Journal of Global Information Management
Volume 29 • Issue 6 • November-December 2021

Addressing Algorithmic Bias in AI-Driven Customer Management

Shahriar Akter, University of Wollongong, Australia
https://orcid.org/0000-0002-2050-9985

Yogesh K. Dwivedi, Swansea University, UK & Symbiosis Institute of Business Management, Symbiosis International (Deemed), India

Kumar Biswas, University of Wollongong, Australia
https://orcid.org/0000-0003-1507-1824

Katina Michael, Arizona State University, USA

Ruwan J. Bandara, University of Wollongong, Australia
https://orcid.org/0000-0003-1983-6097

Shahriar Sajib, University of Technology Sydney, Australia

ABSTRACT

Research on AI has gained momentum in recent years. Many scholars and practitioners have been increasingly highlighting the dark sides of AI, particularly related to algorithm bias.. This study elucidates situations in which AI-enabled analytics systems make biased decisions against customers based on gender, race, religion, age, nationality, or socioeconomic status. Based on a systematic literature review, this research proposes two approaches (i.e., a priori and post-hoc) to overcome such biases in customer management. As part of a priori approach, the findings suggest scientific, application, stakeholder, and assurance consistencies. With regard to the post-hoc approach, the findings recommend six steps: bias identification, review of extant findings, selection of the right variables, responsible and ethical model development, data analysis, and action on insights. Overall, this study contributes to the ethical and responsible use of AI applications.

KEYWORDS

AI Ethics, Algorithm Bias, Artificial Intelligence, Machine Learning, Responsible AI

1. INTRODUCTION

The world is witnessing groundbreaking changes emerging from the application of artificial intelligence (AI). AI has revolutionized many sectors, including healthcare, education, retail, finance, insurance, and law enforcement and becoming increasingly adopted due to its ability to perform complex tasks which are comparable to humans. It is expected that companies will spend around $98 billion on AI in 2023 globally (International Data Corporation, 2019). This makes sense as AI solves critical business issues helping organizations to become more efficient, gaining competitive advantage while also saving on operational costs (Davenport & Ronanki, 2018; Oana, Cosmin, & Valentin, 2017; Rai, 2020). However, the use of AI is not without limitations.

DOI: 10.4018/JGIM.20211101.oa3

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

With the increasing popularity of automating and enhancing business processes with AI, many scholars and practitioners have voiced their concerns regarding the dark sides of AI. Especially concerns over fairness and algorithm bias have increased (Wang, Harper, & Zhu, 2020). Algorithm bias occurs when AI produces systematically unfair outcomes that can arbitrarily put a particular individual or group at an advantage or disadvantage over another (Gupta & Krishnan, 2020; Sen, Dasgupta, & Gupta, 2020). This is an outcome occurring mainly from working with unrepresentative datasets or issues in algorithm design and particularly affects underrepresented minority groups (Gupta & Krishnan, 2020; Mullainathan & Obermeyer, 2017; Obermeyer, Powers, Vogeli, & Mullainathan, 2019). Recently there were many cases that showcased gender, racial and socio-economic biases emanating from AI applications. Some of these include several facial recognition systems, for example, Amazon's AI-based "Rekognition" software, discriminating against darker-skinned individuals and also providing unreliable results in identifying females; Google's AI hate speech detector was found providing racially biased outcomes; Google was showing fewer ads to females compared to males in the recruitment of high paying jobs; Amazon also abandoned an algorithmic human resources recruitment system for reviewing and ranking applicants' resumes since it was biased against women; a racial bias in a medical algorithm developed by Optum was found to favor white patients over sicker black patients; and the robodebt scheme in Australia wrongly and unlawfully pursued hundreds of thousands of welfare clients for the debt they did not owe (Blier, 2019; Hunter, 2020; Johnson, 2019; Martin, 2019).

The impact of algorithm bias can be devastating, asymmetric and oppressive, with individuals discriminated against and businesses negatively impacted. Despite the increasing understanding of algorithm bias and its effects, overall research in this stream lacks a systematic discussion of how it can affect service systems and how we can address algorithm-bias in data-driven decision making. Therefore, this paper responds to the question: 'how to address algorithm bias in AI-driven customer management?' The main objectives of the current study are: 1) to review and analyze the algorithm bias in customer management; 2) to synthesize the systematic literature review findings into a decision-making framework, and 3) to provide future research directions as per the knowledge gap. The systematic literature review in the emerging topic of algorithm bias contributes to AI literature mainly by providing a clear picture of the determinants of algorithm bias and its effects on customer management. Also, this study uniquely contributes to the theory by presenting a theoretical framework that identifies four consistency measures and six post-hoc measures to address algorithm bias in customer management. Further, this study is important as it contributes to the debate of responsible innovation and ethical AI (Ghallab, 2019; Gupta and Krishnan, 2020; Rakova et al. 2020) by scrutinizing the key ethical challenge of algorithm bias in AI applications.

To achieve these goals, we have conducted a systematic review of the literature to synthesize and integrate the body of knowledge of the relevant high impact publications in the field (Palmatier, Houston, & Hulland, 2018). This type of review can identify real facts by critically evaluating and synthesizing the underlying knowledge in a robust, rigorous, transparent, and replicable way (Denyer & Tranfield, 2009; Littell, Corcoran, & Pillai, 2008; Vrontis & Christofi, 2019). As there is a lack of systematic review regarding this topical area, extending the knowledge through such a review process in this field is highly relevant.

The remainder of the study is structured as follows. The next section focuses on defining and conceptualizing AI and algorithm bias. The third section highlights the procedures of exploratory research methods explaining searching, synthesis and thematic analysis techniques. The fourth section develops a conceptual framework highlighting a priori and post-hoc mechanisms to deal with algorithm bias. Finally, we discuss the findings with theoretical and practical contributions and future research directions.

2. LITERATURE REVIEW

2.1 What Is AI?

AI primarily refers to the effort to develop computational technologies that mimic human reasoning, and decision making following the underlying mechanism of the human brain and nervous system guided by psychology and cognitive science (Kreutzer & Sirrenberg, 2020, Mehta & Hamke, 2019, Hassabis, Kumaran, Summerfield, and Botvinick, 2017). Mahmoud, Tehseen and Fuxman (2020) suggest that human intelligence encompass a wide array of approaches that can express logical, spatial and emotional cognition. Furthermore, human intelligence represents a learning ability based on experience, adaptability to new circumstances and has the potential to process abstract concepts with a capacity to apply knowledge to enact changes in the environment (Sternberg 2017). Although computers outperform humans in computational capabilities, the capacity of a machine is constrained and limited, considering human intelligence (Yao, Zhou, & Jia, 2018). Therefore, Mahmoud, Tehseen and Fuxman (2020) describe AI as computer or software intelligence where the software component consisting of a set of commands directs how the computer or the machine will act through electronic signals.

Several subclassifications of AI have emerged to distinguish the different capabilities of AI-enabled machines and also to avoid confusion regarding the general capability of AI. For example, the term Artificial General Intelligence (AGI) or 'Strong AI' refers to AI with human-level or higher intelligence. In contrast, the term Weak AI or Narrow AI refers to the embedded capacity of a machine to handle the specific task (Yao et al., 2018). Furthermore, the notion of machine learning, deep learning and hyper learning implemented by artificial neural network programming focuses on building capacity to simulate learning processes that are similar to the learning mechanisms of biological species, including humans (Akter et al. 2020). Reinforcement learning algorithms can also train themselves based on inputs received, learning via interaction and feedback without requiring hard-wired programming (Luca, Kleinberg and Mullainathan, 2019, Davenport and Ronanki, 2018; Flasinski, 2016; Kreutzer & Sirrenberg, 2020).

The definition of AI is essentially related to our understanding of intelligence. Intelligence is a long-debated concept which has been an enquiry in several disciplines within social science including psychology, philosophy, sociology etc. A practical definition of AI considering the context of business operations is warranted to assist managers and policymakers in determining the scope of AI across their organizational boundaries. Following a systems perspective, AI can be conceptualized as an enabler to foster new capabilities integrating emerging technologies and design paradigms (e.g., machine learning, big data analytics, etc.) to aid decisions, interactions, detections and recommendations (Ransbotham, 2018; Kaplan and Haenlein, 2019; Mckensy and Company, 2018; Davenport, 2018; Davenport, Guha, Grewal and Bressgott 2020; Rai, 2020). Overall, AI is perceived as a technological advancement with the potential to create a meaningful impact on business operations (Davenport, Abhijit, Grewal & Bressgott, 2019; Carmon, Schrift, Wertenbroch, and Yang, 2019; Daugherty, Wilson and Rumman, 2018).

2.2 Dark Side of AI

Business organizations are embracing the applications of AI for three critical business needs, including automating business processes, gaining insight through data analysis, and engaging with customers and employees (Davenport and Ronanki, 2018). The authors reveal that companies are now deploying algorithms using machine learning applications to identify patterns of customers' purchasing behaviour, detecting fraudulent transactions, analyzing warranty data to identify quality problems and provide insurers with more detailed actuarial modelling. Moreover, companies such as Vanguard has deployed AI-enabled cognitive agents to assist customer service employees to respond to frequently asked questions. However, a study reveals that to realize the usefulness of AI implementation, it is important to gain acceptance by consumers as consumers need to develop

confidence into the recommendations produced by AI as well as trust that the use of their personal information will be appropriate (Kaplan and Haenlein, 2019). Based on a study of US customers, Davenport (2018) finds that 41.5% of respondents said they did not trust AI-enabled services including home assistants, financial planning, medical diagnosis, and hiring, only 9% of trusted AI with their financials, and only 4% trusted AI in the employee hiring process. This may be as a result of the lack of user consultation in the development of AI as users perceive AI as a black box.

Managers recognize both the opportunities and risks of using AI (Ransbotham, Gerbert, Reeves, Kiron, and Spira, 2018). Iansiti and Lakhani (2020) highlight the examples of AI applications adopted by companies such as Didi, Grab, Lyft, and Uber to create predictions, insights, and choices through systematically analyzing internal and external data to guide and automate workflows. However, the automation may cause severe damage as evident in the accidents caused by the self-driving cars by UBER (Wakabayashi, 2018) or in the incident of deaths caused by the malfunctioned robot at an Amazon warehouse (Shah, 2018). Polli (2017) observes the incredible capacity of an algorithm for making data-driven decision-making predictions. Companies are increasingly relying on algorithms to make objective and comprehensive choices, however, and Polli (2017) notes while reliance on technology may avoid human bias, the potential to produce biased algorithms opens up a dark side of algorithm-based decisions. Table 1 summarises selected work on the dark side of AI.

2.3 Algorithm Bias and Its Effects On Customers Management

AI will substantially change both marketing strategies and customer behaviours (Davenport, Guha, Grewal & Bressgott 2019). The objective to deploy algorithm-driven AI is to reduce unconscious human bias – however, this may result in algorithmic bias. Therefore, bias within algorithms needs to be carefully evaluated, monitored and may be removed if deemed necessary (Polli, 2017). This research identifies that technology-driven platforms such as Humanyze and HireVue develop processes to remove bias from algorithms resulting in equal access to employment opportunities across demographically diverse applicants. Kaplan and Haenlein (2019) suggest enhancing customers' confidence and trust in AI applications to commensurate disclosure and explainability of the AI application's underlying rules, such as the production of decisions with superior explanation. In an aim to develop a guideline for AI adoption, the Personal Data Protection Commission of Singapore (2018) proposed that decisions of AI applications should be explainable, transparent and fair. The report recommended adopting corporate practices for monitoring automated algorithmic decisions to avoid unintentional discrimination and further warned that improper AI deployments will continue to erode existing consumer trust and confidence.

The potential algorithmic bias that is embedded within an AI application could originate from multiple causes including the data set that is used to train the neural network model (Davenport, Guha, Grewal & Bressgott, 2019, Villasenor 2019). For example, Weissman (2018) reported that Amazon abandoned an AI application for assisting the recruitment process due to discriminating behaviour towards women as it has been revealed that the bias emerged because of the training data used to train the neural network model containing predominantly previous male applicants. Additionally, AI-driven recommendation engines can reduce the perceived autonomy a customer may experience, in addition to the customer feeling that they are constantly under surveillance and being manipulated (Carmon, Schrift, Wertenbroch, and Yang, 2019).

With higher adoption of technology, customers are increasingly aware of releasing and sharing more personal information to obtain the desired products. However, maintaining trust becomes increasingly harder (Bandara, Fernando, and Akter, 2020a, 2020b) as most customers do not feel comfortable. Excessive purchase or browsing history related information gathering might lead to the potential for misuse or deception by a firm to gain a decisive strategic advantage (Bostrom and Yudkowsky, 2014; Mahmoud, Tehseen and Fuxman, 2020). For example, there is growing evidence of dark side Customer Relationship Management (CRM) practices (Frow, Payne, Wilkinson and Young, 2011, McGovern and Moon 2007). Frow et al. (2011) suggest that when service providers apply

powerful, intrusive technologies with a poor understanding of the strategic focus or unethical means or motives, it may result in inappropriate exploitation and abuse of customers. These practices involve distorting, manipulating or hindering the flow of information towards customers to purposefully constrain their decision making. This leads to customer dissatisfaction and the misuse of resources. By using CRM technologies, service providers often engage in a range of activities that extend beyond the ethical practice of the responsible use of technology.

In the ever-growing digital economy, electronic-CRM requires service providers to collect a vast amount of information, which can be misused or, used for purposes without receiving consent from customers or, sold to third parties or can be used for targeted marketing purposes (Bandara, Fernando, and Akter, 2019; Frow et al., 2011). Furthermore, complex pricing comparison algorithms can create alternatives that may create confusion and make it difficult for customers to make appropriate decisions (Frow et al., 2011) that may exploit a vulnerable group of customers including young or elderly (Sheth and Sisodia, 2006). CRM performance measurement systems and employee rewards may encourage buying behaviour without actual necessity. On the contrary, using the data within CRM, firms can promote discriminatory pricing strategies to allow services to a specific segment of customers while depriving others (Payne, Wilkinson and Young, 2011).

Thus, research is warranted to understand how service providers can avoid dark side behaviour to eliminate the dysfunctional economic, social and ethical consequences of such manipulative approaches using emerging digital technologies (Bandara, Fernando, and Akter, 2020b; Frow et al., 2011; R. Wang, Harper, & Zhu, 2020). Moreover, safeguarding customers from bias in AI applications within AI-driven business operations is an important research avenue (Carmon, Schrift, Wertenbroch, and Yang, 2019; Davenport, Guha, Grewal & Bressgott, 2019). Table 2 shows selected studies that focus on AI and algorithm bias.

3. METHODOLOGY

To develop the systematic literature review (SLR) process, we have followed established guidelines provided by Akter et al. (2019); Durach, Kembro, and Wieland (2017); Tranfield, Denyer, and Smart (2003); and Watson, Wilson, Smart, and Macdonald (2018). Based on these guidelines, first, we planned the searching protocols; second, we applied screening techniques with an extraction mechanism and finally, we synthesized and reported the themes of our research enquiry of algorithm bias.

3.1 Discovery

An original research question has driven our research process (Nguyen, de Leeuw, & Dullaert, 2018), which has been derived after careful exploration of various academic databases, newspapers, magazines and industry white papers. We followed the research question: "How to reduce algorithm bias in AI driven customer management?" Using the guidelines of Dada (2018) and C. L. Wang and Chugh (2014), we have addressed this research question, by exploring ScienceDirect, Emerald Insight, EBSCOhost Business Source Complete, and other relevant journals from cross-disciplinary areas. We applied the keywords as follows under systematic search ("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning") AND (algorithm bias* OR dark side*) AND ("customer ethics" OR "customer privacy") from 2000-2020 to capture a wide range of pertinent research from various fields. Our initial search has provided us with 3033 various papers (See Figure 1).

3.2 Screening And Inclusion

In this stage, we excluded a total of 2895 articles from the initial discovery of 3033 studies based on relevance, duplication check and quality. Using the procedures of Fosso Wamba, Akter, Edwards, Chopin, and Gnanzou (2015), Watson et al. (2018) and Pittaway, Robertson, Munir, Denyer, and Neely

Table 1. Selected studies on AI and its dark side

Study type	Study	What is AI	Main findings	Dark Side of AI
Conceptual	Frow, Paine, Wilkinson and Young, 2011	Perceive AI as an information management process.	Reveals three broad categories of dark side behaviour and identifies ten forms within each broad category considering the means and target of usage. Further demonstrates the linkage between key strategic CRM processes and different types of dark side behaviours.	The paper suggests that inadequate understanding of the CRM's strategic focus coupled with the application of intrusive technologies may result in service providers' exploitative dark side behaviour.
Conceptual	Luca, Kleinberg and Mullainathan, 2019 (Porter and Heppleman, 2019)	Define hyper learning, a branch of machine learning that allows systems to learn at machine speed has the capacity to develop novel solutions in specific settings, involving unsupervised learning and reinforcement learning algorithms.	Articulates three possible types of human-machine interaction: augmentation, true human-machine collaboration and hyperlearning. Suggests augmentation as the most popular application of AI that is used for business decision-making, retrieving relevant information; providing superior sales, financial and other forecasts etc.	The study suggests that the transformative potential of AI-based technologies are undermined due to considering human and machine interactions as exclusive to teams of human and machine or the augmentation of humans.
Empirical	Davenport and Ronanki, 2018	Point to the powerful hype surrounding the notion of artificial intelligence.	The study emphasizes on the cognitive technology-based AI. Further suggests that AI can facilitate three important business needs: automation of business processes, data analytics-based insights, and engagement with customers and employees.	The authors suggest that AI applications targeting a specific niche scope are generating superior outcomes over highly ambitious AI projects.
Conceptual	Ransbotham, 2018	Highlight the detection capacity of AI.	Findings suggest that predictions based on AI applications are useful for long term organizational goal and further confirms the considerable progress of business organizations in adopting AI-based prediction in different business operations.	Despite the significant potential, the application of AI based prediction is essentially difficult and may generate a low return on investment (ROI).
Report	Mckensy and Company, 2018	AI is defined as a range of capabilities of a machine to perform cognitive functions related to human minds such as reasoning, learning, problem-solving etc to develop effective solutions to business problems.	The study recommends that an organization's progress on transforming the core business components through digitization is a critical factor application of AI.	The study reveals that foremost challenges and barriers to AI adoption is an absence of a clear AI strategy, lack of appropriate talent, functional boundaries constraining end-to-end AI solutions, and the shortage of leadership ownership and commitment to AI.

continued on next page

Table 1. Continued

Study type	Study	What is AI	Main findings	Dark Side of AI
Empirical	Kaplan and Haenlein, 2019	Artificial intelligence (AI) is defined as a system level ability to appropriately interpret external data, to learn from this data, and to utilize thise data in learnings to accomplish specific tasks and goals through adapting in a flexible manner.	Illustrates the potential and risk of AI using a series of case studies of corporations, governments and universities. Further, the study presents an organizational level framework, the C Model of Confidence, Change, and Control to better manage internal and external implications of AI.	AI needs to be seen either by following the perspective of evolutionary stages of AI such as narrow intelligence, general intelligence and super intelligence or through focusing on different kinds of AI systems such as human-inspired AI, humanized AI and analytical AI.
Empirical	Ransbotham, Gerbert, Reeves, Kiron, and Spira, 2018	Generally accepted conception of AI	The study reveals that innovative organizations with a higher level of AI adoption are assigning higher priority on AI applications that are revenue generating over the cost reduction ones and the study finds that these companies are keen to scale their AI adoption across organization with increasing level of commitment.	The study notes that AI is creating both optimism and anxiety. The study further recommends that organizations can improve the overall understanding of artificial intelligence through having direct experience of working with AI tools and techniques on practical business problems or by recruiting employees having AI expertise.
Report	Gerbert, Ramachandaran, Mohr and Spira, 2018	Generally accepted conception of AI	A systematic and structured approach to realize the value of AI within the organizational context.	Risk and maturity of AI implementation needs to be carefully considered during AI integration and adoption.
Analysis	Davenport, 2018	Generally accepted conception of AI	Based on a study on US customers, this study reveals that trust on AI applications is very low among US customers.	The authors recommend the AI applications vendor must avoid overpromising, and encourage becoming more transparent and to consider third party certification.
Conceptual	Iansiti and Lakhani, 2020	Generally accepted conception of AI	The study finds that AI-driven applications can generate a higher number of users, higher level of engagement, and significant revenue growth if fits with market effectively.	The authors warn of the dangers of unconstrained growth of AI. They further recommend business leaders to become cautious and to explicitly consider the capacity of AI to inflict widespread harm.

continued on next page

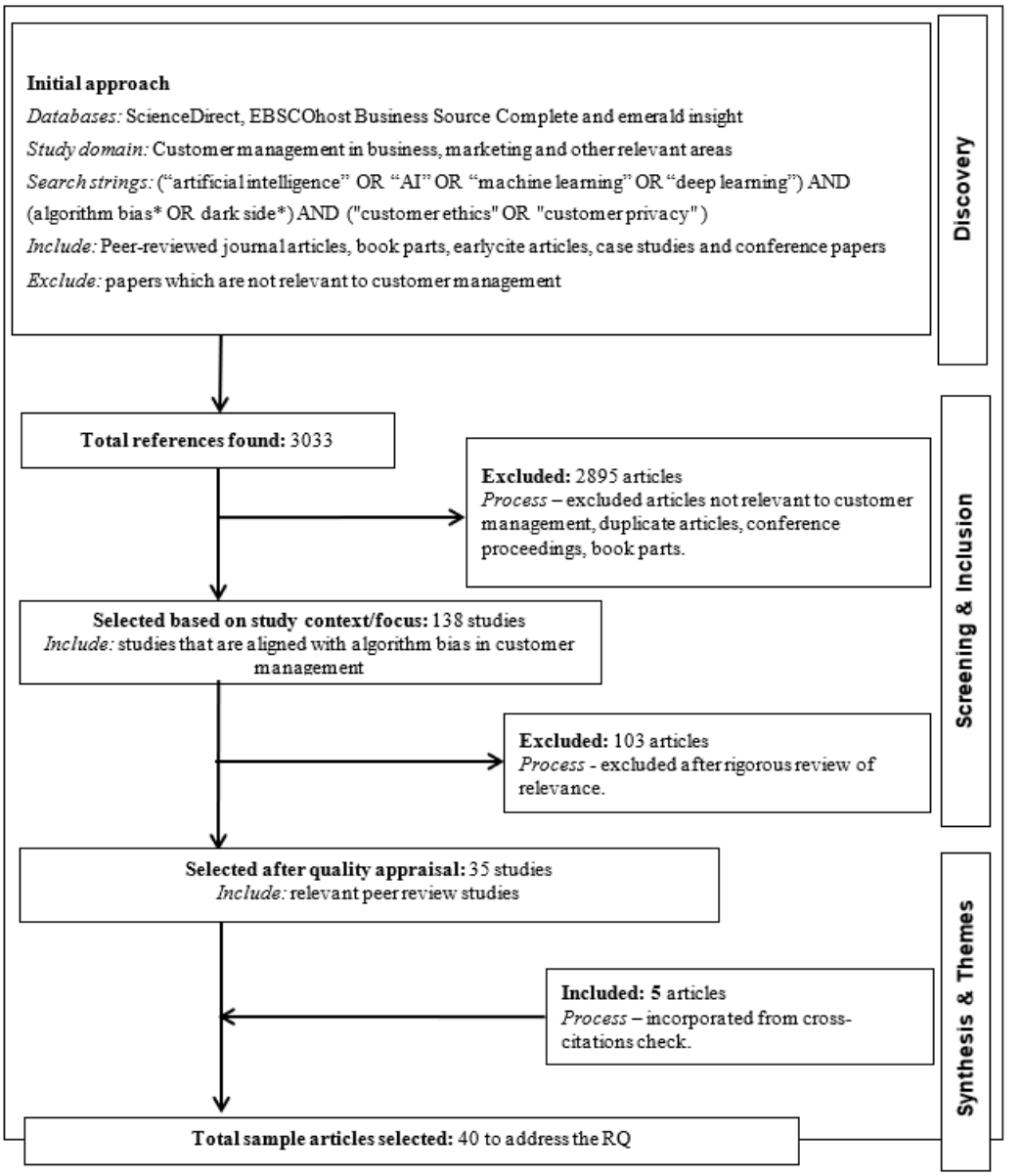
Table 1. Continued

Study type	Study	What is AI	Main findings	Dark Side of AI
Conceptual	Davenport, Guha, Grewal and Bressgott 2020	AI is conceptualized from the perspective of marketing and business applications for example automation of business processes, obtaining insights from data, engagement with the customers and stakeholders.	AI carries the potential for cost reduction and revenue generation. Revenue generation can be achieved through improved marketing decisions, and cost reduction can be achieved through task automation.	The algorithmic bias in AI applications may originate from the training data set, and due to the lack of transparency of algorithm design makes it difficult to identify the exact factors contributing to the algorithmic bias.
Conceptual	Rai, 2020	AI is perceived as technological innovation contemplated as systems, machines and applications. Highlight Explainable AI (XAI) as the class of the AI system that assists the users to understand the underlying mechanism of the decisions or predictions derived by the AI applications.	Technological innovations resulting in a transformative potential, as well as new identifiable risks which require to be understood and effectively managed to realize the benefits and safeguard the downsides of AI adoption.	Due to the inscrutable nature of the mechanism of many machine learning (ML) algorithms, specifically, the deep learning neural network approach causes a lack of trust in AI systems and may lead to the rejection of adoption. Algorithmic bias may result in vulnerability among the specific customer segment or community.
Conceptual	Rust, 2020	Artificial intelligence (AI) is conceptualized as computerized machineries to mimic capabilities that are unique to humans.	Artificial intelligence is playing a significant role in revolutionizing traditional marketing activities.	Marketing professionals have to deal with challenges such as socio-economic factors of diversity and inclusion and major geopolitical threats in adopting AI.

Table 2. Selected studies on algorithm bias

Study type	Study	Main findings	Algorithm Bias
Report	Personal Data Protection Commission Singapore, 2018	The study proposes policies and regulations promoting explainability, transparency, fairness, human-centricity as standard requirements to obtain consumer trust in the deployment of AI.	Risk of bias can be identified based on the inherent or latent authenticity or quality within a dataset. The study recommends organizations to adopt practices that may enable detecting biases within data to take effective steps to address appropriately.
Empirical	Davenport, Abhijit, Grewal & Bressgott, 2019	The study proposes a multidimensional framework to identify whether AI is embedded in a robot and obtains insights on the effects of AI considering intelligence levels and the nature of tasks.	The study suggests the sources or causes of the potential algorithmic bias embedded in AI applications. The study states that the lack of transparency makes it difficult to isolate and identify the exact factors that are under consideration by these algorithms.
Conceptual	Carmon, Schrift, Wertenbroch, and Yang, 2019	Findings suggest that tracking completed purchases carries a higher degree of fairness for consumers over ambiguous online monitoring.	The study highlights the reliance of AI applications on data and notes that the learning capability of automation technologies are causing discomfort among customers and result in concerns about the method of usage of the private data collected by AI-based automated technologies.
Empirical	Wissing and Reinhard, 2018	Examines the individual level differences in the perception of risk of AI and further studies the relationship between different forms of AI risk perception among non-experts and the Dark Triad personality traits.	This study reveals that individuals having self-reported knowledge of machine learning possess higher levels of AGI risk perception, associated with the Dark Triad traits.
Empirical	Vinuesa et al., 2010	The study finds that the understanding about the potential impact of AI on institutions is limited. The research confirms positive impact of AI algorithms in fraud detection, however, notes that algorithmic bias may hinder equality. The authors suggest developing policies and legislations regarding accountability and transparency of AI as well as ethical standards of the scope of AI applications.	The study suggests that the inherent bias in the training data possesses an underlying risk in applying AI in evaluation and prediction of human behaviour. The authors stress on the need to modify the data preparation process and warn about the exclusive adoption of AI-based applications to avoid such bias in areas such as recruitment.
Conceptual	Frow, Paine, Wilkinson and Young, 2011	The study offers understanding about the linkage between the dark side behaviours of service providers and CRM practices.	The authors state that the service providers may distort, manipulate or hinder information flow with poor timing or biased information which may affect decision making capacity of customers resulting in dissatisfaction and misuse of resources.
Empirical	Ransbotham, Kiron, Gerbert and Reeves, 2017	The study notes that data scarcity for training AI applications is a critical issue.	The study recommends to apply negative data along with the positive data set to overcome bias in the training data. Positive data refers to the data indicating intended results, whereas negative data refers to the data set containing failed outcomes.
Discussion paper	Chui et al., 2018	The study suggests that bias in data may result in concerns about privacy, fairness and equity as well as transparency and accountability in the use of extremely complex algorithms. The authors suggest business organizations and other users of data for AI to evolve business models in an ongoing base to address stakeholders' concerns related to data usage.	The study considers the risk of bias in data and algorithms as a limitation of AI. It further explains that the concerns related to bias are societal in nature and require implementing broader steps, including a deeper understanding of the process of collecting the training data.
Conceptual	Daugherty, Wilson and Rumman, 2018	The author recommends an inclusive approach in algorithm design through working with diverse groups to overcome the negative consequences of bias.	The author suggests inclusive training behaviour to the AI program taking into consideration the historical artefacts of bias contained within the training data set.

Figure 1. Protocol for systematic literature review

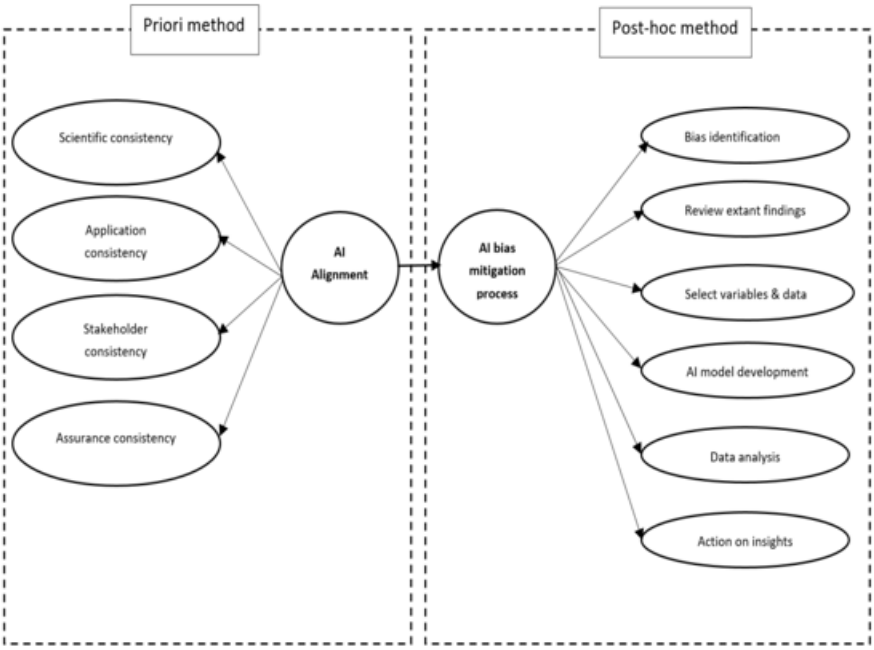


(2004), we excluded another 103 papers based on relevance check and quality appraisal. Finally, we studied 40 papers after a careful review for synthesizing our findings (see Figure 1).

3.2 Synthesis and Themes Identification

This section presents the findings of 40 articles included for thematic analysis and developing the conceptual framework to reduce algorithm bias in AI driven customer management. Following the procedures of Braun and Clarke (2006) and Akter, Bandara, et al. (2019), we examined 40 articles rigorously to identify potential themes. At this stage, we applied a coding method using a vital

Figure 2. A conceptual framework to address algorithmic bias.



analysis technique (Miles, Huberman, Huberman, & Huberman, 1994) to extract significant themes from the datasets (Tuckett, 2005). Finally, we have derived two codes in algorithm bias for customer management: a priori and post-hoc. The following section discusses the subdimensions of the two themes of algorithm bias.

4. FINDINGS

Based on a systematic literature review and thematic analysis, the study proposes two approaches/methods to mitigate algorithmic bias: ***a priori approach and post-hoc approach*** (see Figure 2). A priori approach includes four states of consistencies. The post-hoc approach encapsulates six steps to deal with algorithmic biases. First, a priori approach suggests AI alignment should be in place to overcome algorithmic biases by ensuring the four states of consistencies—scientific, application, stakeholder and assurance consistency. Second, the post-hoc approach recommends six steps to fix algorithm bias—identification of the algorithmic problem, review of extant findings and context, selecting relevant variables and collecting data, development of an ethical and responsible AI model by diverse teams, robust analysis of the training data and finally, act on insights and improve the model based on stakeholder feedback. We have suggested both a priori approach and post-hoc approach to mitigate algorithmic biases in the area of customer management. However, this can be equally applicable in other aspects in order to deal with algorithm biases.

4.1 A Priori Approach

According to Wixom, Someh and Gregory (2020), an adaptive management approach— articulated as an AI alignment— is a prerequisite to ensure a safe and large-scale AI deployment in any organization by orchestrating three overarching states of consistency—scientific, application and stakeholder

consistency. *Scientific consistency* produces a robust AI model capable of generating bias-free, accurate outcomes. To do so, an AI model needs to solve real-world problems by comparing the outcomes with a reality surrogate. If any gaps identified, that is addressed in line with the expectation of the real world. Modifications are brought in to refine labels, classes, variables in the training data and algorithms coded for machine learning. For example, General Electric (GE) ensured scientific consistency in its corporate environment, health and safety (EHS) standard for high-risk operations by developing and implementing an AI-enabled Contractor Document Assessment (CDA) application by 2020 (Jarrahi, 2018). This bolt-on AI-enabled application served all GE EHS professionals for use during the contractor onboarding process to free up time to divert their expertise to field execution and higher-value related EHS work (GE, 2020).

Application consistency creates a valid and reliable AI solution that delivers consistent outcomes over time to achieve the intended goals. To do so, it is important to fully understand the people, process and technology of a particular context and how the AI model is interacting with each sub-system. For example, the Australian Taxation Office (ATO), collecting more than $426 billion worth of net tax every year deployed the Smart Data program analytics in 2015 with a real-time nudging capability to support work-related expense claims (Body, 2008). Nearly 240,000 taxpayers received a pop-up message asking them to review their claim amount. Algorithms used to develop this pop-up message were based on past claims made by other taxpayers who are working in the same industry (Sydney Morning Herald, 2018). This AI-enabled real-time nudging prompted many taxpayers to adjust their work-related claims by around $113 million that benefited taxpayers to claim the right amount and saved time and resources of the ATO to assess the right taxable amount. *Stakeholder consistency* occurs when an AI solution offers a value proposition that is understood and applied by all stakeholders such as managers, frontline workers, and customers.

In addition, with the underpinning of technology adoption (Davis, 1989) and service quality research (Akter, Wamba, & D'Ambra, 2019), we propose that *assurance consistency* of the AI platform can enhance end-users' satisfaction by addressing security, privacy and ease of operation over time. Such assurance consistency is critical to keeping current users loyal to the AI platform and attracting new users through leveraging the power of word of mouth (Dwivedi et al., 2019; 2020). Even though an AI platform promises all consistency if end-users find the AI solution is complex and neither user friendly and nor trustable (Akter et al., 2019), it may prevent employees from using that AI platform. Lack of trust and user-friendliness can create excessive workloads for employees, thus deterring them from achieving their KPIs. For example, the 'Robodebt Scheme' employed by the Australian government in 2015 falsely accused welfare recipients of owing money to the government and issued automated false debt notices through a process of income averaging. This scheme received significant criticism from wider stakeholders such as media, scholars, advocacy groups and politicians (ABC News, 2020). As a result, the 'Robodebt Scheme' had been the subject of an independent investigation by the Commonwealth Ombudsman of the Australian Government and other legal bodies.

The Australian government revoked the robodebt recovery scheme for its gross algorithmic biases that created a disparate impact on welfare recipients and unbearable physical and mental trauma caused by the falsely computer-generated debt notices. The Australian Government also announced that it would repay in full 470,000 victims who received false debt notices, with an estimated A$721 million to be refunded. Very early on in the release of the robodebt scheme, advocacy groups called for evidence that the scheme actually did what it was meant to do, given that it was driven by AI and there was limited consultation with users, non-government organizations (NGOs) and other pertinent stakeholders (e.g. industry advisory). Prior to the robodebt roll-out, procedures were in place to ensure Centrelink was satisfied that debt had occurred before issuing a debt notice given that many citizens relied on their income for survival (Parliament of Australia, 2020). Among the victims of robodebt were significant numbers of vulnerable people, few of whom could not pay back any amount of money. In this context, it was less about trust by the Australian citizenry and more about evidence that the AI-driven scheme did what it was not meant to do from the outset. It became increasingly obvious

that robodebt not only did not work but was a debacle for the Australian government, sending the message to the Australian public that AI was not only functionally incompetent in its effectiveness but financially harmful. Little is known about how the program was developed, tested, and indeed whether it was piloted appropriately. As a public interest technology, robodebt was a large-scale failure. For many observers, the original Centrelink procedures worked, the impetus for the new system is unknown save for the allure of a technology that might reveal more. To realize the full advantage of safe AI deployment, it is important for management to establish alignment across the four states — scientific, application, stakeholders, and assurance consistency amidst dynamic internal and external forces. Given this, we posit:

Proposition 1: Consistency in the AI solution in terms of scientific, application, stakeholders and assurance can reduce algorithm biases.

4.2 Post-Hoc Approach

Following Akter et al. (2019), we propose the following six steps take into account to reactively address algorithmic biases in customer management. We define customer management as the holistic process of relationship management with both existing and new customers using data analytics. These steps can be equally applied in other AI-driven contexts to reduce algorithm biases.

4.2.1 Algorithmic Problem Recognition

Due to the emergence of machine learning and influx of voluminous big data, there has been a widespread reliance on algorithm-driven biased decision-making, for example, to perform mundane to complex decisions such as sorting applications for job-interviews, evaluating mortgage applications, and offering credit products. However, there is an array of evidence indicating that the use of biased algorithms can result in a disparate impact on a certain group in society due to differences in people's gender, race, colour, and socio-economic status. Such unfair algorithmic outcomes that arbitrarily prefer one group over another, whether done intentionally or unintentionally, can deprive vulnerable groups of basic human rights such as accessing loans, mortgages, getting a job, receiving health insurance cover and equal treatment in workplaces and community. Datta, Tschantz, and Datta (2015) found that, in 2014, Google's Ad settings webpage reportedly disadvantaged females over males. It was found that "setting the gender to female resulted in getting fewer instances of an ad related to high paying jobs than setting it to male" (p. 1). The Washington Post (2019) reported that bias-infecting algorithms generated and distributed by the leading US healthcare tech, Optum, favoured white patients over sick black patients in predicting which patients will most benefit from extra medical care. Consequently, as per the decision supplied by Optum, only 17.7% of black patients were eligible to receive additional care; however, correction of this AI bias would increase that figure to 46.5%.

The algorithmic problem leads to biased decision-making against a vulnerable group in society that warrants a thorough investigation of the algorithmic problem by defining and focusing on the specific business problem that a company is experiencing most. This enables the business to verify to what extent algorithms used to generate particular outcomes are unbiased (Appen, 2020). Focusing on the specific algorithmic problem helps draw out a road map depicting— who will do what, when and how (Davenport & Kim, 2013). This above example provides convincing justifications as to why it is important to critically identify the algorithmic problem at the very outset to minimize discriminatory outcomes because earlier detection can pave the way to designing a robust and rigorous AI model ensuring the survival and competitiveness of the business. First, problem identification at an earlier stage allows the business to ensure transparency and equity in all aspects of their business operations. Second, it protects the company from potential reputation damage by aggrieved customers, and monetary penalty imposed by regulators. Therefore, we propose the following proposition that reinforces real-time problem identification to reduce the likelihood of bias in consumer management.

Proposition 2: Real-time problem identification reduces the algorithm bias for customer management.

4.2.2 A Rigorous Review of Extant Findings and Context

Development of an ethical, responsible, and bias-free AI model starts with recognizing the algorithmic problem. However, without a thorough review of past and current biases, it is unlikely to navigate exact algorithmic problems. The extensive review indicates what sort of biases exist in current AI solutions being used by industries, governments and what sorts of study variables, labels, and algorithms are being used in the machine learning for decision-making (Davenport, 2014). For instance, algorithms used by Amazon's recruitment software for hiring senior managers was found biased towards males over females as it downgraded those resumes containing words such as 'women' and 'women's college' (Lavanchy, 2018). Gupta and Krishnan (2020) reviewed several AI-related biased outcomes and concluded that the majority of biases occur due to biased training data. As is the case for Amazon, in which Amazon's global workforce is 60 per cent males, and 74 per cent of them hold management roles. This distribution has been fed into training data, and the ML algorithm identifies that males are preferred candidates for Amazon's leadership roles. Scholarly review points out two major sources of algorithmic problems, which induce algorithm bias—biased training data (Sweeney 2013; O'Neil, 2016) and algorithm design (Obermeyer et al., 2019). Gupta and Krishnan (2020) contend that though algorithmic bias is the most popular term widely used, the identification of the real algorithmic problem is not lying with the 'algorithm' itself, rather it is in the actual data used to run the algorithms. They stress that 'algorithms are not biased, data is!' because algorithms learn from the attributes and persistent patterns in the training data. For example, Amazon's "Rekognition" facial recognition software led to AI bias because it falsely matched 28 US Congress members with a database of criminal mugshots. The study conducted by the American Civil Liberties Union found that "Nearly 40 per cent of Rekognition's false matches in our test were of people of colour, even though they make up only 20 per cent of Congress" (Lexalytics, 2019, p. 1). It signs flaws in the training data that can generate manipulative outcomes (Gupta and Krishnan, 2020, p. 1). This warrants the need to conduct an extensive review of the training data beforehand because the evaluation and detection of potential biases at an early stage can protect vulnerable groups from discrimination. Therefore, we posit:

Proposition 3: Rigorous review of past findings reduces algorithm bias.

4.2.3 Select Relevant Variables and Collect Data

After the identification of the algorithm problem and use of the review-findings, the next step is to select relevant variables and collect the most appropriate and valid data in order to develop an authentic and robust AI model ensuring equity and fairness in its applications. There are many instances where biased decisions are unintentionally made as the AI model favours a particular group of people over others, resulting in discriminatory treatments. Unwanted biases that an AI model generates is due to the extraction of flawed variables from the training data as well as a biased command within the model over which the end-user has no control. For example, Chowdhury (2018) reported that due to existing biases in the training data, many lenders in the US were granting loans to non-eligible white Americans while many eligible African Americans were ineligible to get mortgage applications approved. Scholars at Princeton University used off-the-shelf machine learning AI software to analyze 2.2 million words and found Anglo-Saxon names were perceived as more pleasant compared to those of African-Americans. They also explored that words such as "woman" and "girl" were less likely to be associated with science, mathematics (i.e., STEM subjects) rather than arts (Hadhazy, 2017). Therefore, prior to collecting data, it is important to know the data attributes, particularly in the age of big data where both structured and unstructured data are increasingly being considered together (Michael & Miller 2013).

Big data —classified as structured, semi-structured and unstructured— is derived from many sources such as social media (e.g., Facebook, LinkedIn), government agencies (e.g. Australian Bureau of Statistics), customer transactions (Amazon's online shopping order), click and video streams (Netflix), product reviews (Google review) and click and collect (e.g. Walmart). All these data have

been of great use for generating AI solutions; however, in many cases, the selection of wrong labels/variables lead to biased decisions. For example, Sweeney and Zang (2019) found that online search queries for African-American names more likely came up with a pop-up advertisement offering 'arrest records' and such arrest ads were significantly low when searched for white names. They also found that advertisements relating to higher interest-bearing credit cards and financial products were displayed on the screen once the system detected the subjects were from African-American backgrounds. It is important to have a solid understanding of different types of data and how they are coded and processed to run the AI model. Structured data are highly organized and easier for machine language to solve a particular problem. Structured data usually emanates from an organization's internal documents such as sales reports, customer purchases, transaction history, and view time. Semi-structured data is structured data but unorganized, embedded with some identifiable features, for example, BibTex files, CSV files, tab-delimited text files. Unstructured data is both ill-defined and unorganized such as blogs, wikis, images, graphs, audio, video, emails, streaming. To process and retrieve the meaning of this data requires advanced tools and software that AI algorithms have been leveraging more than any time in the past (Naik et al., 2008; Phllips-Wren et al., 2015). To address algorithmic problems, utmost attention and professionalism should be maintained while collecting reliable and valid data relevant to the selected variables that allows analysts to measure and test the AI model without the influence of confounding factors (Davenport, 2013; Janssen et al., 2017). Therefore, we posit:

Proposition 4: Systematic selection of relevant variables and collection of relevant data reduce algorithm bias.

4.2.4 Development of An Ethical and Responsible AI Model By Engaging Diverse Teams

Despite the plethora of availability of big data from multiple sources, realizing the full benefits from the authentic training data is contingent on the design of a robust and ethical AI model. Adequate precaution should be taken while processing variables, writing codes in the machine learning to make the model bias-free. Angwin et al. (2017) found that Facebook allowed advertisement purchasers to target "Jew-haters" as a category of users. Facebook later acknowledged the incident was an inadvertent outcome of algorithms used in assessing and categorizing data. Similarly, Facebook's use of flawed algorithms permitted ad buyers to block African-Americans from seeing housing ads. Therefore, it is critical to understanding data attributes, the engrained parameters, and machine languages used to develop an ethical and responsible AI model (Sivarajah et al., 2017). In 2010, Nikon received significant criticism because its S630 model digital camera displayed a warning message 'did someone blink?' while capturing images of people of Asian descent. Later, it was found that the use of flawed image-recognition algorithms contributed to this kind of unintentional bias that tarnished the brand reputation of Nikon.

Experts from world's leading AI technology company Appen, suggests that inclusion of diverse AI teams can challenge themselves in evaluating the AI model from different users' perspectives, which can lead to eliminating this kind of algorithmic problem before reaching out to end-users located across the world. Chowdhury (2018) points out that HR departments of many large organizations use AI for hiring and performance-evaluation to make promotional decisions but studies show that gender and race are highly correlated with salary, thus adversely influencing promotional decisions. To eliminate promotional biases, algorithm design should be orchestrated in a manner that excludes employees' race and gender while running the model to ensure meritocracy for leadership roles. The US Equal Credit Opportunity Act instituted in 1974 provides equal access to credit without discriminating people based on their race, colour, religion, national origin, sex, marital status, age, or because a person is receiving public assistance. Using this Act, anyone can challenge biased credit decisions generated by the faulty training data based on consumers' zip codes, socio-economic status, gender, and religion (Chowdhury, 2018). Therefore, the development of an ethical and responsible AI model should engage people from diverse socio-cultural settings to ensure that no one is disadvantaged with AI driven decision making. This leads to the following proposition:

Proposition 5: Development of an ethical and responsible AI model with diverse team members reduces algorithm bias.

4.2.5 Robust Analysis of The Training Data

Once the AI model is developed, the next step is to analyze the training data for testing whether the AI model is delivering critical insights in order to mitigate algorithmic biases. It is very critical to employ advanced analytical tools and techniques to explore underlying relationships between variables to gain meaningful insights (Davenport & Kim, 2013). Scholars have been favouring the use of complex analysis of models that allow for the mitigation of three kinds of biases: descriptive, predictive and prescriptive (Wang, Gunasekaran, Ngai, & Papadopoulos, 2016; Sivarajah et al., 2017).

Descriptive analytics use data aggregation and data mining processes to search out and summarise historical data in order to identify the change in patterns and relationships in the dataset and thereby provides useful insights into identifying a persistent problem and leveraging opportunities (Delen & Demirkan, 2013). Descriptive analytics of AI models help in navigating the problem by answering — 'what happened?' or 'what is happening now'. In the AI context, it could be useful to dig further, for example, monitoring changes in a firm's customer and employee diversity ratio over the last 12 months. This may trigger the next level of analysis - what might be the underlying reasons, which might have contributed to the given downward trend. The purpose of predictive analytics is to forecast what could happen in the future by employing complex algorithms. This answers 'what will happen' and 'why something will happen in the future' (Delen & Demirkan, 2013) if the current situation prevails. Take our previous example, if a company continues to lose a particular ethnic-related customer base over the last 12 months, how that could impact on the company's profitability, and stock price. Prescriptive analytic uses a large volume of data and takes hypothetical situations into account to generate a series of possible pathways to reach the desired outcomes (Watson, 2014). Findings generated by prescriptive modelling offer rich information context and expert opinions to optimize business decisions enhancing overall firm performance. For example, what course of action does a company need to attract more customers and retain them over the next 6-12 months? Besides these classifications, Sivarajah et al. (2017) gave an account of inquisitive analytics used to decide whether to accept or reject business propositions, whereas pre-emptive analytics take precautionary actions should any unexpected events occur to safeguard the business from undesirable influences. Diagnostic analytics originally built on descriptive analytics provide causal reasoning for relationships between variables that shed light on why things happened (Wedel & Kannan, 2016). In the scenario of algorithmic biases, both descriptive and diagnostic analytics are reactive in nature, whereas predictive and prescriptive analytics tend to optimize future decisions. Given this in mind, we propose the following proposition.

Proposition 6: Robust analysis of the training data with an ethical and responsible AI model reduces algorithm biases.

4.2.6 Act on Insights and Improve the Model Based on Stakeholders' Real-Time Feedback

According to Zhong et al. (2016), the main purpose of employing big data-driven complex AI models is to make solid decisions that safeguard the greater interest of diverse stakeholders. This necessitates the results generated by the AI model to be bias-free, reliable and acceptable by experts and end-users. There should be a concrete plan in place to act on insights gained from the feedback provided by end-users, AI experts and independent auditors. To leverage the full advantage of AI solutions, it is important to engage all employees, such as all levels of managers, frontline employees, customers, and suppliers using AI solutions (Wixom, Someh and Gregory, 2020). Employee engagement with AI and real-time communication with them is arguably one of the prime factors why the world's largest companies such as Amazon and Alphabet are benefitting from AI solutions, whereas the majority of other companies who fail on this are unable to have a positive return on investment using AI (Sam et

al. 2019). Furthermore, once feedback is received from key end-users, there should be a diverse data science team in place to address those identified biases both in the training data and algorithms in order to determine whether any modifications should be introduced in the training data and algorithms used to run the AI model. Therefore, we posit:

Proposition 7: Continuous feedback to improve the AI model and action on insights reduces algorithm bias.

Overall, there is a convincing consensus among scholars that the future source of competitive advantage of a firm is dependent on the extent to which it can safely and securely deploy bias-free AI solutions to deliver real-time decisions and solve critical business problems. To remain competitive globally, more companies are leveraging AI solutions which is estimated to reach $97.9 billion (IDC, 2019). Though the world's leading companies such as Google, Facebook and Amazon are leveraging AI benefits to excel their business performance; however, the majority of companies are unable to have a positive rate of return using AI (Sam et al., 2019). This warrants a call for the adoption of robust and ethical AI solutions for companies who are more concerned with sustained long-term profit maximization than short-term profits. To do so, we have suggested two approaches to be considered for a safe AI deployment. First, a priori method that suggests ensuring four states of consistency to be ensured in terms of scientific, application, and stakeholder (Wixom, Someh & Gregory, 2020) along with assurance consistency through adaptive and agile management. As part of a post-hoc method, we suggest six steps as noted above to be considered as a cycle of the continuous controlling process to mitigate algorithmic biases though it can be a challenging task given the inherent existence of deep-rooted social and institutional biases in many societies (Lexalytics, 2019). The General Data Protection Regulation (GDPR) legislation enacted in the EU Parliament on 25 May 2018 is a commendable step toward regulating data privacy and fair usage of private data advancing the adoption of ethical AI solutions. However, there is still a long way to go to protect customers and society from the dark effects of biased AI as many societies are prioritizing technological advancement over the humanistic and ethical aspect of AI. For instance, the *Financial Times* (2019) shares the concern that both China and the US are in favour of looser (or no) AI regulation for the sake of faster technological advancement over compromised and unethical treatment with vulnerable groups. Despite all these arguments, we suggest the ***a priori and post hoc approaches*** that can be a greater value addition to the existing literature of how to address algorithmic biases systematically; however, without an orchestrated global effort, humanity may not be able to eliminate algorithm biases to enjoy the complete advantages of AI solutions.

5. IMPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH

This study was motivated to advance knowledge by examining how organizations can deal with algorithm bias in their customer management efforts. The findings of this study have several implications for both theory and practice. First, the study systematically reviews literature pertinent to algorithm bias and presents key thematic areas relevant to the topic. This type of review enables a team to critically evaluate and synthesize the subject's underlying knowledge in a robust, rigorous, transparent, and replicable way (Denyer & Tranfield, 2009; Littell, Corcoran, & Pillai, 2008; Vrontis & Christofi, 2019). This is a significant contribution considering the importance and relevance of the topical area, and lack of such efforts in this field.

Second, the study proposes a conceptual framework that consists of both a priori and post-hoc measures for addressing algorithm bias. To the best of our knowledge, this is the first study to systematically integrate both a priori and post-hoc approaches to mitigate or overcome algorithm bias. We propose four consistency measures and six post-hoc measures, which can help businesses to deploy AI applications and solutions in an ethical and responsible manner and thereby improve customer management efforts (Michael et al. 2020).

Third, we contribute to the debate of responsible and ethical AI (Ghallab, 2019; Gupta and Krishnan, 2020; Rakova et al. 2020) by scrutinizing the key ethical challenge of algorithm bias in AI applications. Our motivation is to promote the ethical and responsible use of AI that mitigate or overcome discrimination, lack of fairness, and manipulation against certain social or institutional individuals and groups. We provide a theoretical basis to address algorithm bias and discuss potential causes as well as measures to overcome this challenge.

Fourth, our findings also further contribute to practice; we inform firms, AI scientists, and other practitioners to consider both a priori and post-hoc approaches to address algorithm bias. For organizations, we show that addressing ethical issues such as algorithm bias will ensure long-term benefits of AI investments over short-term gains. Businesses can integrate and apply the proposed framework in their customer management practices as well in other functions which involve AI such as recruitment.

Based on the review of the literature and thematic areas found from the analysis, we provide several avenues for future research (see Table 3). We identify that research on algorithm bias is only nascent, and therefore, the research agenda presented in this paper can immensely contribute to advance the research in this area. Especially, we highlight the necessity of research to dig deep into causes and determinants of algorithmic bias, and also further measures, apart from what we have identified to address those causes. Moreover, we call for extensive research in this area to address fairness, non-discrimination, non-manipulation, and trust in AI algorithms to deliver unbiased AI-driven outcomes. Further, we identify the need for taking an inclusive approach where different stakeholders are involved to ensure responsible and ethical deployment of AI applications that can bring sustainable growth to organizations.

Table 3. Future research directions from the review of extant literature

Future research area	Reference
Understand the impact and ways to address endogeneity bias, as AI-based approaches are very likely to exacerbate this issue.	De Bruyn et al. (2020)
Examine ways to transfer tacit knowledge from various marketing stakeholders to the AI algorithm and also from the AI algorithm back to the experts.	
Identifying different causes that can induce algorithm bias and testing for bias in AI application remain a non-trivial issue.	Davenport et al. (2020), Campbell et al. (2020), Conick (2017)
Identify different stages of the AI adaptation process and identify specific issues in each stage that may induce bias and discrimination for certain groups (e.g., data preparation stage, variable selection).	Vinuesa et al (2010); Ransbotham et al (2017);
Address individual and societal consequences emanating from biased training data and algorithm design for effective AI deployment.	Gupta and Krishnan (2020); Obermeyer et al. (2019)
Explore how fairness of AI systems can be established through 'Explainable AI' to prevent and detect algorithm bias in marketing applications.	Rai (2020); Feng et al. (2020); Grewal et al. (2019); Huang et al. (2020), Kumar et al. (2020), Ma & Sun (2020)
Examine the levels of explainability and transparency in AI systems to cater for the needs of different users.	
Develop automated decision-support capabilities which combine scale and insights.	Ma and Sun (2020), Rust (2020)
Explore how AI and related systems ensure the quality of life and well-being of consumers (e.g., ensure fairness and eliminate social biases, and safety-related concerns).	Kumar, Ramachandran and Kumar (2020)
Examine different effects on consumers such as discrimination, manipulation and loss of autonomy resulting from AI applications.	Carmon et al. (2019)
Build trust in all stages in AI life cycle for ensuring fair and non-discriminatory consumer outcomes.	Toreini et al. (2019)
Understand ways to balance between achieving organizational benefits of using AI and addressing dark sides of AI for gaining sustainable benefits.	Frow et al. (2011); Ransbotham (2018)
Given that the algorithm biases reflect certain social biases, research should design an inclusive approach to AI.	Chui et al, (2018); Daugherty, Wilson and Rumman (2018)
Investigate means of deploying safe and large-scale AI solutions using three interdependent states of consistency, namely, scientific consistency, application consistency, and stakeholder consistency.	Wixom, Someh and Gregory (2020)
Develop an AI-culture in organizations where employees at all levels and diverse stakeholders are engaged to ensure that AI applications are properly deployed (e.g., overcome social biases in AI algorithms).	Appen (2020); Wixom, Someh and Gregory (2020)
Address different ethical issues that are related and can augment algorithm biases such as the violation of consumer data privacy and security, intensive profiling, lack of transparency, and consumer autonomy and decision choices.	Tschider (2018); Qiu et al. (2019); Bandara, Fernando, and Akter (2019); Shams et al. (2020)

6. CONCLUSION

Although the growth of AI is unprecedented, the machine learning-based data analytics has resulted in situations in which many customers have been unfairly targeted due to algorithm bias. This is the dark side of AI that has been sporadically documented in the context of customer management. Both the digital giants (e.g., Facebook, Amazon, Google) and small specializing companies have applied either socially biased training data or algorithm design, which often reflect deep-rooted institutional discrimination or intolerance. The findings of the study propose two approaches (a priori and post-

hoc) to reduce algorithm bias in customer management. AI is often deployed with the company in mind, rather than customers. In large-scale government-driven AI deployments, the interaction with citizenry prior to the feasibility study is necessary to ensure that trust is maintained as users are the target of the AI rather than traditional "customers". It is important to make this distinction in the application of AI given the scale and the emphasis. What both private and public stakeholders must do is to consult more with end-users, and one another to ensure the most responsible and ethical AI is designed and implemented with rigorous testing and evidence for success. In this manner, businesses and government agencies established brands among their end-users that are positive in the adoption of new technologies.

REFERENCES

ABC News. (2020). *How the Robodebt settlement softens five years of pain for welfare recipients.* Retrieved from https://www.abc.net.au/news/2020-11-16/robodebt-settlement-explained/12888178

Akter, S., Bandara, R., Hani, U., Fosso Wamba, S., Foropon, C., & Papadopoulos, T. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. *International Journal of Information Management*, *48*, 85–95. doi:10.1016/j.ijinfomgt.2019.01.020

Akter, S., Michael, K., Uddin, M. R., McCarthy, G., & Rahman, M. (2020). Transforming business using digital innovations: The application of AI, blockchain, cloud and data analytics. *Annals of Operations Research*, 1–33. doi:10.1007/s10479-020-03620-w

Akter, S., Motamarri, S., Hani, U., Shams, S. M. R., Fernando, M., Babu, M. M., & Shen, K. N. (2020). Building dynamic service analytics capabilities for the digital marketplace. *Journal of Business Research*, *118*, 177–188. Advance online publication. doi:10.1016/j.jbusres.2020.06.016

Akter, S., Wamba, S. F., & D'Ambra, J. (2019). Enabling a transformative service system by modeling quality dynamics. *International Journal of Production Economics*, *207*, 210–226. doi:10.1016/j.ijpe.2016.08.025

Angwin, J., Varner, M., & Tobin, A. (2017). *Facebook Enabled Advertisers to Reach 'Jew Haters'*. ProPublica.

Appen. (2020). *How to Reduce Bias in AI*. Retrieved from https://appen.com/blog/how-to-reduce-bias-in-ai/

Bandara, R., Fernando, M., & Akter, S. (2019). Privacy concerns in E-commerce: A taxonomy and a future research agenda. *Electronic Markets*, *30*(3), 629–647. doi:10.1007/s12525-019-00375-6

Bandara, R., Fernando, M., & Akter, S. (2020a). Explicating the privacy paradox: A qualitative inquiry of online shopping consumers. *Journal of Retailing and Consumer Services*, *52*, 101947. doi:10.1016/j.jretconser.2019.101947

Bandara, R., Fernando, M., & Akter, S. (2020b). Managing consumer privacy concerns and defensive behaviours in the digital marketplace. *European Journal of Marketing*, *55*(1), 219–246. Advance online publication. doi:10.1108/EJM-06-2019-0515

Blier, N. (2019). *Bias in AI and machine learning: Sources and solutions*. Retrieved from https://www.lexalytics.com/lexablog/bias-in-ai-machine-learning

Body, J. (2008). Design in the Australian taxation office. *Design Issues*, *24*(1), 55–67. doi:10.1162/desi.2008.24.1.55

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. The Cambridge Handbook of Artificial Intelligence, 1, 316-334.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. doi:10.1191/1478088706qp063oa

Campbell, C., Sands, S., Ferraro, C., Tsao, H. Y. J., & Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. *Business Horizons*, *63*(2), 227–243. doi:10.1016/j.bushor.2019.12.002

Carmon, Z., Schrift, R., Wertenbroch, K., & Yang, H. (2019). Designing AI systems that customers won't hate. *MIT Sloan Management Review*.

Chowdhury R. (2018), Auditing Algorithms for Bias. *Harvard Business Review*.

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). *Notes from the AI frontier: Insights from hundreds of use cases*. McKinsey Global Institute.

Conick, H. (2017). The past, present and future of AI in marketing. *Marketing News*, *51*(1), 26–35.

Dada, O. (2018). A model of entrepreneurial autonomy in franchised outlets: A systematic review of the empirical evidence. *International Journal of Management Reviews*, *20*(2), 206–226. doi:10.1111/ijmr.12123

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *Proceedings on Privacy Enhancing Technologies, 2015*(1), 92-112.

Daugherty, P. R., Wilson, H. J., & Chowdhury, R. (2019). Using artificial intelligence to promote diversity. *MIT Sloan Management Review*, *60*(2), 1.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42. doi:10.1007/s11747-019-00696-0

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42. doi:10.1007/s11747-019-00696-0

Davenport, T. H. (2014). Keep up with your quants. *Harvard Business Review*, *91*(7/8).

Davenport, T. H. (2018). *Can We Solve AI's 'Trust Problem'?* Retrieved from https://alfredopassos.wordpress.com/tag/artificial-intelligence/

Davenport, T. H., & Kim, J. (2013). *Keeping up with the quants: Your guide to understanding and using analytics*. Harvard Business Review Press.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108–116.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, *13*(3), 319–340. doi:10.2307/249008

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K. U., & von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of Interactive Marketing*, *51*, 91–105. Advance online publication. doi:10.1016/j.intmar.2020.04.007

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*(1), 412–421. doi:10.1016/j.dss.2012.05.048

Denyer, D., & Tranfield, D. (2009). *Producing a systematic review*. Sage Publications Ltd.

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. *The Journal of Supply Chain Management*, *53*(4), 67–85. doi:10.1111/jscm.12145

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Galanos, V. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994. Advance online publication. doi:10.1016/j.ijinfomgt.2019.08.002

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., & Kumar, V. (2020). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, 102168. Advance online publication. doi:10.1016/j.ijinfomgt.2020.102168

Express, F. (2019). *EU backs AI regulation while China and US favour technology*. https://www.ft.com/content/4fd088a4-021b-11e9-bf0f-53b8511afd73

Feng, C. M., Park, A., Pitt, L., Kietzmann, J., & Northey, G. (2020). Artificial intelligence in marketing: A bibliographic perspective. *Australasian Marketing Journal*. Advance online publication. doi:10.1016/j.ausmj.2020.07.006

Flasiński, M. (2016). *Introduction to artificial intelligence*. Springer. doi:10.1007/978-3-319-40022-8

Folse, K. (2020). *What are the Other Types of Data Analytics?* Available at https://accent-technologies.com/2020/06/18/examples-of-prescriptive-analytics/

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246. doi:10.1016/j.ijpe.2014.12.031

Frow, P., Payne, A., Wilkinson, I. F., & Young, L. (2011). Customer management and CRM: Addressing the dark side. *Journal of Services Marketing*, *25*(2), 79–89. doi:10.1108/08876041111119804

GE. (2020), *Environment, Health, and Safety at GE*. Available at https://www.ge.com/sites/default/files/GEA20002_Environment,Health,and_Safety_at_GE_2020.pdf

Gerbert, P., Ramachandran, S., Mohr, J. H., & Spira, M. (2018). *The big leap toward AI at scale*. Academic Press.

Ghallab, M. (2019). Responsible AI: Requirements and challenges. *AI Perspectives*, *1*(1), 1–7. doi:10.1186/s42467-019-0003-z

Grewal, D., Hulland, J., Kopalle, P. K., & Karahanna, E. (2019). The future of technology and marketing: A multidisciplinary perspective. *Journal of the Academy of Marketing Science*, *48*(1), 1–8. doi:10.1007/s11747-019-00711-4

Gross, A., Murgia, M., & Yang, Y. (2019). Chinese tech groups shaping UN facial recognition standards. *Financial Times*. Retrieved from https://www.ft.com/content/c3555a3c-0d3e-11ea-b2d6-9bf4d1957a67

Gupta, D., & Krishnan, T. S. (2020). *Algorithmic bias: Why bother?* Retrieved from https://cmr.berkeley.edu/2020/11/algorithmic-bias/

Hadhazy, A. (2017). *Biased Bots: Artificial-Intelligence Systems Echo Human Prejudices*. Princeton University. Available at https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), 245–258. doi:10.1016/j.neuron.2017.06.011 PMID:28728020

Huang, M.H. & Rust, R.T. (2020). Engaged to a Robot? The Role of AI in Service. *Journal of Service Research*. 10.1177/1094670520902266

Hunter, F. (2020). *Human Rights Commission warns government over 'dangerous' use of AI*. Retrieved from https://www.smh.com.au/politics/federal/human-rights-commission-warns-government-over-dangerous-use-of-ai-20200813-p55lgn.html

Iansiti, M., & Lakhani, K. R. (2020). Competing in the Age of AI. *Harvard Business Review*, *98*(1), 60–67.

International Data Corporation. (2019). *Worldwide Spending on Artificial Intelligence Systems Will Be Nearly $98 Billion in 2023, According to New IDC Spending Guide*. Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS45481219

Janssen, M., Matheus, R., Longo, J., & Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. *Transforming Government: People, Process and Policy, 11*(1).

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. doi:10.1016/j.bushor.2018.03.007

Johnson, C. Y. (2019). *Racial bias in a medical algorithm favors white patients over sicker black patients*. Retrieved from https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. doi:10.1016/j.bushor.2018.08.004

Khadem, N. (2018). How the ATO is nudging Australians to pay more tax. *Sydney Morning Herald*. Retrieved from https://www.smh.com.au/money/tax/how-the-ato-is-nudging-australians-to-pay-more-tax-20180813-p4zx8x.html

Kreutzer, R. T., & Sirrenberg, M. (2020). What Is Artificial Intelligence and How to Exploit It? In *Understanding Artificial Intelligence* (pp. 1–57). Springer. doi:10.1007/978-3-030-25271-7_1

Kumar, V., Ramachandran, D., & Kumar, B. (2020). Influence of new-age technologies on marketing: A research agenda. *Journal of Business Research*. Advance online publication. doi:10.1016/j.jbusres.2020.01.007

Lavanchy, M. (2018). Amazon's sexist hiring algorithm could still be better than a human. *The Conversation*. Retrieved from https://theconversation.com/amazons-sexist-hiring-algorithm-could-still-be-better-than-a-human-105270

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press. doi:10.1093/acprof:oso/9780195326543.001.0001

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing*, *37*(3), 481–504. Advance online publication. doi:10.1016/j.ijresmar.2020.04.005

Mahmoud, A. B., Tehseen, S., & Fuxman, L. (2020). The Dark Side of Artificial Intelligence in Retail Innovation. In *Retail Futures*. Emerald Publishing Limited. doi:10.1108/978-1-83867-663-620201019

Martin, N. (2019). *Google's artificial intelligence hate speech detector Is 'racially biased,' study finds.* Retrieved from https://www.forbes.com/sites/nicolemartin1/2019/08/13/googles-artificial-intelligence-hate-speech-detector-is-racially-biased/?sh=2eefbdb326c4

McGovern, G., & Moon, Y. (2007). Companies and the customers who hate them. *Harvard Business Review*, *85*(6), 78. PMID:17580650

McKinsy and Company. (2018). *Global AI Survey: AI proves its worth, but few scale impact*. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Global%20AI%20Survey%20AI%20proves%20its%20worth%20but%20few%20scale%20impact/Global-AI-Survey-AI-proves-its-worth-but-few-scale-impact.pdf

Mehta, D., & Hamke, A. K. (2019). *In-depth: B2B eCommerce 2019*. eCommerce.

K. Michael, R. Abbas, G. Roussos, E. Scornavacca and S. Fosso-Wamba, "Ethics in AI and Autonomous System Applications Design," in IEEE Transactions on Technology and Society, vol. 1, no. 3, pp. 114-127, Sept. 2020, doi: .10.1109/TTS.2020.3019595

Michael, K., & Miller, K. W. (2013). Big Data: New Opportunities and New Challenges. *Computer, 46*(6), 22-24. https://ieeexplore.ieee.org/document/652725910.1109/MC.2013.196

Miles, M. B., Huberman, A. M., Huberman, M. A., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. doi:10.1016/j.artint.2018.07.007

Mullainathan, S., & Obermeyer, Z. (2017). Does machine learning automate moral hazard and error? *The American Economic Review*, *107*(5), 476–480. doi:10.1257/aer.p20171084 PMID:28781376

Mullainathan, S., & Obermeyer, Z. (2019). Who is tested for heart attack and who should be: Predicting patient risk and physician error (No. w26168). National Bureau of Economic Research.

Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2020). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. *Journal of Business Research*. Advance online publication. doi:10.1016/j.jbusres.2020.10.044

Naik, G. N., Gopalakrishnan, S., & Ganguli, R. (2008). Design optimization of composites using genetic algorithms and failure mechanism based failure criterion. *Composite Structures*, *83*(4), 354–367. doi:10.1016/j.compstruct.2007.05.005

Nguyen, D. H., de Leeuw, S., & Dullaert, W. E. (2018). Consumer behaviour and order fulfilment in online retailing: A systematic review. *International Journal of Management Reviews*, *20*(2), 255–276. doi:10.1111/ijmr.12129

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Oana, O., Cosmin, T., & Valentin, N. C. (2017). Artificial Intelligence-a new field of computer science which any business should consider. *Ovidius University Annals. Economic Sciences Series*, *17*(1), 356–360.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. doi:10.1126/science.aax2342 PMID:31649194

Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: Purpose, process, and structure. *Journal of the Academy of Marketing Science*, *46*(1), 1–5. doi:10.1007/s11747-017-0563-4

Parliament of Australia. (n.d.). *Chapter 2: Data Matching*. Senate Standing Committees on Community Affairs Design, Scope, Cost-Benefit Analysis, Contracts Awarded and Implementation Associated with the Better Management of the Social Welfare System Initiative. https://www.aph.gov.au/parliamentary_business/committees/senate/community_affairs/socialwelfaresystem/Report/c02

Personal Data Protection Commission Singapore. (2018). *Discussion paper on Artificial Intelligence (AI) and personal data – fostering responsible development and adoption of AI*. Retrieved from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf

Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. *Journal of Decision Systems*, *24*(1), 87–102. doi:10.1080/12460125.2015.994333

Pittaway, L., Robertson, M., Munir, K., Denyer, D., & Neely, A. (2004). Networking and innovation: A systematic review of the evidence. *International Journal of Management Reviews*, *5*(3-4), 137–168. doi:10.1111/j.1460-8545.2004.00101.x

Polli, F. (2017). *AI And Corporate Responsibility: Not Just For The Tech Giants*. Retrieved from https://www.forbes.com/sites/fridapolli/ 2017/11/08/ai-and-corporate-responsibility not-just-for-the-tech-giants/790804f95d4b

Qiu, M., Dai, H. N., Sangaiah, A. K., Liang, K., & Zheng, X. (2019). Guest Editorial: Special Section on Emerging Privacy and Security Issues Brought by Artificial Intelligence in Industrial Informatics. *IEEE Transactions on Industrial Informatics*, *16*(3), 2029–2030. doi:10.1109/TII.2019.2953884

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. doi:10.1007/s11747-019-00710-5

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2020). *Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices.* arXiv preprint arXiv:2006.12358.

Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018). *Artificial intelligence in business gets real*. MIT Sloan Management Review and Boston Consulting Group.

Ransbotham, S., Khodabondeh, S., Fehling, R., LaFountain, B., & Kiron, D. (2019). Winning with AI. *MIT Sloan Management Review*. Available at https://sloanreview.mit.edu/projects/winning-with-ai/

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT Sloan Management Review*, *59*(1).

Rose, A. (2010). Are Face-Detection Cameras Racist? *Time*.

Rust, R. T. (2020). The future of marketing. *International Journal of Research in Marketing*, *37*(1), 15–26. doi:10.1016/j.ijresmar.2019.08.002

Sen, S., Dasgupta, D., & Gupta, K. D. (2020). *An empirical study on algorithmic bias.* Paper presented at the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain.

Shah, S. (2018). *Amazon workers hospitalized after warehouse robot releases bear repellent*. Retrieved from https://www.engadget.com/2018-12-06-amazon-workers-hospitalized-robot.html

Shams, S. M. R., & Solima, L. (2019). Big data management: Implications of dynamic capabilities and data incubator. *Management Decision*, *57*(8), 2113–2123. doi:10.1108/MD-07-2018-0846

Sheth, J. N., Sisodia, R. S., & Barbulescu, A. (2006). The image of marketing. *Does Marketing Need Reform*, 26-36.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. doi:10.1016/j.jbusres.2016.08.001

Sternberg, R. J. (2017). Intelligence and competence in theory and practice. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 9–24). The Guilford Press.

Sweeney, L. (2013). Discrimination in online ad delivery. *Queue*, *11*(3), 10–29. doi:10.1145/2460276.2460278

Sweeney, L., & Zang, J. (2014). *How appropriate might big data analytics decisions be when placing ads?* Powerpoint presentation presented at the Big Data: A tool for inclusion or exclusion, Federal Trade Commission conference, Washington, DC. Available at https://www.ftc.gov/systems/files/documents/public_events/313371/bigdata-slides-sweeneyzang-9_15_14.pdf

Sydney Morning Herald. (2018). *How the ATO is nudging Australians to pay more tax*. Available at https://www.smh.com.au/money/tax/how-the-ato-is-nudging-australians-to-pay-more-tax-20180813-p4zx8x.html

The Washington Post. (2019). *Racial bias in a medical algorithm favors white patients over sicker black patients.* Available at https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2019). *The relationship between trust in AI and trustworthy machine learning technologies.* arXiv preprint arXiv:1912.00782.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222. doi:10.1111/1467-8551.00375

Tschider, C. A. (2018). Regulating the internet of things: Discrimination, privacy, and cybersecurity in the artificial intelligence age. *Denver Law Review*, *96*(1), 87–143.

Tuckett, A. G. (2005). Applying thematic analysis theory to practice: A researcher's experience. *Contemporary Nurse*, *19*(1-2), 75–87. doi:10.5172/conu.19.1-2.75 PMID:16167437

Villasenor, J. (2019). Artificial intelligence and bias: Four key challenges. Academic Press.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 1–10. doi:10.1038/s41467-019-14108-y PMID:31932590

Vrontis, D., & Christofi, M. (2019). R&D internationalization and innovation: A systematic review, integrative framework and future research directions. *Journal of Business Research*. Advance online publication. doi:10.1016/j.jbusres.2019.03.031

Wakabayashi, D. (2018). Self-driving Uber car kills pedestrian in Arizona, where robots roam. *The New York Times, 19*.

Wang, C. L., & Chugh, H. (2014). Entrepreneurial learning: Past research and future challenges. *International Journal of Management Reviews*, *16*(1), 24–61. doi:10.1111/ijmr.12007

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98–110. doi:10.1016/j.ijpe.2016.03.014

Wang, R., Harper, F. M., & Zhu, H. (2020). Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3313831.3376813

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*(1), 65. doi:10.17705/1CAIS.03465

Watson, R., Wilson, H. N., Smart, P., & Macdonald, E. K. (2018). Harnessing difference: A capability-based framework for stakeholder engagement in environmental innovation. *Journal of Product Innovation Management*, *35*(2), 254–279. doi:10.1111/jpim.12394

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97–121. doi:10.1509/jm.15.0413

Weissman. (2018). *Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women, Slate*. https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html

Wissing, B. G., & Reinhard, M. A. (2018). Individual differences in risk perception of artificial intelligence. *Swiss Journal of Psychology*, *77*(4), 149–157. doi:10.1024/1421-0185/a000214

Wixom, Someh, & Gregory. (2020). *AI Alignment: A New Management Paradigm*. Available at https://cisr.mit.edu/publication/2020_1101_AI-Alignment_WixomSomehGregory

Yao, M., Zhou, A., & Jia, M. (2018). *Applied artificial intelligence: A handbook for business leaders*. Topbots Inc.

Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*, *101*, 572–591. doi:10.1016/j.cie.2016.07.013

Shahriar Akter is an Associate Professor in the Sydney Business School at the University of Wollongong, Australia. Shahriar received his doctorate from University of New South Wales (UNSW) Business School. As part of his doctoral program, Shahriar received his methodological training from Oxford Internet Institute, University of Oxford. He has published in the leading international journals including Information & Management, Journal of Business Research, International Journal of Production Economics, International Journal of Production Research, International Journal of Operations and Production Management, Electronic Markets, Journal of the American Society for Information Science & Technology, Behaviour & IT, Communication of AIS, Journal of Selected Areas in Communication etc. Shahriar's research interests include service systems evaluation, big data and business analytics, and complex modelling using PLS.

Yogesh K. Dwivedi is a Professor of Digital Marketing and Innovation, Founding Director of the Emerging Markets Research Centre (EMaRC) and Co-Director of Research at the School of Management, Swansea University, Wales, UK. Professor Dwivedi is also currently leading the International Journal of Information Management as its Editor-in-Chief. His research interests are at the interface of Information Systems (IS) and Marketing, focusing on issues related to consumer adoption and diffusion of emerging digital innovations, digital government, and digital and social media marketing particularly in the context of emerging markets. Professor Dwivedi has published more than 300 articles in a range of leading academic journals and conferences that are widely cited (more than 21 thousand times as per Google Scholar). Professor Dwivedi is an Associate Editor of the Journal of Business Research, European Journal of Marketing, Government Information Quarterly and International Journal of Electronic Government Research, and Senior Editor of the Journal of Electronic Commerce Research. More information about Professor Dwivedi can be found at: https://www.swansea.ac.uk/staff/som/academic-staff/y.k.dwivedi/.

Kumar Biswas received his PhD in Management in 2013 from The University of Newcastle, Australia. Dr Biswas has been researching in the area of Big-Data, artificial intelligence and sustainability, gender diversity, firm performance and HR practices. Dr Biswas has published in the top tier and high impact journals such as International Journal of Human Resource Management Journal, Asia Pacific Journal of Human Resources, British Accounting Review, Journal of Strategic Marketing, and Evidence-based HRM: a global forum for empirical scholarship, Australian Journal of Environmental Education, Education + Training.

Katina Michael is a professor at Arizona State University, holding a joint appointment in the School for the Future of Innovation in Society and School of Computing, Informatics and Decisions Systems Engineering. She is also the director of the Society Policy Engineering Collective (SPEC) and the Founding Editor-in-Chief of the IEEE Transactions on Technology and Society. Katina is a senior member of the IEEE and a Public Interest Technology advocate who studies the social implications of technology. Katina has been funded by the Australian Research Council and the National Science Foundation.

Ruwan J. Bandara (PhD) is a Researcher at the Faculty of Business and Law in the University of Wollongong, Australia. His main research interests focus on responsible business and leadership, future of work, ethical issues of big data and new technologies, and impact of technology on employee / stakeholder wellbeing.

Shahriar Akter is a researcher and academic based in University of Technology Sydney. After completing PhD in dynamic capabilities, he has engaged researches that investigate innovative technologies including platform, additive manufacturing and artificial intelligence. Apart from scholarly journey, Dr. Shahriar is a CEO of Novorium, a Sydney based Digital Agency.