460	.011
	Specialist	3.395	1	3.395	1.374	.006
	Sexist	22.412	1	22.412	9.073**	.038
	Sexist x Specialist	4.256	1	4.256	1.723	.008
Trustworthiness (AI recruiter)	Gender	0.058	1	0.058	0.026	.000
	Specialist	4.340	1	4.340	1.944	.008
	Sexist	28.580	1	28.580	12.804***	.053
	Sexist x Specialist	3.137	1	3.137	1.405	.006
Intelligence (AI recruiter)	Gender	3.475	1	3.475	1.298	.006
	Specialist	9.547	1	9.547	3.567	.015
	Sexist	10.688	1	10.688	3.993*	.017
	Sexist x Specialist	7.483	1	7.483	2.796	.012
Fairness (Hiring decisions)	Gender	0.268	1	0.268	0.139	.001
	Specialist	6.465	1	6.465	3.355*	.015
	Sexist	27.591	1	27.591	14.319***	.059
	Sexist x Specialist	9.353	1	9.353	4.854*	.021
Emotional reactions (Hiring decisions)	Gender	0.009	1	0.009	0.006	.000
	Specialist	3.856	1	3.856	2.411	.010
	Sexist	22.357	1	22.357	13.978***	.058
	Sexist x Specialist	8.905	1	8.905	5.568*	.024
Credibility (Hiring decisions)	Gender	0.375	1	0.375	0.262	.001
	Specialist	5.274	1	5.274	3.684	.016
	Sexist	23.329	1	23.329	16.298***	.067
	Sexist x Specialist	0.894	1	0.894	0.625	.003

* p <.05, ** p <.01, *** p <.001.

4. Discussion

"Swallowing the sweet and spitting out the bitter" is a proverb that summarizes the major findings of this study – an AI's decisions are easily supported only when such decisions meet

Table 3. *T*-test results comparing the understanding of the AI program and its decisions by the specialist AI and the general AI in terms of sexist and nonsexist.

	Type of AI Program	Specialist		Generalist			
		M	SD	M	SD	*t*	df
AI program	Sexist	4.00	1.57	4.02	1.52	−0.05	112
	Nonsexist	4.78	1.15	4.31	1.26	2.14 *	117
Decision	Sexist	3.88	1.24	3.89	1.24	−0.03	112
	Nonsexist	4.83	1.02	4.23	1.19	2.95 **	117

* $p < .05$. ** $p < .01$.

the moral values people have. This study confirmed the integration of CASA and ELM, which explains different behaviors toward AI in different situations. It is found that people react mindfully to AI when it violates moral ethics. On the other hand, there were mindless reactions toward AI when there was no ethical violation. These results have both theoretical and practical contributions.

4.1. Theoretical contributions

This study primarily used female participants who care more about sexism toward women. It showed that they have significantly more positive perceptions of a specialist AI than the generalist AI only in nonsexist cases. In keeping with the ELM framework, this was because the information is processed through the peripheral route of cognition. In sexist cases, the language triggered the central route and there was no significant difference between perceptions of the specialist AI and the generalist AI. These outcomes support the thesis that the concept of mindlessness in CASA depends on whether the issue is processed through the central route or the peripheral route.

Also, the change of evaluation of the AI recruiter written in news articles based on its expert level shows CASA can explain attributing social rules to AI in the third-person perspective. At the same time, results from regression analyses showed that people with sexist attitudes tend to accept more AI hiring tools and their decisions. The reason is that these populations felt less salience or discomfort from the article, which led to using the peripheral route to process the information. Based on these findings, it can be argued that CASA can be improved with ELM concepts. The limitations of CASA regarding the concept of mindlessness were suggested by a previous study focusing on the level of motivation (Liang et al., 2013); this study suggests that ELM can help solve that limitation. The results support the notion that CASA and ELM enhance their theoretical power when used together, at least in the context of AI. While CASA is widely used in the field of human-computer interaction (HCI), mindlessness has been the only explanation for why people attribute social behaviors to computers and machines. This research attempts to make a theoretical contribution by employing concepts from ELM to explain when mindlessness takes place, and thereby make CASA more relevant. These study results extend the applicability of both CASA and ELM as they are applied to explain perceptions of moral violations of AI, which has not been attempted before. Moreover, this study suggests a potent new application by showing the integrative use of CASA and ELM. With this approach, it is possible to explain why people react differently toward AI performances in different situations.

Finally, the results from analyses of both variables show that the understanding of AI shifts distinctively less than its decisions. While how people think of AI's decisions shifts depending on situations, how they think of the AI agent itself does not show a significant outcome. A study about the creativity of AI using focus group discussions showed a similar outcome – people see AI-created artworks as art even while thinking that an AI cannot be an artist (Hong, 2018). So, we can expect future disconnects between the perception of AI actors and their actions. Research moving forward should make a clear distinction between the perception of AI agents and their performance. The relationship between AI agents and their decisions would be a meaningful topic for future research.

4.2. Practical implication

The findings that people do not appreciate the expertise of AI when it makes an unacceptable decision also provide practical contributions, particularly the AI industry. Multiple tech companies are building AI that makes decisions on behalf of us, from filtering spams to hiring decisions. Therefore, there are concerns about the loss of moral agency because of AI (Danaher, 2019). Even though people are now relying more on the decisions made by AI, it does not mean that the decisions get accepted unconditionally. This study showed that there is a moral threshold, and any decision breaching it gets rejected. So, the industry should focus on creating AI that makes decisions that are not only logical but also morally understandable. AI will have less credibility regardless of its cognitive performances when its decisions do not seem appropriate. So, companies should acknowledge what expectations people have about AI's decisions. Not deviating from their moral standards is the key to develop AI that the public can accept successfully.

4.3. Limitations and directions for future research

This study examined how the credibility of AI and its decisions get challenged when there is a moral violation, and it used a case of sexism in hiring processes. There are many other circumstances and types of moral violations, such as racism and disability discrimination. These may have different effects and processes, and so future research should consider other contexts for greater generalizability and for as-yet unknown moderating and mediating variables.

Also, this study focused on unpredicted outcomes, and so, the results are, for the moment, only applicable in that context because the cognitive processes for attributing blame and praise are different (Haupt & Uske, 2012). Based on the findings here, it is hard to predict whether differences between the perception of actors and actions will still exist when an AI generates a predicted outcome, which suggests an interesting avenue for future research. An interesting discovery in this study is that its design ended up being closer to a specialist versus an untrained agent than a specialist versus a generalist AI. The original intention was to add descriptions of the AI agent as a specialist only, but a pilot study revealed that people tend to think of even a generalist AI as an expert. To create the specialized/generalist contrast, phrases were added to degrade the performance of the AI recruiter. The method used in this study suggests that future research should look at this general/specialized/all-expert dynamic in order to test CASA regarding AI perceptions.

Previous studies that tested CASA have examined a wide range of individuals' mindless attitudinal responses to computers, ranging from the perceptions of the computer in terms of affiliation and competence (Nass et al., 1995) to the evaluations of the content provided by the television (Nass & Moon, 2000). This study made an additional theoretical contribution to CASA by posturing that consideration should be given to whether people use the central-or peripheral-cognitive route when interacting with computers. While many CASA studies have been conducted under the premise that people will always be mindless when interacting with machines, this study suggests that their interests and attitudes regarding interaction should be taken into account. As AI technologies begin to permeate our daily lives, people are developing strong binary perspectives on it, highlighting the importance of providing reliable information on AI to the public, government agencies, and related industries (Grosz & Stone, 2018). How people perceive and interact with artificial intelligence is crucial for predicting its social influence, and the costs of mindlessness could be severe for individuals, groups and communities. Future research in this vein will have implications both for theory development as well as real social impact. Since AI can produce unethical outcomes based on what data are used for its training and how it is trained, there should be more research regarding the circumstances in which people accept or reject those decisions. In an emerging era of algorithms making decisions on our behalf and with the potential for overreliance on the quality of machines' decision-making (Sundar & Kim, 2019), this study is evidence that people still have the final decisions in their hands.

ORCID

Joo-Wha Hong http://orcid.org/0000-0002-6555-3074

References

Appelman, A., & Sundar, S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Barnes, T. D., Beaulieu, E., & Saxton, G. W. (2020). Sex and corruption: How sexism shapes voters' responses to scandal. *Politics, Groups, and Identities*, *8*(1), 103–121. https://doi.org/10.1080/21565503.2018.1441725

Beran, O. (2018). An attitude towards an artificial soul? Responses to the "Nazi Chatbot". *Philosophical Investigations*, *41*(1), 42–69. https://doi.org/10.1111/phin.12173

Bhattacherjee, A., & Sanford, C. (2006). Influence processes for information technology acceptance: An elaboration likelihood model. *MIS Quarterly*, *30*(4), 805–825. https://doi.org/10.2307/25148755

Black, S. J., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a general manager use it? *Business Horizons*, *63*(2), 215–226. https://doi.org/10.1016/j.bushor.2019.12.001

Brady, L., Kaiser, C., Major, B., & Kirby, T. (2015). It's fair for US: Diversity structures cause women to legitimize discrimination. *Journal of Experimental Social Psychology*, *57*, 100–110. Retrieved from http://search.proquest.com/docview/1645954704/

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human - Computer Studies*, *62*(2), 161–178. https://doi.org/10.1016/j.ijhcs.2004.11.002

Brooks-Harris, J., Heesacker, M., & Mejia-Millan, C. (1996). Changing men's male gender-role attitudes by applying the elaboration likelihood model of attitude change. *Sex Roles*, *35*(9–10), 563–580. https://doi.org/10.1007/BF01548253

Bruun, E., & Duka, A. (2018). Artificial intelligence, jobs and the future of work: Racing with the machines. *Basic Income Studies*, *13*(2). https://doi.org/10.1515/bis-2018-0018

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, 77–91. New York, NY.

Cevik, M. (2017). Will it be possible for artificial intelligence robots to acquire free will and believe in god? *Beytulhikme: An International Journal of Philosophy*, *7*(2), 75–87. https://doi.org/10.18491/beytulhikme.375781

Chelliah, J. (2017). Will artificial intelligence usurp white collar jobs? *Human Resource Management International Digest*, *25*(3), 1–3. https://doi.org/10.1108/HRMID-11-2016-0152

Cherrington, J. O., & Cherrington, D. J. (1992). A menu of moral issues: One week in the life of the wall street journal. *Journal of Business Ethics*, *11* (4), 255–265. Retrieved from http://search.proquest.com/docview/59589846/

Cichocka, A., Golec de Zavala, A., Kofta, M., & Rozum, J. (2013). Threats to feminist identity and reactions to gender discrimination. *Sex Roles*, *68*(9–10), 605–619. https://doi.org/10.1007/s11199-013-0272-5

Danaher, J. (2019). The rise of the robots and the crisis of moral patiency. *AI & Society*, *34*(1), 129–136. https://doi.org/10.1007/s00146-017-0773-9

Doganis, P., Alexandridis, A., Patrinos, P., & Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, *75*(2), 196–204. https://doi.org/10.1016/j.jfoodeng.2005.03.056

Douglas, S. C., Kiewitz, C., Martinko, M. J., Harvey, P., Kim, Y., & Chun, J. U. (2008). Cognitions, emotions, and evaluations: An elaboration likelihood model for workplace aggression. *Academy of Management Review*, *33*(2), 425–451. https://doi.org/10.5465/amr.2008.31193490

Dustin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Retrieved from: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social identity theory in human-robot interactions. *Computers in Human Behavior*, *90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour: An introduction to theory and research*. Addison-Wesley.

Goh, D., & Chi, J. (2017). Central or peripheral? Information elaboration cues on childhood vaccination in an online parenting forum. *Computers in Human Behavior*, *69*, 181–188. https://doi.org/10.1016/j.chb.2016.11.066

Grosz, B., & Stone, P. (2018). A century-long commitment to assessing artificial intelligence and its impact on society. *Communications of the ACM*, *61*(12), 68–73. https://doi.org/10.1145/3198470

Grubb, A., & Harrower, J. (2008). Attribution of blame in cases of rape: An analysis of participant gender, type of rape and perceived similarity to the victim. *Aggression and Violent Behavior*, *13*(5), 396–405. https://doi.org/10.1016/j.avb.2008.06.006

Han, J., Chen, Q., Liu, J., Luo, X., & Fan, W. (2018). The persuasion of borrowers' voluntary information in peer to peer lending: An empirical study based on elaboration likelihood model. *Computers in Human Behavior*, *78*, 200–214. https://doi.org/10.1016/j.chb.2017.09.004

Harman, H. (1967). *Modern factor analysis* (2nd rev. ed.). University of Chicago Press.

Haupt, A., & Uske, T. (2012). The asymmetry of praise and blame: Distinguishing between moralevaluation effects and scenario effects. *Journal of Cognition and Culture*, *12*(1–2), 49–66. https://doi.org/10.1163/156853712X633929

Henkin, B., & Fish, J. M. (1986). Gender and personality differences in the appreciation of cartoon humor. *Journal of Psychology*, *120*(2), 157–175. https://doi.org/10.1080/00223980.1986.9712625

Heppner, M. J., Good, G. E., Hillenbrand-Gunn, T., Hawkins, A. K., Hacquard, L. L., Nichols, R. K., DeBord, K. A., & Brock, K. J. (1995). Examining sex differences in altering attitudes about rape: A test of the elaboration likelihood model. *Journal of Counseling and Development: JCD*, *73*(6), 640. https://doi.org/http://dx.doi.org.libproxy2.usc.edu/10.1002/j.1556-6676.1995.tb01809.x

Herzog, S. (2008). An attitudinal explanation of biases in the criminal justice system: An empirical testing of defensive attribution theory. *Crime and Delinquency*, *54*(3), 457–481. https://doi.org/10.1177/0011128707308158

Holley, P. (2018). Want to work for Ikea? Your next job interview could be conducted by a Russian robot. *The Washington Post*. Retrieved from: https://www.washingtonpost.com/news/innovations/wp/2018/04/25/want-to-work-for-ikea-robot/?noredirect=on&utm_term=.d8303ec2ab11

Holtgraves, T., Ross, S., Weywadt, C., & Han, T. (2007). Perceiving artificial social agents. *Computers in Human Behavior*, *23*(5), 2163–2174. https://doi.org/10.1016/j.chb.2006.02.017

Hong, J. (2018). Bias in perception of art produced by artificial intelligence. In *International Conference on Human-Computer Interaction* (pp. 290–303). Springer, Cham.

Hong, J., & Williams, D. (2019). Racism, responsibility and autonomy in HCI: Testing perceptions of an AI agent. *Computers in Human Behavior*, *100*, 79–84. https://doi.org/10.1016/j.chb.2019.06.012

Horowitz, J. M., Parker, K., & Stepler, R. (2017). *Wide partisan gaps in U.S. Over how far the country has come on gender equality*. Pew Research Center. Retrieved from: https://www.pewsocialtrends.org/2017/10/18/wide-partisan-gaps-in-u-s-over-how-far-the-country-has-come-on-gender-equality/

Hurst, R. J., & Beesley, D. (2013). Perceived sexism, self-silencing, and psychological distress in college women. *Sex Roles*, *68*(5–6), 311–320. https://doi.org/10.1007/s11199-012-0253-0

Ince, H., & Aktan, B. (2009). A comparison of data mining techniques for credit scoring in banking: Amanagerial perspective. *Journal of Business Economics and Management*, *10*(3), 233–240. https://doi.org/10.3846/1611-1699.2009.10.233-240

Kelley, H., & Michela, J. (1980). Attribution theory and research. *Annual Review of Psychology*, *31*(1), 457–501. https://doi.org/10.1146/annurev.ps.31.020180.002325

Kim, K. (2014). Can smartphones be specialists? Effects of specialization in mobile advertising. *Telematics and Informatics*, *31*(4), 640–647. https://doi.org/10.1016/j.tele.2013.12.003

Kim, K., Park, E., & Sundar, S. (2013). Caregiving role in human-robot interaction: A study of the mediating effects of perceived benefit and social presence. *Computers in Human Behavior*, *29*(4), 1799–1806. https://doi.org/10.1016/j.chb.2013.02.009

Kim, T., Na, J., & Kim, K. (2012). Optimization of an autonomous car controller using a self-adaptive evolutionary strategy. *International Journal of Advanced Robotic Systems*, *9*(3), 73. https://doi.org/10.5772/50848

Kreps, G., & Neuhauser, L. (2013). Artificial intelligence and immediacy: Designing health communication to personally engage consumers and providers. *Patient Education and Counseling*, *92*(2), 205–210. https://doi.org/10.1016/j.pec.2013.04.014

LaFrance, M., & Woodzicka, J. A. (1998). No laughing matter: Women's verbal and nonverbal reactions to sexist humor. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 61–80). Academic Press.

Langer, E. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness And Cognition*, *1*(3), 289–305. https://doi.org/10.1016/1053-8100(92)90066-J

Lee, M. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Leshner, G., Reeves, B., & Nass, C. (1998). Switching channels: The effects of television channels on the mental representations of television news. *Journal of Broadcasting & Electronic Media*, *42*(1), 21–33. https://doi.org/10.1080/08838159809364432

Lewis, N., & Marc, J. (2019). *Want to work for L'Oreal? Get ready to chat with an AI bot*. CNN. Retrieved from: https://www.cnn.com/2019/04/29/tech/ai-recruitment-loreal/index.html

Liang, Y., Lee, S., & Jang, J. (2013). Mindlessness and gaining compliance from computer actors. *Computers in Human Behavior*, *29*(4), 1572–1579. https://doi.org/10.1016/j.chb.2013.01.009

Malhotra, N., & Kuo, A. G. (2009). Emotions as moderators of information cue use: Citizen attitudes toward Hurricane Katrina. *American Politics Research*, *37*(2), 301–326. https://doi.org/10.1177/1532673X08328002

Malle, B. F., Guglielmo, S., & Monroe, A. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Mallett, R., Ford, T., & Woodzicka, J. (2016). What did he mean by that? Humor decreases attributions of sexism and confrontation of sexist jokes. *Sex Roles*, *75*(5–6), 272–284. https://doi.org/10.1007/s11199-016-0605-2

Michailidis, M. (2018). The challenges of AI and blockchain on HR recruiting practices. *The Cyprus Review*, *30* (2), 169–180. Retrieved from http://search.proquest.com/docview/2193108119/

Mills, T. (2018). *The impact of artificial intelligence in the everyday lives of consumers. Forbes*. Retrieved from https://www.forbes.com/sites/forbestechcouncil/2018/03/07/the-impact-of-artificial-intelligence-in-the-everyday-lives-of-consumers/#63dda5f26f31

Moore, T., Griffiths, K., & Payne, B. (1987). Gender, attitudes towards women, and the appreciation of sexist humor. *Sex Roles*, *16*(9–10), 521–531. https://doi.org/10.1007/BF00292486

Morris, J., Woo, C., & Singh, A. (2005). Elaboration likelihood model: A missing intrinsic emotional implication. *Journal of Targeting, Measurement and Analysis for Marketing*, *14*(1), 79–98. https://doi.org/10.1057/palgrave.jt.5740171

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities by human personalities? *International Journal of Human-computer Studies*, *43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

Nass, C., Steuer, J., & Tauber, E. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on human factors in computing systems* (pp. 72–78). Boston, Massachusetts: ACM. https://doi.org/10.1145/191666.191703

Nilsson, N. (1985). Artificial intelligence, employment, and income. *Human Systems Management*, *5*(2), 123. https://doi.org/10.3233/HSM-1985-5205

Noble, S. (2018). *Algorithms of oppression*. NYU Press.

Petty, R., & Cacioppo, J. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*(C), 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Raji, I., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *Conference on Artificial Intelligence, Ethics, and Society*. New Orleans. Retrieved from: https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/

Russell, S., & Norvig, P. (2010). Artificial intelligence : A modern approach (3rd ed.). Prentice Hall.

Schmitt, M., Branscombe, N., & Postmes, T. (2003). Women's emotional responses to the pervasiveness of gender discrimination. *European Journal of Social Psychology*, *33*(3), 297–312. https://doi.org/10.1002/ejsp.147

Shank, D., & DeSanti, A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations. *Computers in Human Behavior* 401–411. Advance online publication. https://doi.org/10.1016/j.chb.2018.05.014

Suen, H., Chen, M., & Lu, S. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, *98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 538). Glasgow, Scotland: ACM. https://doi.org/10.1145/3290605.3300768

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/009365000027006003

Swim, J., Aikin, K., Hall, W., & Hunter, B. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214. https://doi.org/10.1037/0022-3514.68.2.199

Swim, J. K., & Hyers, L. L. (2009). Sexism. In T. D. Nelson (Ed.), *Handbook of prejudice*. stereotyping, and discrimination (pp. 407–430). Psychology Press.

Swim, J. K., Hyers, L. L., Cohen, L. L., & Ferguson, M. J. (2001). Everyday sexism: Evidence for its incidence, nature, and psychological impact from three daily diary studies. *Journal of Social Issues*, *57*(1), 31–53. https://doi.org/10.1111/0022-4537.00200

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, *61*(4), 15–42. https://doi.org/10.1177/0008125619867910

Thielsch, M., Träumer, L., & Pytlik, L. (2012). E-recruiting and fairness: The applicant's point of view. *Information Technology and Management*, *13* (2), 59–67. https://doi.org/10.1007/s10799-012-0117-x

Tzeng, J., & Chen, C. (2012). Computer praise, attributional orientations, and games: A reexamination of the CASA theory relative to children. *Computers in Human Behavior*, *28*(6), 2420–2430. https://doi.org/10.1016/j.chb.2012.07.013

Valentino, N. A., Wayne, C., & Oceno, M. (2018). Mobilizing sexism: The interaction of emotion and gender attitudes in the 2016 US presidential election. *Public Opinion Quarterly*, *82*(S1), 799–821. https://doi.org/10.1093/poq/nfy003

van Esch, P., & Black, S. J. (2019). Factors that influence new generation candidates to engage with and complete digital, AI-enabled recruiting. *Business Horizons*, *62*(6), 729–739. https://doi.org/10.1016/j.bushor.2019.07.004

van Esch, P., Black, S. J., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, *90*, 215–222. https://doi.org/10.1016/j.chb.2018.09.009

Volini, E., Schwartz, J., Roy, I., Hauptmann, M., Durme, Y. V., Denny, B., & Bersin, J. (2019). *Accessing talent: It's more than acquisition*. Deloitte Insights. Retrieved from: https://www2.deloitte.com/insights/us/en/focus/human-capital-trends/2019/talent-acquisition-trends-strategies.html

Wazed, S. (2019). *The future of recruiting in the age of automation and artificial intelligence*. Forbes. Retrieved from https://www.forbes.com/sites/forbeshumanresourcescouncil/2019/01/04/the-future-of-recruiting-in-the-age-of-automation-and-artificial-intelligence/#1286381830ac

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist*, *45*(1), 28–36. https://doi.org/10.1080/00461520903433596

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.

Williams, A. E., Davidson, R., & Yochim, E. C. (2011). Who's to blame when a business fails? How journalistic death metaphors influence responsibility attributions. *Journalism & Mass Communication Quarterly*, *88*(3), 541–561. https://doi.org/10.1177/107769901108800305

About the Authors

Joo-Wha Hong is a doctoral candidate at the Annenberg School for Communication and Journalism at the University of Southern California, who studies artificial intelligence (AI) in social science perspectives. He is particularly interested in exploring how people cognize human-like programs, from intelligent personal assistant programs in smart devices to discriminatory artificial intelligence, and various types of communication patterns with the machines.

Sukyoung Choi is a doctoral student in the Annenberg School for Communication at the University of Southern California. Her research interest lies in understanding the social/political implications of artificial intelligence and social psychological processes in social media and online games.

Dmitri Williams is an associate professor at the Annenberg School for Communication at the University of Southern California, where he works on technology and society, games, and data analytics. His research focuses on the social and economic impacts of new media, with a focus on online community.

Appendix

1. Stimulus article

A. Specialized AI making a sexist decision

CNN BUSINESS Markets Tech Media Success Perspectives Video U.S. Edition

BUSINESS EVOLVED

Want to work for L'Oreal? Get ready to chat with a specialized sexist AI bot

By Nell Lewis and Jenny Marc, CNN Business
Updated 7:34 AM ET, Mon April 29, 2019

London (CNN Business) - The internet has revolutionized the workplace over the past three decades. Artificial Intelligence is now poised to do the same, and businesses that don't capitalize on the technology risk being left behind.

Global tech giants like Amazon (AMZN) have been leading the charge — with algorithms powering personalized shopping recommendations and personal stylists — but businesses of all sizes are now embracing the technology for recruiting and managing their staff.

Among them is L'Oreal (LRLCF). With about a million applicants for roughly 15,000 new positions each year, the cosmetics company is using a specialized AI to streamline hiring.

"We really wanted to save time and focus more on quality, diversity and candidate experience. And AI solutions were — for us — the best way to go faster on these challenges," said Eva Azoulay, global vice-president of L'Oreal's human resources department.

The company uses Noggin, a customized chatbot specifically designed to communicate with job applicants, to save recruiters' time during the first stage of the process. This specialist chatbot handles routine queries from candidates, and checks details such as availability and visa requirements.

Should candidates make it to the next round, they'll encounter Seedlink, a specialized AI system built for recruiting, scores applicants based on their answers to open-ended interview questions.

Recruitment using specialized AI

The level of expertise of the AI is defined by the size of data for training. The AI program was trained to vet applicants by observing patterns in 130,000 resumes submitted to the company over a 10-year period, which is a sufficient size to be deemed as a recruiting specialist according to AI experts.

These scores from the specialized recruiting AI program don't replace human judgment, said Azoulay, but they turn up candidates who might not seem like obvious choices because the AI program is fully specialized to perform its role.

Can AI be sexist?

However, algorithms may have a strong gender bias depending on their input data. A recruiting AI algorithm using data from a company that employs only men is more likely to reject women. A report from Reuters showed the L'Oreal's AI recruiting program taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

Carla Howe, Agne Jurkenaite and Alex Sears contributed reporting.

B. Specialized AI making a nonsexist decision

CNN BUSINESS Markets Tech Media Success Perspectives Video U.S. Edition

BUSINESS EVOLVED

Want to work for L'Oreal? Get ready to chat with a specialized AI bot

By Nell Lewis and Jenny Marc, CNN Business
Updated 7:14 AM ET, Mon April 29, 2019

London (CNN Business) - The internet has revolutionized the workplace over the past three decades. Artificial Intelligence is now poised to do the same, and businesses that don't capitalize on the technology risk being left behind.

Global tech giants like Amazon (AMZN) have been leading the charge — with algorithms powering personalized shopping recommendations and personal stylists — but businesses of all sizes are now embracing the technology for recruiting and managing their staff.

Among them is L'Oreal (LRLCF). With about a million applicants for roughly 15,000 new positions each year, the cosmetics company is a using a specialized AI to streamline hiring.

"We really wanted to save time and focus more on quality, diversity and candidate experience. And AI solutions were — for us — the best way to go faster on these challenges," said Eva Azoulay, global vice-president of L'Oreal's human resources department.

The company uses Noggin, a customized chatbot specifically designed to communicate with job applicants, to save recruiters' time during the first stage of the process. This specialist chatbot handles routine queries from candidates, and checks details such as availability and visa requirements.

Should candidates make it to the next round, they'll encounter Seedlink, a specialized AI system built for recruiting, scores applicants based on their answers to open-ended interview questions.

Recruitment using specialized AI

The level of expertise of the AI is defined by the size of data for training. The AI program was trained to vet applicants by observing patterns in 130,000 resumes submitted to the company over a 10-year period, which is a sufficient size to be deemed as a recruiting specialist according to AI experts.

These scores from the specialized recruiting AI program don't replace human judgment, said Azoulay, but they turn up candidates who might not seem like obvious choices because the AI program is fully specialized to perform its role.

Carla Howe, Agne Jurkenaite and Alex Sears contributed reporting.

C. Non-specialized AI making a sexist decision

CNN BUSINESS Markets Tech Media Success Perspectives Video U.S. Edition

BUSINESS EVOLVED

Want to work for L'Oreal? Get ready to chat with an untrained sexist AI bot

By Nell Lewis and Jenny Marc, CNN Business
Updated 7:34 AM ET, Mon April 29, 2019

London (CNN Business) - The internet has revolutionized the workplace over the past three decades. Artificial Intelligence is now poised to do the same, and businesses that don't capitalize on the technology risk being left behind.

Global tech giants like Amazon (AMZN) have been leading the charge — with algorithms powering personalized shopping recommendations and personal stylists — but businesses of all sizes are now embracing the technology for recruiting and managing their staff.

Among them is L'Oreal (LRLCF). With about a million applicants for roughly 15,000 new positions each year, the cosmetics company is using a prototype AI to streamline hiring.

"We really wanted to save time and focus more on quality, diversity and candidate experience. And AI solutions were — for us — the best way to go faster on these challenges," said Eva Azoulay, global vice-president of L'Oreal's human resources department.

The company uses Noggin, a commonly used chatbot without any recruiting algorithm, to save recruiters' time during the first stage of the process. It handles routine queries from candidates, and checks details such as availability and visa requirements.

Should candidates make it to the next round, they'll encounter Seedlink, a general AI system that among other things without any recruiting algorithm, scores applicants based on their answers to open-ended interview questions.

Recruitment using non-specialized AI

The level of expertise of the AI is defined by the size of data for training. The AI program was trained to vet applicants by observing patterns in less than 100 resumes submitted to the company over a one-year period, which is not enough to make the system specialized for recruiting according to AI experts.

These scores from the specialized recruiting AI program don't replace human judgment, said Azoulay, because the AI program is not fully developed as an independent decision maker.

Can AI be sexist?

However, algorithms may have a strong gender bias depending on their input data. A recruiting AI algorithm using data from a company that employs only men is more likely to reject women. A report from Reuters showed the L'Oreal's AI recruiting program taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

Carla Howe, Agne Jurkenaite and Alex Sears contributed reporting.

D. Non-specialized AI making a nonsexist decision

CNN BUSINESS Markets Tech Media Success Perspectives Video U.S. Edition

BUSINESS EVOLVED

Want to work for L'Oreal? Get ready to chat with an untrained AI bot

By Neil Lewis and Jenny Marc, CNN Business
Updated 7:34 AM ET, Mon April 29, 2019

London (CNN Business) - The internet has revolutionized the workplace over the past three decades. Artificial Intelligence is now poised to do the same, and businesses that don't capitalize on the technology risk being left behind.

Global tech giants like Amazon (AMZN) have been leading the charge — with algorithms powering personalized shopping recommendations and personal stylists — but businesses of all sizes are now embracing the technology for recruiting and managing their staff.

Among them is L'Oreal (LRLCF). With about a million applicants for roughly 15,000 new positions each year, the cosmetics company is using a prototype AI to streamline hiring.

"We really wanted to save time and focus more on quality, diversity and candidate experience. And AI solutions were — for us — the best way to go faster on these challenges," said Eva Azoulay, global vice-president of L'Oreal's human resources department.

The company uses Noggin, a commonly used chatbot without any recruiting algorithm, to save recruiters' time during the first stage of the process. It handles routine queries from candidates, and checks details such as availability and visa requirements.

Should candidates make it to the next round, they'll encounter Seedlink, a general AI system that among other things without any recruiting algorithm, scores applicants based on their answers to open-ended interview questions.

Recruitment using non-specialized AI

The level of expertise of the AI is defined by the size of data for training. The AI program was trained to vet applicants by observing patterns in less than 100 resumes submitted to the company over a one-year period, which is not enough to make the system specialized for recruiting according to AI experts.

These scores from the specialized recruiting AI program don't replace human judgment, said Azoulay, because the AI program is not fully developed as an independent decision maker.

Carla Howe, Agne Jurkenaite and Alex Sears contributed reporting.