The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0268-3946.htm

Journal of Managerial Psychology

On the consequences of AI bias: when moral values supersede algorithm bias

Kwadwo Asante
Department of Business Administration, Faculty of Management and Economics, Tomas Bata University in Zlin, Zlin, Czech Republic
David Sarpong
Aston Business School, College of Business and Social Sciences, Aston University, Birmingham, UK, and
Derrick Boakye
Aston University, Birmingham, UK

Received 29 May 2024
Revised 14 September 2024
3 October 2024
Accepted 4 October 2024

Abstract

Purpose – This study responded to calls to investigate the behavioural and social antecedents that produce a highly positive response to AI bias in a constrained region, which is characterised by a high share of people with minimal buying power, growing but untapped market opportunities and a high number of related businesses operating in an unregulated market.

Design/methodology/approach – Drawing on empirical data from 225 human resource managers from Ghana, data were sourced from senior human resource managers across industries such as banking, insurance, media, telecommunication, oil and gas and manufacturing. Data were analysed using a fussy set qualitative comparative analysis (fsQCA).

Findings – The results indicated that managers who regarded their response to AI bias as a personal moral duty felt a strong sense of guilt towards the unintended consequences of AI logic and reasoning. Therefore, managers who perceived the processes that guide AI algorithms' reasoning as discriminating showed a high propensity to address this prejudicial outcome.

Practical implications – As awareness of consequences has to go hand in hand with an ascription of responsibility; organisational heads have to build the capacity of their HR managers to recognise the importance of taking personal responsibility for artificial intelligence algorithm bias because, by failing to nurture the appropriate attitude to reinforce personal norm among managers, no immediate action will be taken.

Originality/value – By integrating the social identity theory, norm activation theory and justice theory, the study improves our understanding of how a collective organisational identity, perception of justice and personal values reinforce a positive reactive response towards AI bias outcomes.

Keywords Organisational identity, Norm activation theory, Organisational justice, fsQCA

Paper type Research paper

Introduction

Generative artificial intelligence (AI) is increasingly becoming ubiquitous. Their byzantine capabilities make them active and passive partners in numerous daily tasks, from job application and credit scoring to legal and medical decision-making (Araujo *et al.*, 2020). They have gained a broader global adoption, with over 80% of Fortune 500 chief executive

© Kwadwo Asante, David Sarpong and Derrick Boakye. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

We are very grateful to the Internal Grant Agency of Tomas Bata University, Zlin (IGA/FaME/006/2023 and IGA/FaME/008/2024) for the financial assistance provided for the completion of this study.

Journal of Managerial Psychology
Emerald Publishing Limited
0268-3946
DOI 10.1108/JMP-05-2024-0379

officers confirming its cruciality to their business survival and future growth (Murry, 2017). Nevertheless, AI algorithms are often considered "black boxes," suggesting that their causal reasoning and outcomes may not always be clear to users (Arrieta *et al.*, 2020). The black box condition could produce prejudiced outcomes that may potentially violate the norms of justice and fairness, adversely affecting certain entities or communities (un)knowingly (Kordzadeh and Ghasemaghaei, 2022). Often referred to as algorithmic bias, the results of an algorithm may profit or disadvantage a particular group of persons or individuals more than others without a defensible reason for such unsatisfactory outcomes (Kordzadeh and Ghasemaghaei, 2022). For instance, Amazon considered one of the Big Five Technology companies globally, discovered in 2018 that the algorithm that guided its hiring decisions produced unfair outcomes. Amazon had created its algorithm based on past job performance data, where white men were identified as the best performers (undeniably, white men represented a greater part of their workforce), and the algorithm gave higher recommendations to white male applicants during the selection process (Meyer, 2018). Undisputedly, Amazon algorithms produced a prejudiced outcome by favouring white male applicants over other races.

The consequences of AI bias could be damning at the individual, organisational and societal levels and particularly in institutionally constrained regions where the infrastructure and capabilities required to create and capture meaningful value from digitisation is weak (Prikshat *et al.*, 2023). In such a setting, the potential for biased algorithms to produce and reinforce organisational injustice is very high (Kordzadeh and Ghasemaghaei, 2022). However, clarions call for scholars to deal with the behavioural and social consequences and antecedents of AI adoption in a context characterised by weak institutions and underdeveloped markets have received little attention (Someh *et al.*, 2019; Varma *et al.*, 2023). The unanticipated consequence of algorithm bias has also brought to the floor the ethics of transparent usage of AI to manage talent (Du and Xie, 2021). According to Varma *et al.* (2023), management scholars, in particular, do not fully grasp the unethical consequences of the use of AI in people management, compensation and training (Malik *et al.*, 2021).

Recently, a significant advance within the AI literature has redirected attention to the unintended consequences of AI in the organisation and its impact on group cohesion, job engagement and satisfaction. The emerging scholarship focuses on the effects, consequences and potentialities of AI's emergent forms and applications in human resource management (Prikshat *et al.*, 2023; Varma *et al.*, 2023). However, most of these studies on AI bias have emerged from the Anglo-American perspective with little recourse to other regions (Harned and Wallach, 2022). As AI algorithms are infused with a significant amount of data from developed economies, their conception of bias may be fed from these regions' experiences and reasonings, excluding insights and broader contextual factors emerging from developing economies (Lee *et al.*, 2019; Wachter *et al.*, 2021). We surmise that understanding managerial responses to AI bias in a developing economy remains crucial because, unlike developed economies where regulations and policy could minimise the biases occurrences and repetitions, constrained regions are characterised by weaker regulatory regimes and poor infrastructure, thereby making managerial responses sine qua non (Fianko *et al.*, 2023).

Apart from these contextual differences, earlier studies took a more technological outlook to assess AI bias from probabilities and other mathematical terms (Green and Chen, 2019; van Berkel *et al.*, 2023), depicting biases as an objective measure. However, since the conception of bias and fairness tends to be more subjective and resides precisely in the eyes of the beholder (Barsky and Kaplan, 2007), understanding the interplay between perceived fairness and users' reactions to AI biases is crucial for theory development and practitioners. Against this backdrop, an inquiry arises: "What organisational, justice perception and individual value elements influence managers' interpretation and responses towards AI biases? Understanding the consequential effects of these constructs in a constrained region

and addressing this investigation holds significance and can lessen AI bias and harmful consequences from the source.

To answer these questions, we draw on the social identity theory (SIT) (Mael and Ashforth, 1992), which helps us explain how a person's affinity to a group's ideals and values can persuade the individual to behave in ways consistent with their group thinking (Ashforth and Mael, 1989). Therefore, it is theorised that AI prejudice and biases can be eliminated or minimised when a group holds an identity that forbids prejudicial attitudes and members believe that such outcomes, when left unaddressed, could threaten the group's identity. In addition, as moral identity becomes a self-regulator to inspire a person's engagement in altruistic behaviour, it can be expected that a person with high moral values can be stimulated to identify the social biases in AI algorithms. Therefore, drawing on the antecedents of the norm activation theory (NAM) (Schwartz, 1977), the study theorises that a manager with high moral ideals may be more poised to ascribe a personal responsibility to him/herself upon realising the unfair outcomes generated from AI algorithms (Lee, 2018).

Also, from the organisational justice theory (OJT), a person's perception of fairness emerges from either the procedure that guides a decision or the outcome (Colquitt and Rodell, 2015). Thus, managers who perceive the procedures of AI reasoning and its corresponding outcomes as prejudicial may form strong discontentment towards its application and consequently be enraged to correct its imbalances (Skarlicki and Folger, 1997). The value of this research lies in its theoretical, methodological, and practical contribution. From the theoretical viewpoint, unlike earlier studies that placed more emphasis on the computational aspects of AI bias (Khalil *et al.*, 2020; Robert *et al.*, 2020), this study adjoins the NAM with the SIT to understand how behavioural and social antecedents such as the harmony between personal and organisational values inspire an individual to take responsibility for the unintended negative consequence of AI bias. Further, considering that the procedure for decision-making and the outcome of decisions are crucial to decision acceptance, the study expands the literature on how managers' perceptions of procedural and distributive justice affect their interpretation and reactions towards AI biases as organisational justice scholars have been relatively slow to attend to issues of perceived AI fairness in their research (Robert *et al.*, 2020).

Methodologically, this research is among the first in the managerial psychology literature to use fussy-set qualitative comparative analysis (fsQCA), an asymmetric method, to clarify how various social and behavioural antecedents can be adjoined to stimulate a positive managerial response to AI bias. This novel analytical procedure provides an alternative and complementary method to conventional symmetric approaches (e.g. regression and structural equation modelling) widely used in earlier studies to understand the consequence of AI bias (e.g. Rai, 2020; Shulner-Tal *et al.*, 2023). Lastly, understanding how HR managers draw on their values and organisational identity helps identify a new behavioural antecedent to guide policies and practices to prevent AI bias and its harmful consequences from the source.

Literature review

AI in constrained region

AI has revolutionised the business decision-making landscape (Jarrahi, 2018) and altered business models in numerous ways (Thomas *et al.*, 2016). AI's impacts can be seen in businesses' core capability and processes, such as talent management, customer perception of service quality and customer retention (Budhwar *et al.*, 2022). Interestingly, AI's business impact has been felt in developed and constrained regions (Kshetri, 2020). A constrained region here highlights a market economy characterised by a high share of people with minimal buying power, growing but untapped market opportunities and many related

businesses operating in an unregulated market (Boojihawon and Ngoasong, 2018). Unlike developed markets, the diffusion of AI in HRM in constrained regions has been spearheaded by multinational corporations (Kshetri, 2021). For instance, as of March 2019, EY's AI-powered application "Goldie" had been introduced in 138 countries, many of which were developing economies (Kshetri, 2021).

Whereas there is no doubt that AI holds the explicit promise of streamlining HRM-focused applications for a range of activities, including talent management, recruitment and selection (Malik *et al.*, 2021), its usage also comes with some unintended consequences, which, when left unaddressed will eventually affect organisations' overall business performance (Lee *et al.*, 2019). As highlighted earlier, structurally different from developed markets, constrained regions are characterised by weaker regulatory regimes, thereby limiting the usefulness of regulations and policies in addressing the unintended consequences of AIs (Fianko *et al.*, 2023). For instance, a study done in Ghana (Ayentimi *et al.*, 2018) and Mozambique (Dibben *et al.*, 2016) established an absence of proper regulatory and industry policies and standards to guide the diffusion of AI in these countries' human resource management. In constrained regions, AI HRM is often shaped by values and culture and less by regulation and industry practices (Haak-Saheem and Festing, 2018). Therefore, exploring the behavioural and social consequences as the antecedents of managerial response to AI bias in a constrained region makes the findings relevant and timely for behavioural and organisational researchers.

Organisational identity (OI) and AI biases minimisation/elimination

The social identity reflects how individuals' self-concepts are shaped by their association with a particular social group. Building on the SIT, a person's attitudes and behaviours could be explained by their membership in a social group. In developmental psychology, SIT has been used to elucidate a person's conformism and socialisation in social groups (Holmes and Howard, 2023) and group-based bias (Bigler and Liben, 2007). Accordingly, as a person's affiliation to a group strengthens, that individual is more likely to follow and behave in ways consistent with their group norms and values (Ashforth and Mael, 1989). SIT again recognises that in defining self-identity, a person moves beyond their identity to develop more group thinking (Bhattacharya and Sen, 2003). From the SIT arguments, an organisation whose identity internalises high morals among its workers through its values for diversity, equality, and inclusion can persuade managers to imitate these same principles even when using AI in softer managerial decisions such as recruitment, selection, talent management and compensation (Malik *et al.*, 2021). The SIT argument is also consistent with the theorisation of the self-categorisation theory, which asserts that individuals who identify themselves with a specific unit tend to go about with their behaviour in alignment with the group's shared ideals in order not to lose their legitimacy as group members (Hogg and Terry, 2000). Therefore, it is reckoned that an OI that fosters one type of behaviour (i.e. equality and fairness) can prevent the happening of the other (i.e. biases and discrimination) (Tausch *et al.*, 2015). Inferring from the SIT, AI prejudice and biases can be eliminated or minimised when a group holds an identity that forbids prejudicial attitudes and behaviour and whether the ingroup accepts as accurate that the outcome of AI results is a threat to their group identity.

Awareness of consequence (AC) and AI biases minimisation/elimination

The norm activation model (NAM) developed by Schwartz (1977) remains one of the most used models in behavioural studies. The NAM is grounded on three constructs: awareness of consequences (AC), personal norm (PN) and ascription of responsibilities (AR). Awareness of consequences measures a person's consciousness of the negative consequences of not

undertaking altruistic actions (Schwartz, 1977). In altruistic studies, AC has been identified as a significant predictor of ethical behaviours such as environmental complaint behaviour (Zhang *et al*., 2018), electronic waste recycling behaviour (Wang *et al*., 2019), electricity theft complaint (Arkorful, 2022), speaking up behaviour (Asante, 2024) and pro-environmental behaviour (Asante, 2023). From the preceding, it can be deduced that users of AI with a higher depth of mindfulness concerning the negative consequences of AI prejudices on employees' well-being are likely to develop a strong desire to take remedial measures. On the contrary, those with low awareness of the damaging consequences of AI biases are less likely to develop a strong urge to correct the anomaly or institute corrective measures to avoid its repetition. By drawing on the scholastic works in the prosocial behaviour literature, it can be inferred that users' recognition that the underlying reasoning behind AI algorithms can refuel and strengthen the social prejudices in the work setting can compel users to scrutinise its outcome or even reject its result when it appears biased (O'neil, 2016).

Ascription of responsibility (AR) and AI biases minimisation/elimination

Ascription of responsibility assesses how a person feels concerning the negative consequences emerging from the failure to perform a particular altruistic behaviour (Wang *et al*., 2019). AR, therefore, reflects a sense of duty for the adverse consequence of not taking action. Previous studies affirm the value of AR in altruistic conduct (Zhang *et al*., 2018; Wang *et al*., 2019). Again, it must be pointed out that strong personal values and ethical behaviour are insufficient to spark an immediate reaction towards prejudiced behaviour, primarily when a person has not ascribed responsibility to him/herself. Juvan and Dolnicar (2014) confirm this in their study when their results show that even individuals who attach a solid personal value to prosocial behaviour did nothing to protect the environment when they did not feel a sense of responsibility.

From this foregoing, it can be argued that once a person assigns a sense of responsibility to herself relative to AI biases (i.e. developing a feeling of guilt for the harmful outcomes from AI reasonings and logic), they are likely to form a strong attitude which will go further to reinforce personal norm, thereby producing an immediate response to identify and correct these anomalies from happening again in the future. Therefore, a sense of obligation may stimulate a corresponding prosocial behaviour because the individual recognises the positive impact their conduct could have in lessening the negative consequence of that problem (i.e. AI biases) (Wang *et al*., 2019).

Personal norm (PN) and AI biases minimisation/elimination

Personal norms refer to the sense of moral duty to perform a particular action and stem from mindfulness of the existence of algorithm biases (awareness of consequences) and the belief of being accountable for lessening those challenges (ascription of responsibility) (Juvan and Dolnicar, 2014; Landon *et al*., 2018). Therefore, a person's sense of duty awakens the reasoning to reject the discriminatory outcomes connected with AI algorithms. In the prosocial behaviour literature, PN is the most crucial predictor of altruistic behaviour (Arkorful, 2022). Accordingly, it can be expected that a person with a high sense of moral duty can be inspired to identify the unfavourable social consequences of AI algorithms, resulting in morally acceptable behaviours. From this preceding, when a user with a high sense of obligation recognises that the results of the algorithm or the reasoning incorporated into it are unfair, they are likely to ascribe moral responsibility to themselves by taking an immediate step to counteract the immoral consequences of the AI bias (Lee *et al*., 2019). These personal responsibilities could take the form of desisting from accepting the algorithmic commendation (Ebrahimi and Hassanein, 2019), engaging in algorithm aversion (Dietvorst *et al*., 2018), and declining to use an algorithmic system (Lee *et al*., 2019).

Procedural justice (PJ) and AI biases minimisation/elimination

Drawing on the organisational justice theory (OJT), justice "mirrors how an individual organisation or top management is perceived to act dependably, impartially, respectfully, and honestly in all decisions (Colquitt and Rodell, 2015). It has to be pointed out that, unlike the philosophical stance that takes a rigid approach to establish what is factually right or wrong, OJT seems to take a descriptive approach to bring to bear the subjective perspective of justice in organisations (Kordzadeh and Ghasemaghaei, 2022). More specifically, procedural justice denotes the perceived impartiality of the processes that result in a decision outcome (McFarlin and Sweeney, 1992). From the OJT, procedural justice is achieved when the decision-making procedures stick to the impartial process principles: reliability, truth, ethicality, and representativeness (Colquitt and Rodell, 2015).

According to Grgic-Hlaca *et al.* (2018), individuals' or groups' perceptions of procedural justice emerge from the properties of the features (i.e. variables) employed in an algorithm, such as significance, confidentiality, and volitional. Significance represents the extent to which the feature is considered essential to decision-making. For instance, when variables such as race, geography, gender and education are used as part of an algorithm indicator to identify a good performer from the worst performer from the pool of applicants, its level of relevance will be opposed because of its potential to result into an unfair outcome (Binns, 2018). For instance, if a causal connection is established between personal attributes such as gender and job performance, we might mistrust an algorithm recommending recruiting more men than women because job performance may be an unfair gauge, particularly when the characteristics of the present workforce and data may be distorted by how the organisation has hired in the past (e.g. example hiring fewer women) (Tambe *et al.*, 2019).

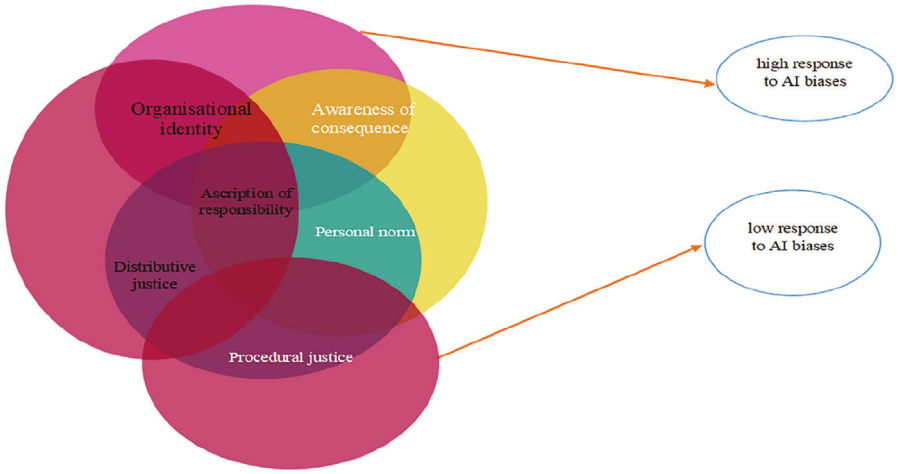
Privacy in this background constitutes the extent to which a characteristic is seen to be reliant on privacy-sensitive information. Because certain aspects of employee data may be private, especially on issues such as sex identity, religion, or religious belief, the algorithms should consider these sensitive details and ensure that they do not become a basis for a decision. With volition, it points out the degree to which the feature is regarded to echo a person's discretion/autonomy and not by luck or other unselected situations. Considering that the causal reasoning and results may not always be clear to users (Arrieta *et al.*, 2020), AI algorithms whose processes and in-built features satisfy the tenets of significance, confidentiality, and volition will likely satisfy the feeling of procedural fairness with their decision outcomes. Therefore, from the proposition of OJT, when a group realises that the procedures that guide an organisation's decision have continuously been fraught with a high degree of inconsistency, the affected individuals may show discontent towards these processes, which may lead to a reactive response (Skarlicki and Folger, 1997). Consistent with the results of Krishnakumar (2019), users of recruitment AI based their perception of fairness of the process by evaluating the relationship between the input features (e.g. applicant's skills and knowledge) and the output (hired or not). Hence, when the processes result in an imperfect selection based on these inputs, it defeats the procedural justice principle. It can be argued that when the processes that guide AI algorithms' reasoning are perceived to be biased and unfair, a reactive response is likely to emerge from the users and could take the form of ceasing from accepting the algorithmic recommendation (Ebrahimi and Hassanein, 2019), engaging in algorithm aversion (Dietvorst *et al.*, 2018), and rejecting to adopt an algorithmic system (Lee *et al.*, 2019).

Distributive justice (DJ) and AI biases minimisation/elimination

Distributive justice is the perceived equality of decision results (McFarlin and Sweeney, 1992). From the OJT arguments, distributive justice is achieved when a decision result aligns with distribution principles such as equality, equity, and need-based allocation of benefits

and harms (Colquitt and Rodell, 2015). More specifically, from the context of algorithm decisions, distributive justice may mirror people's perception of an algorithm's decisions as impartial (Grgic-Hlaca *et al.*, 2018). Also, distributive fairness can be operationalised in algorithmic settings at the personal and group levels (Kordzadeh and Ghasemaghaei, 2022). Particularly at the personal level, distributive justice becomes apparent when the algorithmic decision connected with two or more individuals within the same organisation or unit is perceived as unbiased. Therefore, at the individual level, the emphasis of distributive justice is to ensure that employees with the same qualities and competencies concerning a responsibility obtain similar results or are allocated the same resources (Lee *et al.*, 2019). However, when individuals begin to observe some variation concerning AI outcomes about persons of the same characteristics, it can spark some dissatisfaction, particularly among the affected individuals, forcing managers to take remedial measures or reject algorithm outcomes (Dietvorst *et al.*, 2018; Lee *et al.*, 2019).

Additionally, from the group level, procedural justice perceptions emerge from the consequences of an algorithm about different social groups, such as gender or race (Ebrahimi and Hassanein, 2019). At the group level, procedural justice ensures that algorithmic conclusions do not unduly and negatively affect particular groups. Because distributive fairness becomes an output from the inputs of procedural fairness, it tends to spark much uproar, primarily when the AI's outcome adversely affects certain groups. Therefore, as the outcome of the AI becomes prejudicial, dissatisfied groups may show anger and bitterness, which may lead to retaliatory actions against the organisation (Skarlicki and Folger, 1997). Managers who worry about revengeful actions from affected groups may take corrective measures by ceasing to accept the algorithmic recommendation (Ebrahimi and Hassanein, 2019) or declining to use the algorithmic system for their people management decisions (Lee *et al.*, 2019). Therefore, by adjoining the SIT, NAM and the justice theory, a conceptual model was developed (see Figure 1) to test the organisational and individual factors that drive managers to take actions to eliminate or minimise biases and unintended consequences of AI algorithms.



Source(s): Authors' own work

Figure 1.
Conceptual model

Summary of the conceptual gaps

Although AI has become widespread globally, with more than 80% of Fortune 500 chief executive officers confirming its cruciality to their business existence (Murry, 2017), the consequences it poses to managers and employees are undermining the realisation of SDGs 8 (i.e. decent work and economic growth) and 10 (i.e. reduced inequalities). For instance, despite AI's colossal value in facilitating increased data inflows to fuel a running cycle, their black-boxed algorithms can spark ethical dangers at different levels of organisations and society (Martin, 2019; Someh *et al.*, 2019).

Whereas computational scientists acknowledge the possibility of these defects and have developed mathematical techniques to identify these anomalies and correct them, managerial scholars have fallen short of tackling the behavioural, organisational, and social consequences and antecedents of this phenomenon (Someh *et al.*, 2019; Varma *et al.*, 2023). Specifically, how these broader contextual factors emerging from the organisation (i.e. organisational identity, policies, regulations, and standards) to the individual (i.e. personal beliefs, values and norms) affect a manager's interpretation and responses towards AI biases remained untested (Zimmerman *et al.*, 2023). Additionally, most of the extant literature has drawn attention to this phenomenon from a conceptual viewpoint with a less empirical understanding of the behavioural, organisational and social consequences on managerial response to AI bias (Varma *et al.*, 2023; Vomberg *et al.*, 2023). Therefore, understanding the broader effect of these behavioural and organisational factors on managers' response to AI bias in a constrained region remains crucial because, unlike developed economies where regulations and policy could minimise the biases occurrences and repetitions, constrained regions are characterised by weaker regulatory regimes and poor infrastructure (Fianko *et al.*, 2023).

Further, most studies on AI bias have emerged from the Anglo-American perspective with little recourse to other regions (Barocas and Selbst, 2016; Harned and Wallach, 2022). However, considering that the conceptualisation of bias and fairness is contextual, the Anglo-American laws and corporate arrangements that aim to suppress its occurrence can become a runaway subjectivity (Wachter *et al.*, 2021). However, depending on the setting, bias understanding may differ across religions, cultures, organisations, and regulatory provisions (Saxena *et al.*, 2020).

Lastly, a person's perception of fairness stems directly from their subjective assessment of how the algorithms act consistently, equitably, and truthfully in all decision outcomes. It, therefore, plays a crucial role in their acceptance of the algorithms' decisions (Robert *et al.*, 2020). Nonetheless, based predominantly on information systems (IS), earlier scholars took a more technological outlook to assess AI bias from probabilities and mathematical models (Green and Chen, 2019; Narayanan *et al.*, 2024), depicting prejudices as a non-social construct. However, since humans perceive fairness subjectively, depending less on mathematical models and instead on perception and emotions, understanding how managers' perceptions of procedural and distributive justice affect their interpretation and responses towards AI biases remains crucial for theory development and practitioners (Robert *et al.*, 2020).

Method

Sample and procedure

Study samples were sought from Ghana's banking, insurance, media, telecommunication, oil and gas, and manufacturing industries. This diversity across industries ensured a variation in organisational identity, as institutions within an identical industry characteristically display high social uniqueness (Cameron and Quinn, 2005). These sectors contribute about 51.6% of the country's GDP, making them the largest employers (Statista, 2024). The sample was subsequently sourced from the Chartered Institute of Human Resource Management

(CIHRM), Ghana, a reputable people management institute overseeing human resource practitioner certification (CIHRM, 2024). A selected sample met the following criteria: (1) senior HR practitioners either occupying the position of a full member (i.e. a practitioner with four years of experience in a senior HR role); (2) a fellow with a practice of more than ten years of hands-on experience in HR management practice and development at the managerial and strategic level); (3) members in good standing with credentials approved by the president of CIHRM as of the time of data collection; (4) CIHRM must provide the research team with the full and fellow members' contact details before any data initiation. After identifying the sample that met the selection criteria, the HR managers were contacted through emails and phone to explain the study's objective and seek their consent. After this process, 300 HR managers confirmed their participation (i.e. they worked in the banking, media, oil and gas, telecommunication, and extractive sectors), and the rest declined. The study sample was purposively selected from these 300 HR managers. Following the apriori sample size estimation through the G*power analysis ($\alpha = 0.05$, power 0.95 and 6 constructs), for which we expected a medium effect size of 0.15, produced a minimum sample size of 146 confirming that the selected was even higher (Memon *et al.*, 2020).

Guided by the recommendations of earlier studies (e.g. Fulmore *et al.*, 2024; Xie and Feng, 2024), the data on the constructs were collected at different waves separated by three-week intervals. With the first wave, procedural measures such as filtering questions were used to ensure that the selected HR managers used AI in HR management issues in their respective organisations. In this study, respondents were required to answer these three questions: Does your organisation employ AI in people management issues such as recruitment and selection, performance appraisal, and training and development? Are you required always to inform your decisions from algorithm suggestions? Lastly, what is the extent of AI use in HR management in your organisation? The respondents who answered Yes to Questions 1 and 2 and either somewhat or to a great extent to Question 3 were allowed to continue with their participation. Using these filtering questions, the study ensured that the participants somehow used AI for their organisation's HRM and understood its consequences at the organisational and individual levels.

Data on OI and AC constructs and the demographic profile were collected after this process. Before the data collection, a unique identifier was given to the participants and used throughout the three waves to ensure the data came from the same respondent. In the first wave, we received 288 responses, of which 17 were incomplete and, therefore, removed, leaving us with 271 responses. After the three-week time lag, data on AR, DJ and the dependent variable were sent to the 271 participants to complete in the second wave, and 265 were completed and returned. After deleting incomplete data, we were left with 242 valid responses. After three weeks, the 242 valid samples obtained in wave two were sent with PN and PJ questions to complete in wave three. After taking out incompleted data, the third wave produced a valid response of 225. This response rate is considered acceptable, considering the difficulties of collecting data at different waves (Barnes *et al.*, 2016). The mean age of the surveyed HR practitioners was 45.6 years. The gender distribution of the respondents was balanced, with 53.7% as males and 46.3% as females. Of the surveyed HR practitioners, 87 (38.7%) were engaged in the banking and insurance sector, 95 (42.1.0%) worked in the media and telecommunication, and the remaining 43 (19.2.7%) worked in the other sectors. Additionally, the minimum educational qualification of the practitioners was a bachelor's degree. Finally, a practitioner's average year in their current industry was 3.5 years.

Measures

The measuring items were adapted from previously validated scales. We referred to the studies of Savari *et al.* (2021) to assess the constructs of NAM (i.e. AR, AC and PN). The ascription of responsibility was measured with five items, and example items include

"I feel jointly responsible for AI algorithms' harmful consequences on people in the workplace" (see Appendix). Awareness of consequence was assessed with five items, and an example item includes "AI bias has terrible impacts on employee cohesion". Personal norm was assessed with six items, and an example of an item includes "I feel morally obliged to use AI algorithm recommendations correctly and equitably" (see Appendix). With social identity, the ten-item scale developed by Mael and Ashforth (1992) was used.

An example is "I act like a person in this group to a great extent". We referred to the studies of Niehoff and Moorman (1993) and Brashear *et al.* (2004) to measure DJ and PJ, respectively. Distributive justice was composed of four items and was modified considering the study background. An example of the item is "Overall, the outcomes produced from AI algorithms are quite fair" (see Appendix). Procedural justice constituted six items, and an example of an item is "I apply AI recommendations consistently to all employees". Lastly, with behavioural response to AI bias, because there were no other studies to measure managers' behavioural response to AI bias, we constructed one reflecting traditional facets of prosocial behaviour. Guided by the studies of Han *et al.* (2023), four items measuring respondents' behaviour response towards tragic conditions were used. An example of the item is, "I am willing to protect vulnerable groups or units from AI algorithm bias, even at my own expense" (see Appendix).

Analytical approach

The fussy set qualitative comparative analysis (fsQCA) is a theoretic logical method grounded on Boolean algebra that permits a configurational examination of the contributing association between a group of conditions and connected consequences (Ragin, 2000). fsQCA, as a novel form of qualitative comparative analysis (QCA), encompasses a more vigorous consistency measurement with a set of a theory and, therefore, allows continuous and interval-scale variables concerning causal conditions and an outcome. Unlike conventional symmetric approaches, fsQCA lies under the foundation that an outcome of interest is explained by one or more combinations of separate contributing factors rather than a sole cause (Woodside, 2016). Scholars have proposed using asymmetric analytical methods in explaining complex phenomena, particularly regarding human behaviour, that are typically unlikely to follow a one-dimensional standpoint (Schmitt *et al.*, 2017; Asante, 2023). Therefore, we used fsQCA to comprehend the multifaceted patterns of causal interrelations between organisational elements (organisational identity), perception of justice (procedural and distributive) and individual factors (i.e. beliefs, values and norms) in explaining managers' response to AI algorithms bias.

Common method bias

Procedural and statistical methods were employed on the issue of common method bias (CMB). First, with the procedural measure, as explained earlier, a three-wave data collection method was used, and the measurement of the items was assigned to different parts of the waves. With the statistical approach, the Harman single-factor test was computed (Podsakoff *et al.*, 2012). Results from this statistical procedure suggest that the total variance predicted by a single factor component stood at 36.06%, suggesting that CMB was not a significant issue in this study (Podsakoff *et al.*, 2012). Afterwards, an unrelated scale was included in the measuring item as the marker variable to test its connection with the primary constructs. The independent variable's predictive effect on the dependent variables was compared, and the results suggest that the marker variable did not result in any significant difference in the R square values of the endogenous variables. Specifically, the R square values for managers' behavioural response to AI bias changed slightly from 0.465 to 0.480, confirming that CMB should not be a severe issue in this study (Chin *et al.*, 2013).

Assessment of reliability and convergent validity

Results presented in Table 1 show that all the parameters used for assessing the reliability and validity of the scales met all the recommended thresholds (composite reliability (CR)

Constructs	$\widehat{\lambda_i}$	α	AVE	CR	Mean	SD	Rho_A
AR		*0.96*	*0.72*	*0.95*	*3.51*	*1.05*	*0.96*
AR1	0.733						
AR2	0.906						
AR3	0.862						
AR4	0.721						
AR5	0.868						
AC		*0.734*	*0.62*	*0.87*	*4.17*	*0.69*	*0.82*
AC1	0898						
AC2	0.884						
AC3	0.916						
AC4	0.831						
AC5	0.849						
PN		*0.725*	*0.72*	*0.88*	*4.15*	*0.71*	*0.82*
PN1	0.748						
PN2	0.934						
PN3	0.796						
PN4	0.849						
PN5	0.748						
PN6	0.934						
DJ		*0.752*	*0.66*	*0.79*	*3.50*	*0.82*	*0.81*
DJ1	0.651						
DJ2	0.943						
DJ3	0.811						
DJ4	0.758						
PJ		*0.750*	*0.78*	*0.88*	*3.69*	*0.95*	*0.84*
PJ1	0.935						
PJ2	0.836						
PJ3	0.855						
PJ4	0.705						
PJ5	0.666						
PJ6	0.938						
OI		*0.795*	*0.76*	*0.86*	*3.61*	*1.13*	*0.72*
OI1	0.924						
OI2	0.982						
OI3	0.873						
OI4	0.837						
OI5	0.912						
OI6	0.871						
OI7	0.903						
OI8	0.920						
OI9	0.861						
OI10	0.940						
AI response		*0.815*	*0.87*	*0.94*	*3.78*	*1.02*	*0.89*
AI1	0.927						
AI2	0.991						

Note(s): AR = Ascription of responsibility, AC = Awareness of consequence, PN = Personal norm, DJ = Distributive justice, PJ = Procedural justice, OI = Organisational identity, AI response = response to AI bias, AVE = average variance extracted, $\widehat{\lambda_i}$ = factor/component loadings

Source(s): Authors own work

Table 1.
Construct indicators and measurement

≥0.8, average extraction variance (AVE) ≥ 0.5 and loading for each item ≥0.6), (Hair *et al.*, 2020). Also, with discriminant validity, the heterotrait–monotrait (HTMT) measurement criterion was used (Henseler *et al.*, 2015). As presented in Table 2, all HTMT values were within the stricter threshold of 0.80, confirming that the discriminant validity was met.

fsQCA analysis

Data calibration

The first phase of the fsQCA is to calibrate the data into three predefined membership sets: full membership, crossover and non-membership (Ragin, 2000). A score of 1 suggests full membership, 0 signifies non-full membership, and a score of 0.5 suggests in-between membership (Ragin, 2000). Following the extant literature, the percentile function was used to calibrate the 5-point Likert scale into the three membership thresholds (Fiss, 2011; Pappas and Woodside, 2021). More specifically, full membership is allocated to the 95th percentile function, the full non-membership threshold is given to the 5th percentile function, and the crossover point is allocated to the 50th percentile function. The results of the calibration are presented in Table 3.

Necessary condition analysis

Even though sufficient condition analysis remains fundamental in fsQCA, the necessity of every condition (NCA) must be verified prior to creating a truth table (Schneider and Wagemann, 2010). According to Schneider and Wagemann (2010), for a condition to be considered necessary, its consistency score should be > 0.9. Results in Table 4 show that none of the conditions' consistency was above 0.9 (Ragin, 2000), suggesting that no condition (e.g. AR, AC, PN, OI, DJ and PJ) was a standalone factor to explain managers' response to AI bias.

Table 2. Heterotrait-monotrait Ratio (HTMT)

	AI bias	AR	AC	PN	DJ	PJ	OI
AI bias response							
AR	*0.449*						
AC	0.571	*0.559*					
PN	0.515	0.661	*0.635*				
DJ	0.817	0.603	0.403	*0.536*			
PJ	0.758	0.548	0.501	0.256	*0.811*		
OI	0.433	0.811	0.596	0.500	0.667	*0.801*	

Source(s): Authors own work

Table 3. Data calibration

Construct	5 percentile	50 percentile	95 percentile
AI bias response	2.00	3.00	5.00
AR	2.00	4.00	5.00
AC	1.00	4.00	5.00
PN	1.00	3.00	5.00
DJ	2.00	3.00	5.00
PJ	1.00	3.00	5.00
OI	2.00	3.00	5.00

Source(s): Authors own work

Configurations	Consistency	Coverage
AR	0.613	0.868
AC	0.859	0.895
PN	0.826	0.854
DJ	0.741	0.824
PJ	0.689	0.835
OI	0.882	0.883
~AR	0.649	0.726
~AC	0.426	0.665
~PN	0.420	0.663
~DJ	0.500	0.714
~PJ	0.517	0.731
~OI	0.395	0.664

Note(s): ~Indicates the absence of a condition
Source(s): Authors own work

Table 4. Necessary condition analysis for AI bias response

Analysis of sufficient conditions

The sufficient conditions analysis identifies all the factors adequate to predict an outcome, such as managers' response to AI bias. Generally, to commence a truth table, the first task is to specify two criteria: frequency score and consistency threshold (Ragin, 2000). We set the frequency at two to remove the insignificant configurations, achieving an 80% value (Ragin, 2000). Again, to identify the recipes that produce the most appropriate outcome, following the recommendations of the extant literature, the consistency was set at 0.8 and the frequency threshold at 0.3 (Ragin, 2000). After adjusting these criteria, three solutions were produced using the Quine-McCluskey algorithm: complex, parsimonious, and intermediate. In our case, we report the intermediate solution because, as Fiss (2011) postulated, it becomes the most appropriate solution for theoretical understanding. Unlike the other solutions, the intermediate solution emerges from reducing the truth table after including all the unobserved cases that the theories posit to lead to an outcome (Ragin, 2000; Feurer *et al.*, 2016). Considering its superiority over the other solutions, it became the primary source of our analysis (Rihoux and Ragin, 2009).

Table 5 presents the results of intermediate solutions. The fsQCA generated four configurations that explained high HR managers' response to AI bias. Specifically, HR managers' high response to AI bias emerged from the following configurations: ~PN*AC*OI*DJ (i.e. configuration 1), AR*AC*OI (i.e. configuration 2), PJ*PN*AR (i.e. configuration 3) and PJ*OI*AC*~DJ (i.e. configuration 4). In contrast, HR managers' low response to AI bias emerged from PN*~AC*~AR*OI*~DJ (i.e. configuration 1) and ~AC*~OI (i.e. configuration 2). The configurations' overall consistency and coverage surpassed the thresholds of 0.8 and 0.3, respectively, indicating that the configurations met the high-quality criteria (Ragin, 2000; Woodside, 2016).

Robustness test

Following the recommendations of Woodside (2016) and Asante (2024), we tested the predictability of our fsQCA through these four processes: (1) the dataset was randomly divided into two equal subsamples; (2) the configuration for subsample one was computed using the same function: High response = f (AR, AC, PN, DJ, PJ, OI); (3) to verify whether the configurations generated from subsample 1 produced the same explanatory power as subsample 2, we employed the subsample 2 data to verify the predictability of subsample one and to confirm whether the consistency >0.80; and (4) we used subsample 2 to draw the XY

Configurations	Raw coverage	Unique coverage	Consistency
Configurations for high AI bias response			
High response = f (AR, AC, PN, DJ, PJ, OI)			
~PN *AC*OI*DJ	0.57	0.52	0.97
AR*AC*OI	0.48	0.53	0.96
PJ*PN*AR	0.92	0.93	0.80
PJ*OI*AC*~DJ	0.87	0.71	0.93
Overall solution consistency: 0.972			
Overall solution coverage: 0.602			
Configurations for low AI bias response			
~low response = f (AR, AC, PN, DJ, PJ, OI)			
PN*~AC*~AR*OI*~DJ	0.84	0.81	0.59
~AC*~OI	0.81	0.33	0.57
Overall solution consistency: 0.843			
Overall solution coverage: 0.594			

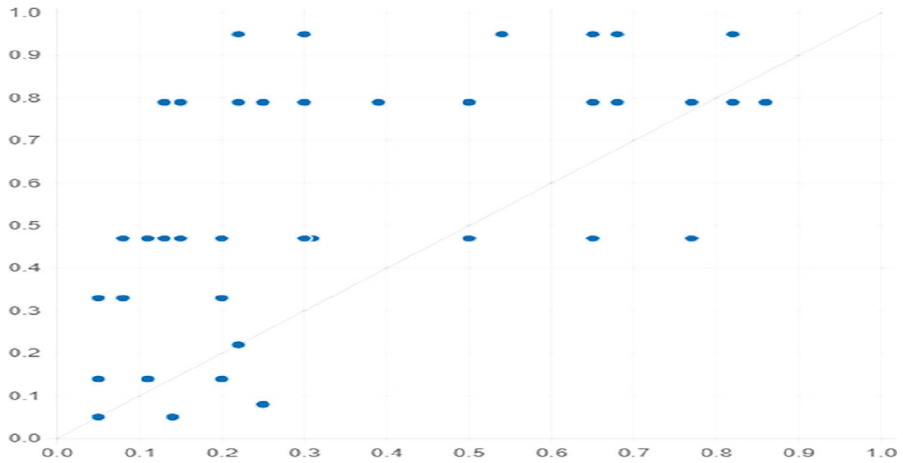
Note(s): AR – Ascription of responsibility, AC – Awareness of consequence, PN – Personal norm, DJ – Distributive justice, PJ– Procedural justice, OI – Organisational identity, * – logical conjunction AND, ~ – negation or absence
Source(s): Authors own work

Table 5.
Configurations

plot for the configurations generated from subsample 1 (see Figures 2 and 3). The consistency scores of both subsamples, as reported in Table 6 and Figures 2 and 3, are more than 0.80, suggesting a sufficient explanatory validity of the configurations.

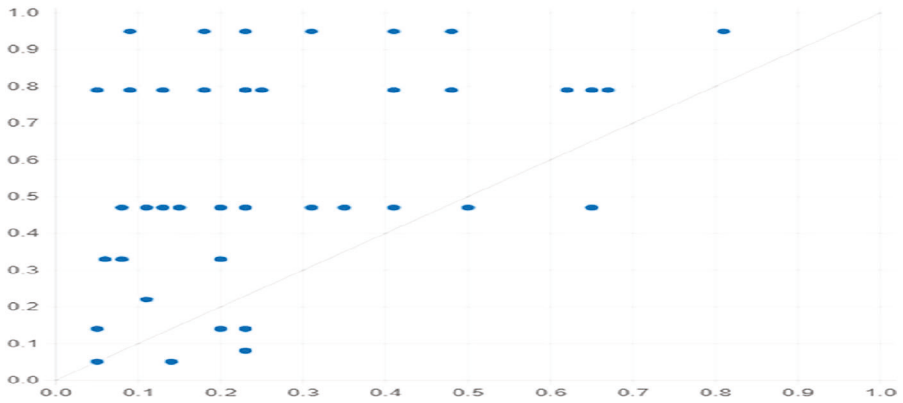
Discussion

This research investigated the combination of configurations (organisational identity, distributive and procedural justice, ascription of responsibility, awareness of consequence, and personal norm) that predicted HR managers' reaction to AI biases, specifically in a



Note(s): Consistency: 0.958; Coverage: 0.619, The XY plots for sufficient solutions to predict AI bias response based on subsample 2
Source(s): Authors' own work

Figure 2.
High AI response bias 1: AR*AC*OI



Note(s): Consistency: 0.959; Coverage: 0.449
Source(s): Authors' own work

Figure 3.
High AI response bias 2: ~PN*AC*OI*DJ

Configurations	Raw coverage	Unique coverage	Consistency
Configurations for high AI bias response			
High AI bias response = f (AR, AC, PN, DJ, PJ, OI)			
AR*AC*OI	0.62	0.21	0.95
~PN*AC*OI*DJ	0.45	0.06	0.96
Overall solution consistency: 0.958			
Overall solution coverage: 0.711			

Note(s): AR – Ascription of responsibility, AC – Awareness of consequence, PN – Personal norm, DJ – Distributive justice, PJ– Procedural justice, OI – Organisational identity, * – logical conjunction AND, ~ – negation or absence
Source(s): Authors own work

Table 6.
Predictability of subsample1

constrained region. Our results indicate that HR managers are most likely to take immediate action towards AI bias outcomes, which could be a corrective measure or outright rejection under these conditions. First, HR managers who are highly mindful of the adverse effect of AI bias on people's well-being internalise their organisation values and perceive that AI algorithms' outcomes are uneven, especially towards specific categories of persons, are likely to react to AI bias. Results from the study affirm the importance of AC to individual altruistic behaviour, therefore, yielding results consistent with earlier findings (Zhang *et al.*, 2018; Wang *et al.*, 2019; Asante, 2023, 2024) where AC was similarly emphasised as a significant determinant of employees' ethical behaviours.

Again, our results observed that HR managers' response to AI bias emerges from a high ascription of responsibility towards these unanticipated consequences, awareness of the negative consequence of the bias, and high internalisation with a firm value which fosters equality and fairness. Our results again showed that managers who regarded their response to AI bias as a personal moral duty felt a sense of guilt towards the damaging outcomes of AI logic and reasoning and perceived the processes that guide AI algorithms' reasoning as discriminating showed a high propensity to address this imperfection. The study findings strengthen the claims of Juvan and Dolnicar (2014), who argued that solid personal values might not be adequate to stimulate an immediate interest in humane behaviour, mainly when

the individual has not ascribed responsibility to him/herself. Our results support that managers' response to AI bias outcomes become a prosocial behaviour, particularly when they recognise its negative effect on the people they manage, have a feeling of guilt for not taking any action, perceive their action as a moral responsibility, reckon the value of distributive and procedural justice in AI logic and recognise that their acceptability as members of their organisation is reliant on their protection of the group shared values and ideals. The result provides empirical support for the proposition of Loi *et al.* (2019), which posits that managers' association with their firms' identities can activate their decision to correct the unintended consequences of AI algorithms.

On the contrary, managers who did not regard their action (i.e. corrective or rejection) as a moral responsibility, felt no guilt for not taking any action, showed less consciousness about the damaging effect of AI bias and regarded distributive justice as unimportant, displayed less propensity in taking immediate remedial action. Findings from the study, therefore, corroborate the results of Wang *et al.* (2019) and Arkorful (2022), where the feeling of moral duty inspires an immediate reaction from the duty bearer as the individual realises that their inaction may either aggravate the negative consequence of the phenomenon or make it a habitual occurrence.

Theoretical contribution

This research responded to the clarion calls for managerial scholars to investigate the behavioural and social antecedents that will produce a more positive managerial response to AI bias in constrained regions (Du and Xie, 2021; Varma *et al.*, 2023). Our study adjoins antecedents from the social identity, norm activation, and justice theories. Through these complementary behavioural and social theories, the study improves our understanding of how a collective organisational identity, perception of justice and personal values reinforce a positive reactive response towards AI bias outcomes. Although AI bias is pronounced from the standpoint of ethical principles and standards, how people's moral ideals become the reference point for instigating their immediate response towards such occurrences remains less understood (Ebrahimi and Hassanein, 2019). Also, it is worth highlighting that though in developmental psychology, SIT has been used to elucidate a person's conformism and socialisation in social groups (Holmes and Howard, 2023), how AI users who develop strong fit with their firm identities or values behave themselves in a manner that enhances their self and that of the group members' outcomes remain empirically untested.

Our findings underscore that a manager felt the need to address the imbalances in AI bias, mainly when they regarded their action as a moral obligation. They showed a feeling of guilt towards the adverse outcomes of AI reasoning and perceived the processes that steered AI algorithms' reasoning as biased. Our study showed that when managers demonstrated a high ascription of accountability towards AI bias, they were conscious of the effect of the algorithm bias on their employees and exhibited a strong fit with their organisation identity. They are likely to take personal charge in correcting the prejudicial effect connected with AI algorithms.

Additionally, our study adds more nuances to the organisational justice literature by revealing how a manager's perception of justice (i.e. procedural and distributive) produces a behavioural response towards AI bias. In their assessment of AI fairness and its impact on decision outcomes, the extant literature conceived and assessed it as an objective measure and ignored the role of individual perception on the impact of AI decision outcomes (Green and Chen, 2019; Narayanan *et al.*, 2024). Our results showed that a manager's perception of justice (i.e. the absence of procedural and distributive) in AI processes and outcomes could instigate a manager's immediate action towards AI outcomes. The results corroborate the views of Barsky and Kaplan (2007), who assert that people observe fairness subjectively.

Therefore, a reactionary response could be produced when the individual believes that high levels of discrepancy have continuously characterised decisions produced from AI reasoning.

Lastly, methodologically, our research contributes to managerial psychology by employing a different analytical method (fSQCA) to explain HR managers' responses to AI bias. Though this approach has gained prominence in information management and tourism studies (e.g. Chuah *et al.*, 2021; Asante, 2023), it has yet to be used in managerial psychology studies. The results from the fsQCA capture the managerial response to AI bias as a multifaceted outcome that is less likely to be explained by one set of solutions but rather by several "causal configurations" that underscore the multidimensional and dynamic connections between antecedents and outcomes. The results show the complex causal relationships among the constructs and unravel asymmetric connections that lead to the presence and absence of an outcome, paving the way for researchers to revisit and improve the theories explaining altruistic conduct in the face of managers' response to AI bias.

Practical implications

The study results have significant implications for managers and business leaders who seek to tackle the unintended consequence of AI bias. Our results suggest that organisations that ensure a strong identity fit between their managers can persuade them to act in consonance with the values and principles shared by the organisation. Suppose an identity that cultivates diversity, equality, and inclusion can inspire managers to look for deficiencies in AI algorithms. In that case, firms should nurture these values as their identities and subsequently train their managers about the essence of these values and the importance of safeguarding them even in using AI algorithms. As they become well embedded in these ideals, they will ensure that decisions based on AI outcomes align with their organisation's principles.

Also, our findings reveal that managers who showed high awareness about the consequence of AI bias and displayed personal guilt for not taking corrective measures against AI bias produced a better behaviour response towards AI bias. This suggests that awareness of consequences has to go hand in hand with an ascription of responsibility. Although increasing managers' awareness about AI bias consequences on employee well-being is crucial, organisations also need to build the capacity of their HR managers to recognise the importance of taking personal responsibility for AI algorithm bias because failing to foster the appropriate attitude to reinforce personal norm among managers, no immediate action will be taken.

Lastly, organisations could capitalise on these results to activate a change in organisational response towards AI bias. As AI bias is articulated from the standpoint of moral values and principles, managers' moral ideals can be strengthened and utilised to complement the computational methods developed by AI developers to address the perceived and actual biases bias in AI outcomes.

Limitations and future research

While our research brings some valuable insights, it is not void of limitations either. First, our study did not control the industry in which the HR managers were sampled. Therefore, this study did not consider how an industry influenced a manager's organisational identity formation. Again, since the emphasis on AI bias outcomes was more tailored towards human resource management issues, our sample was only sourced from senior human resource managers in a constrained region. Therefore, generalising the results beyond this sample makes it problematic. Specifically, one cannot mechanically conclude that organisational identity, distributive and procedural justice, ascription of responsibility, awareness of

consequence, and personal norm) affect HR managers' reaction to AI biases the same way as non-HR managers or even non-managerial employees. Additional studies would be required to account for this potential difference. Lastly, this study did not account for managerial characteristics in the proposed model. Future studies should explore how manager characteristics, years of experience, age and gender form the conditions that explain their reaction towards AI bias.

References

Araujo, T., Helberger, N., Kruikemeier, S. and de Vreese, C.H. (2020), "In AI we trust? Perceptions about automated decision-making by artificial intelligence", *AI & Society*, Vol. 35 No. 3, pp. 611-623.

Arkorful, V.E. (2022), "Unravelling electricity theft whistleblowing antecedents using the theory of planned behaviour and norm activation model", *Energy Policy*, Vol. 160, 112680, pp. 1-12.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020), "Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, Vol. 58, pp. 82-115, doi: 10.1016/j.inffus.2019.12.012.

Asante, K. (2023), "Hotels' green leadership and employee pro-environmental behaviour, the role of value congruence and moral consciousness: evidence from symmetrical and asymmetrical approaches", *Journal of Sustainable Tourism*, Vol. 32 No. 7, pp. 1-23, doi: 10.1080/09669582.2023.22.

Asante, K. (2024), "To speak up or not to speak up, organisational and individual antecedents that undergird this behaviour in resource constrained region", *Journal of Advanced Nursing*, pp. 1-13, doi: 10.1111/jan.16446.

Ashforth, B.E. and Mael, F. (1989), "Social identity theory and the organisation", *Academy of Management Review*, Vol. 14 No. 1, pp. 20-39, doi: 10.5465/amr.1989.4278999.

Ayentimi, D., Burgess, J. and Brown, K. (2018), "HRM development in post-colonial societies: the challenges of advancing HRM practices in Ghana", *International Journal of Cross-Cultural Management*, Vol. 18 No. 2, pp. 125-147, doi: 10.1177/1470595818765863.

Barnes, S., Mattsson, J. and Sørensen, F. (2016), "Remembered experiences and revisit intentions: a longitudinal study of safari park visitors", *Tourism Management*, Vol. 57, pp. 286-294, doi: 10.1016/j.tourman.2016.06.014.

Barocas, S. and Selbst, A.D. (2016), "Big data's disparate impact", *California Law Review*, Vol. 3 No. 104, pp. 671-732.

Barsky, A. and Kaplan, S.A. (2007), "If you feel bad, it's unfair: a quantitative synthesis of affect and organisational justice perceptions", *Journal of Applied Psychology*, Vol. 92 No. 1, pp. 286-295, doi: 10.1037/0021-9010.92.1.286.

Bhattacharya, C. and Sen, S. (2003), "Consumer–company identification: a framework for understanding consumers' relationships with companies", *Journal of Marketing*, Vol. 67 No. 2, pp. 76-88, doi: 10.1509/jmkg.67.2.76.18609.

Bigler, R. and Liben, L. (2007), "Developmental intergroup theory: explaining and reducing children's social stereotyping and prejudice", *Current Directions in Psychological Science*, Vol. 16 No. 3, pp. 162-166, doi: 10.1111/j.1467-8721.2007.00496.x.

Binns, R. (2018), "What can political philosophy teach us about algorithmic fairness?", *IEEE Security and Privacy*, Vol. 16 No. 3, pp. 73-80, doi: 10.1109/MSPP.2018.2701147.

Boojihawon, D.K. and Ngoasong, Z.M. (2018), "Emerging digital business models in developing economies: the case of Cameroon", *Strategic Change*, Vol. 27 No. 2, pp. 129-137.

Brashear, T.G., Brooks, C.M. and Boles, J.S. (2004), "Distributive and procedural justice in a sales force context: scale development and validation", *Journal of Business Research*, Vol. 57 No. 1, pp. 86-93, doi: 10.1016/S0148-2963(02)00288-6.

Budhwar, P., Malik, A., De Silva, M.T. and Thevisuthan, P. (2022), "Artificial intelligence – challenges and opportunities for international HRM: a review and research agenda", *The International Journal of Human Resource Management*, Vol. 33 No. 6, pp. 1065-1097.

Cameron, K. and Quinn, R. (2005), *Diagnosing and Changing Organisational Culture: Based on the Competing Values Framework*, John Wiley and Sons, San Francisco, CA.

Chartered Institute of Human Resource Management (2024), "About the institute", Accra: CIHRM, available at: https://cihrmghana.org/

Chin, W., Thatcher, J., Wright, R.T. and Steel, D. (2013), *Controlling for Common Method Variance in PLS Analysis: the Measured Latent Marker Variable Approach*, Springer Proceedings in Mathematics and Statistics, New York, doi: 10.1007/978-1-4614-8283-3_16.

Chuah, S.H.W., Aw, E.C.X. and Yee, D. (2021), "Unveiling the complexity of consumers' intention to use service robots: an fsQCA approach", *Computers in Human Behavior*, Vol. 123, pp. 1-13, 106870, doi: 10.1016/j.chb.2021.106870.

Colquitt, J. and Rodell, J.B. (2015), "Measuring justice and fairness", *The Oxford Handbook of Justice in the Workplace. s.L.*, Oxford University Press, Walton and Oxford, pp. 187-202.

Dibben, P., Brewster, C., Brookes, M., Cunha, R., Webster, E. and Wood, G. (2016), "Institutional legacies and HRM: similarities and differences in HRM practices in Portugal and Mozambique", *The International Journal of Human Resource Management*, Vol. 28 No. 18, pp. 2519-2537.

Dietvorst, B., Simmons, J. and Massey, C. (2018), "Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them", *Management Science*, Vol. 64 No. 3, pp. 1155-1170, doi: 10.1287/mnsc.2016.2643.

Du, S. and Xie, C. (2021), "Paradoxes of artificial intelligence in consumer markets: ethical challenges and opportunities", *Journal of Business Research*, Vol. 129, pp. 961-974, doi: 10.1016/j.jbusres.2020.08.024.

Ebrahimi, S. and Hassanein, K. (2019), "Empowering users to detect data analytics discriminatory recommendations", *Proceedings of the 40th International Conference on Information Systems*, Munich, Germany.

Feurer, S., Baumbach, E. and Woodside, A. (2016), "Applying configurational theory to build a typology of ethnocentric consumers", *International Marketing Review*, Vol. 33 No. 3, pp. 351-375, doi: 10.1108/imr-03-2014-0075.

Fianko, A.O., Essuman, D., Boso, N. and Muntaka, A.S. (2023), "Customer integration and customer value: contingency roles of innovation capabilities and supply chain network complexity", *Supply Chain Management: An International Journal*, Vol. 28 No. 2, pp. 385-404, doi: 10.1108/scm-12-2020-0626.

Fiss, P.P. (2011), "Building better causal theories: a fussy set approach to typologies inorganisation research", *Academy of Management Journal*, Vol. 54 No. 2, pp. 393-420, doi: 10.5465/amj.2011.60263120.

Fulmore, J.A., Nimon, K. and Reio, T. (2024), "The role of organisational culture in the relationship between affective organisational commitment and unethical pro-organisational behaviour", *Journal of Managerial Psychology*, Vol. 39 No. 7, pp. 845-862, doi: 10.1108/JMP-11-2022-0581.

Green, B. and Chen, Y. (2019), "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. s.l", *FAT 2019 -Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, ACM, pp. 90-99, doi: 10.1145/3287560.3287563.

Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P. and Weller, A. (2018), "Human perceptions of fairness in algorithmic decision making: a case study of criminal risk prediction. s.l", *Proceedings of the 2018 World Wide Web conference*.

Haak-Saheem, W. and Festing, M. (2018), "Human resource management – a national business system perspective", *The International Journal of Human Resource Management*, Vol. 31 No. 14, pp. 1863-1890.

Hair, J.F.Jr, Howard, M.C. and Nitzl, C. (2020), "Assessing measurement model quality in PLS-SEM using confirmatory composite analysis", *Journal of Business Research*, Vol. 109, pp. 101-110, doi: 10.1016/j.jbusres.2019.11.069.

Han, Q., Zheng, B., Cristea, M., Agostini, M., Bélanger, J.J., Gützkow, B., Kreienkamp, J. and Leander, N.P. (2023), "Trust in government regarding COVID-19 and its associations with preventive health behaviour and prosocial behaviour during the pandemic: a cross-sectional and longitudinal study", *Psychological Medicine*, Vol. 53 No. 1, pp. 149-159, doi: 10.1017/S0033291721001306.

Harned, Z. and Wallach, H. (2022), "Stretching human laws to apply to machines: the dangers of a 'colorblind' computer", *Florida State University Law Review*, Vol. 47 No. 3, pp. 1-33, available at: https://ir.law.fsu.edu/lr/vol47/iss3/3

Henseler, J., Ringle, C. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115-135, doi: 10.1007/s11747-014-0403-8.

Hogg, M.A. and Terry, D.J. (2000), "Social identity and self-categorization processes in organizational contexts", *The Academy of Management Review*, Vol. 25 No. 1, pp. 121-140.

Holmes, P.P.E. and Howard, M.C. (2023), "The duplicitous effect of organisational identification: applying social identity theory to identify joint relations with workplace social courage and unethical pro-organisational behaviours", *The Journal of Positive Psychology*, Vol. 18 No. 5, pp. 784-797, doi: 10.1080/17439760.2022.2109199.

Jarrahi, M.H. (2018), "Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making", *Business Horizons*, Vol. 61 No. 4, pp. 577-586.

Juvan, E. and Dolnicar, S. (2014), "The attitude–behaviour gap in sustainable tourism", *Annals of Tourism Research*, Vol. 48, pp. 76-95, doi: 10.1016/j.annals.2014.05.012.

Khalil, A., Ahmed, S.G., Khattak, A.M. and Al-Qirim, N. (2020), "Investigating bias in facial analysis systems: a systematic review", *IEEE Access*, Vol. 8, pp. 130751-130761.

Kordzadeh, N. and Ghasemaghaei, M. (2022), "Algorithmic bias: review, synthesis, and future research directions", *European Journal of Information Systems*, Vol. 31 No. 3, pp. 388-409, doi: 10.1080/0960085X.2021.1927212.

Krishnakumar, A. (2019), "Assessing the fairness of AI recruitment systems" Master thesis, Delq: TU Delq.

Kshetri, N. (2020), "Artificial intelligence in developing countries", *IT Prof.*, Vol. 22 No. 4, pp. 63-68.

Kshetri, N. (2021), "Evolving uses of artificial intelligence in human resource management in emerging economies in the global south: some preliminary evidence", *MRR*, Vol. 44 No. 7, pp. 970-990.

Landon, A.C., Woosnam, K.M. and Boley, B.B. (2018), "Modeling the psychological antecedents to tourists' pro-sustainable behaviors: an application of the value-belief-norm model", *Journal of Sustainable Tourism*, Vol. 26 No. 6, pp. 957-972.

Lee, M.K. (2018), "Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management", *Big Data & Society*, Vol. 5 No. 1, pp. 1-16, doi: 10.1177/2053951718756684.

Lee, M.K., Jain, A., Cha, H.J., Ojha, S. and Kusbit, D. (2019), "Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. s.l", *Proceedings of the ACM on Human-Computer Interaction*, Vol. 3 CSCW, pp. 1-26, doi: 10.1145/3359284.

Loi, M., Heitz, C., Ferrario, A., Schmid, A. and Christen, M. (2019), "Towards an ethical code for data-based business", *2019 6th Swiss Conference on Data Science (SDS)*, Bern, Switzerland, pp. 6-12, doi: 10.1109/sds.2019.00-15.

Mael, F. and Ashforth, B.E. (1992), "Alumni and their alma mater: a partial test of the reformulated model of organisational identification", *Journal of Organisational Behavior*, Vol. 13 No. 2, pp. 103-123, doi: 10.1002/job.4030130202.

Malik, A., De Silva, M., Budhwar, P. and Srikanth, N. (2021), "Elevating talents' experience through innovative artificial intelligence-mediated knowledge sharing: evidence from an IT-multinational enterprise", *Journal of International Management*, Vol. 27 No. 4, 100871, doi: 10.1016/j.intman.2021.100871.

Martin, K. (2019), "Designing ethical algorithms", *MIS Quarterly Executive*, Vol. 18 No. 5, pp. 129-142, doi: 10.17705/2msqe.00012.

McFarlin, D.B. and Sweeney, P.P.D. (1992), "Distributive and procedural justice as predictors of satisfaction with personal and organisational outcomes", *Academy of Management Journal*, Vol. 35 No. 3, pp. 626-637, doi: 10.2307/256489.

Memon, M., Ting, H., Cheah, J.H., Thurasamy, R., Chuah, F. and Cham, T.H. (2020), "Sample size for survey research: review and recommendations", *Journal of Applied Structural Equation Modeling*, Vol. 4 No. 2, pp. 1-10, doi: 10.47263/jasem.4(2)01.

Meyer, D. (2018), "Amazon reportedly killed an AI recruitment system because it couldn't stop the tool from discriminating against women", *Fortune*, available at: https://fortune.com/2018/10/10/amazon-ai-recruitment-bias-women-sexist (accessed 10 October 2018).

Murry, A. (2017), *Fortune 500 CEOs See AI as a Big Challenge*, s.l., available at: https://fortune.com/2017/06/08/fortune-500-ceos-survey-ai/

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S. and De Cremer, D. (2024), "Fairness perceptions of artificial intelligence: a review and path forward", *International Journal of Human-Computer Interaction*, Vol. 40 No. 1, pp. 4-23, doi: 10.1080/10447318.2023.2210890.

Niehoff, B.P.P. and Moorman, R.H. (1993), "Justice as a mediator of the relationship between monitoring methods and organisational citizenship behaviour", *Academy of Management Journal*, Vol. 36 No. 3, pp. 527-556, doi: 10.5465/256591.

O'neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, s.l., Broadway Books, New York, NY.

Pappas, I. and Woodside, A. (2021), "Fussy-set qualitative comparative analysis (FsQCA): guidelines for research practice in information systems and marketing", *International Journal of Information Management*, Vol. 58, pp. 1-23, 102310.

Podsakoff, P.P., MacKenzie, S. and Podsakoff, N. (2012), "Sources of method bias in social science research and recommendations on how to control it", *Annual Review of Psychology*, Vol. 63 No. 1, pp. 539-569, doi: 10.1146/annurev-psych-120710-100452.

Prikshat, V., Malik, A. and Budhwar, P. (2023), "AI-augmented HRM: antecedents, assimilation, and multilevel consequences", *Human Resource Management Review*, Vol. 33, pp. 1-18, 100860, doi: 10.1016/j.hrmr.2021.100860.

Ragin, C. (2000), *Fussy-set Social Science*, University of Chicago Press, Chicago, IL.

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, doi: 10.1007/s11747-019-00710-5.

Rihoux, B. and Ragin, C. (2009), *Configurational Comparative Methods: QualitativeComparative Analysis (QCA) and Related Techniques*, Sage Publications, London.

Robert, L.P., Pierce, C., Marquis, L., Kim, S. and Alahmad, R. (2020), "Designing fair AI for managing employees in organisations: a review", *Human-Computer Interaction*, Vol. 35 Nos 5-6, pp. 545-575, doi: 10.1080/07370024.2020.1735391.

Savari, M., Abdeshahi, A., Gharechaee, H. and Nasrollahian, O. (2021), "Explaining farmers' response to water crisis through the theory of the norm activation model: evidence from Iran", *International Journal of Disaster Risk Reduction*, Vol. 60, pp. 1-10, 102284.

Saxena, N.A., Huang, K., De Filippis, E., Radanovic, G., Parkes, D.C. and Liu, Y. (2020), "How do fairness definitions fare? Testing public attitudes towards three algorithmic definitions of fairness in loan allocations", *Artificial Intelligence*, Vol. 283, 103238, pp. 1-15.

Schmitt, A., Grawe, A. and Woodside, A.G. (2017), "Illustrating the power of fsQCA in explaining paradoxical consumer environmental orientations", *Psychology and Marketing*, Vol. 34 No. 3, pp. 323-334, doi: 10.1002/mar.20991.

Schneider, C. and Wagemann, C. (2010), "Standards of good practice in qualitative comparative analysis (QCA) and fussy sets", *Comparative Sociology*, Vol. 9 No. 3, pp. 397-418, doi: 10.1163/156913210X12493538729793.

Schwartz, S.H. (1977), "Normative influence on altruism", in Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology*, Academic Press, New York, pp. 221-279.

Shulner-Tal, A., Kuflik, T. and Kliger, D. (2023), "Enhancing fairness perception – towards human-centred AI and personalised explanations understanding the factors influencing laypeople's fairness perceptions of algorithmic decisions", *International Journal of Human-Computer Interaction*, Vol. 39 No. 7, pp. 1455-1482, doi: 10.1080/10447318.2022.2095705.

Skarlicki, D.P.P. and Folger, R. (1997), "Retaliation in the workplace: the roles of distributive, procedural, and interactional justice", *Journal of Applied Psychology*, Vol. 82 No. 3, pp. 434-443, doi: 10.1037/0021-9010.82.3.434.

Someh, I., Davern, M., Breidbach, C.F. and Shanks, G. (2019), "Ethical issues in big data analytics: a stakeholder perspective", *Communications of the Association for Information Systems*, Vol. 44 No. 1, pp. 34-46, doi: 10.17705/1CAIS.04434.

Statista (2024), "Ghana: Share of economic sectors in the gross domestic product (GDP) from 2012 to 2022, s.l",available at: https://www.statista.com/statistics/447524/share-of-economic-sectors-in-the-gdp-in-ghana/ (accessed 6 May 2024).

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, doi: 10.1177/0008125619867910.

Tausch, N., Saguy, T. and Bryson, J. (2015), "How does intergroup contact affect social change? Its impact on collective action and individual mobility intentions among members of a disadvantaged group", *Journal of Social Issues*, Vol. 71 No. 3, pp. 536-553.

Thomas, R., Amico, R. and Kolbjørnsrud, V. (2016). "How artificial intelligence will redefine management". *Harvard Business Review*, pp. 1-15, available at: https://hbr.org/2016/11/how-artificial-intelligence-will-redefine-management

van Berkel, N., Sarsenbayeva, Z. and Goncalves, J. (2023), "The methodology of studying fairness perceptions in Artificial Intelligence: contrasting CHI and FAccT", *International Journal of Human-Computer Studies*, Vol. 170 C, 102954, doi: 10.1016/j.ijhc.

Varma, A., Dawkins, C. and Chaudhuri, K. (2023), "Artificial intelligence and people management: a critical assessment through the ethical lens", *Human Resource Management Review*, Vol. 33 No. 1, 100923, doi: 10.1016/j.hrmr.2022.100923.

Vomberg, A., Schauerte, N., Krakowski, S., Bogusz, C.I., Gijsenberg, M.J. and Bleier, A. (2023), "The cold-start problem in nascent AI strategy: kickstarting data network effects", *Journal of Business Research*, Vol. 168, 114236, pp. 1-10, doi: 10.1016/j.jbusres.2023.114236.

Wachter, S., Mittelstadt, B. and Russell, C. (2021), "Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI", *Computer Law and Security Review*, Vol. 41, 105567, pp. 1-33, doi: 10.1016/j.clsr.2021.105567.

Wang, S., Wang, J., Zhao, S. and Yang, S. (2019), "Information publicity and resident's waste separation behaviour: an empirical study based on the norm activation model", *Waste Manag*, Vol. 87, pp. 33-42, doi: 10.1016/j.wasman.2019.01.038.

Woodside, A. (2016), "The good practices manifesto: overcoming bad practices pervasive in current research in business", *Journal of Business Research*, Vol. 69 No. 2, pp. 365-381, doi: 10.1016/j.jbusres.2015.09.008.

Xie, H. and Feng, X. (2024), "Feeling stressed but in full flow? Leader mindfulness shapes subordinates' perseverative cognition and reaction", *Journal of Managerial Psychology*, Vol. 39 No. 3, pp. 323-351, doi: 10.1108/JMP-03-2022-0140.

Zhang, X., Liu, J. and Zhao, K. (2018), "Antecedents of citizens' environmental complaint intention in China: an empirical study based on norm activation model", *Resour. Conserv. Recycle.*, Vol. 134, pp. 121-128, doi: 10.1016/j.resconrec.2018.03.003.

Zimmerman, A., Janhonen, J. and Beer, E. (2023), "Human/AI relationships: challenges, downsides, and impacts on human/human relationships", *AI and Ethics*, pp. 1-13, doi: 10.1007/s43681-023-00348-8.

Appendix

Constructs
AR
Apart from AI developers, companies and policymakers, managers are responsible for the adverse effects of AI algorithms
Managers can do something to mitigate or even eliminate the negative impacts of AI algorithms
I am responsible for minimising the ethical dangers AI algorithms produce in the workplace
Managers have nothing to do with the arbitrariness in algorithmic processes and outcomes
I feel jointly responsible for AI algorithms' damaging consequences on people in the workplace
AC
AI bias has damning consequences at individual, organisational and societal levels
AI bias has terrible impacts on employee cohesion
AI algorithm bias worsens social justice
AI bias leads to reproduction and strengthens social prejudices
Excessive use of AI algorithm recommendations can lead to high employee inequalities
PN
Managers have the responsibility to reject AI recommendations when they appear biased
Managers have the responsibility to correct AI bias results with better alternatives
I must do something positive to combat the unintended consequences of AI bias
I am morally responsible for protecting employees
If I use fewer AI recommendations, I feel like a manager farmer
I feel morally obliged to use AI algorithm recommendations correctly and equitably
DJ
My use of AI algorithms for work roles and responsibilities is fair
I think that the extent of my dependence on AI for work decisions is fair
Overall, the outcomes produced for AI algorithms are quite fair
I feel that the decisions I make out of AI algorithms are fair
PJ
The standards set by management to supervise AI algorithms are enforced equally among all human resource managers
I treat all employees equally when applying AI algorithm recommendations
I follow different rules when using AI results to deal with different units or individuals in this organisation
I do not favour one employee over another, mainly when guiding my decision on AI outcomes
I consistently apply AI recommendations to all employees
I follow fair procedures when using AI algorithms to make a decision
OI
When someone criticises this group, it feels like a personal insult
I do not act like the typical person of this group (reversed)
I am very interested in what others think about this group
The limitations associated with this group also apply to me
When discussing this group, I usually say "we" rather than "they."
I have several qualities typical of members of this group
The group's successes are my successes
If a story in the media criticised this group, I would feel embarrassed
When someone praises this group, it feels like a good compliment
I act like a person in this group to a great extent
AI response
I am willing to help others who suffer from AI bias algorithm outcomes
I am willing to make alternative decisions to help others who suffer from AI algorithm bias outcomes
I am willing to make personal sacrifices to prevent the spread of AI bias consequences in the workplace

Source(s): Authors own work

Table A1.
Measuring scales

Filtering questions:

Does your organisation employ AI in people management issues such as recruitment and selection, performance appraisal, and training and development?

Are you required always to inform your decisions from algorithm suggestions?

What is the extent of AI use in HR management in your organisation?

Table A2. Industry composition of the study sample

Industries	Selected samples across industries	Per cent
Banking and Insurance	105	35
Media and telecommunication	85	28.3
Oil and gas (i.e. upstream and downstream)	43	14.3
Manufacturing	67	22.3

Source(s): Authors own work

Corresponding author
Kwadwo Asante can be contacted at: asante@utb.cz

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**