

South African Journal of Accounting Research

ISSN: 1029-1954 (Print) 2376-3981 (Online) Journal homepage: www.tandfonline.com/journals/rsar20

Building an ethical artificial intelligence corporate governance framework for the integration of emerging technologies into business processes

Husain Coovadia, Ben Marx, Ilse Botha & Nusirat O. Gold

To cite this article: Husain Coovadia, Ben Marx, Ilse Botha & Nusirat O. Gold (24 Jul 2025): Building an ethical artificial intelligence corporate governance framework for the integration of emerging technologies into business processes, South African Journal of Accounting Research, DOI: 10.1080/10291954.2025.2523661

To link to this article: https://doi.org/10.1080/10291954.2025.2523661

© 2025 The Author(s). Co-published by NISC Pty (Ltd) and Informa UK Limited, trading as Taylor & Francis Group

Published online: 24 Jul 2025.

Submit your article to this journal

Article views: 571

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rsar20

South African Journal of Accounting Research, 2025
https://doi.org/10.1080/10291954.2025.2523661

Routledge
Taylor & Francis Group
Check for updates

Building an ethical artificial intelligence corporate governance framework for the integration of emerging technologies into business processes

Husain Coovadia *

Department of Commercial Accounting, University of Johannesburg, Johannesburg, South Africa

Ben Marx

Department of Accountancy, University of Johannesburg, Johannesburg, South Africa

Ilse Botha

Department of Accountancy, University of Johannesburg, Johannesburg, South Africa

Nusirat O. Gold

Department of Commercial Accounting, University of Johannesburg, Johannesburg, South Africa; Department of Accounting and Finance, Kwara State University, Malete, Nigeria

(*Received 17 April 2024; accepted 17 June 2025*)

Purpose: To develop an ethical artificial intelligence (AI) corporate governance (CG) framework to guide South African business leaders in deploying and integrating AI into business processes, thus providing practical guidance to ensure responsible, transparent, and stakeholder-centric AI adoption.
Motivation: AI governance remains largely underdeveloped across Africa, particularly in South Africa, where businesses experience significant dilemmas in adopting and implementing an ethical AI framework. This study addresses that gap by developing a structured approach to ethical AI CG that supports responsible business practices.
Design/Methodology/Approach: A sequential mixed-methods approach was employed, combining a systematic literature review based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, with quantitative insights from a questionnaire.
Main findings: The study identified five essential elements for an effective ethical AI CG framework, namely transparency, machine bias, privacy, beneficial AI, and responsible AI, all of which must be stakeholder-centric.
Practical implications: A robust ethical CG framework tailored for South African business environments will encourage ethical AI adoption, strengthen adherence to the King IV Code, and enhance stakeholder trust while mitigating AI-driven risks inherent in technologies. The study emphasises continuous monitoring, stakeholder engagement, and compliance with legal frameworks like the Protection of Personal

*Corresponding author. Email: hcoovadia@uj.ac.za

© 2025 The Authors

South African Journal of Accounting Research is co-published by NISC Pty (Ltd) and Informa Limited (trading as Taylor & Francis Group)

This is 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 reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Information Act (POPIA). While the framework is tailored for South Africa, its principles can enjoy broader applications in other African business contexts.
Novelty/Contribution: This study developed a novel ethical AI CG framework for South African businesses using a sequential mixed-methods approach incorporating stakeholder views.

Keywords: AI; ethics; corporate governance framework; stakeholders; AI risk management

1. Introduction

The rapid advancement of Artificial intelligence (AI) technologies driven by increased data availability and computational advancements, is transforming industries**,** as businesses now leverage AI for efficiency, decision-making, innovation, competitiveness**,** and value creation (Enholm et al., 2022). Although AI adoption presents significant business opportunities, its widespread usage for daily activities has introduced unprecedented ethical issues for stakeholders (de Almeida et al., 2021). Specifically, concerns are being raised about ethical AI usage, transparency**,** and accountability (Bankins & Formosa, 2023; Turlapati et al., 2024), owing to its design, development**,** and deployment (Floridi, 2019). These concerns have resulted in a substantial gap between AI's potential and the ethical risks associated with its implementation (Martin et al., 2020).

In response to stakeholders' concerns, a growing debate has emerged globally on the ethical implications of AI principles and guidelines (Blanchard et al., 2024; Camilleri, 2024; Laine et al., 2024), and around the framework needed to ensure AI benefits for both humanity and the environment (Floridi, 2019). Although this awareness is promising, the focus needs to shift from "*what"* ethical AI entails to "*how"* it can be effectively implemented (Floridi, 2019). This is crucial as businesses struggle with integrating AI ethics into their business processes (Crawford et al., 2019), to foster adherence to corporate governance (CG) and ensure meaningful ethical oversight (Blanchard et al., 2024).

According to extant debate, the global North faces challenges with defining ethical AI CG, while the global South**,** including Africa, grapples with even greater challenges owing to lack of comprehensive AI CG frameworks (Jobin et al., 2019). As a result, African countries remain largely underrepresented in global AI CG discourse. This void presents a unique opportunity for South Africa to build its own AI CG framework that specifies *what* ethical governance should entail and offers practical guidance on *how* to operationalise it. This therefore underscores the need for a well-structured CG framework tailored for businesses deploying AI in their operations. This would enable businesses to operate responsibly, understand potential risks to stakeholders**,** and proactively address ethical concerns.

An ethical AI CG framework is an essential tool that emphasises real-time oversight, bias detection**,** and transparency, ensuring responsible AI development and use, while establishing ethical AI watchdogs and dynamic compliance algorithms to promote safe, secure, humane and sustainable AI development (Ejjami, 2024). Beyond risk mitigation, such a framework serves as an essential guide for corporate and legal operations (de Almeida et al., 2021), promoting accountability among employees, regulators and other stakeholders. Furthermore, a strong AI CG framework can help avoid bias, protect users' privacy**,** and reduce or mitigate environmental risks (KPMG, 2021).

The foregoing discussions provoked interest to develop an ethical AI CG framework for South African businesses deploying or planning to integrate AI into their operations. Consequently, the study develops an ethical AI CG framework to guide South African businesses in ensuring responsible AI technology adoption. This study is important,

given stakeholders' demand for a CG framework with clear guidance. Moreover, an ethical AI CG framework is critical to mitigating the risks of the black-box effect, where users and developers lack transparency around AI-generated outputs (Bankins & Formosa, 2023). More importantly, it helps guard against ethics "blue washing" (Floridi, 2019), where businesses project an image of responsibility without necessarily implementing meaningful action, ultimately undermining trust and slowing progress towards authentic ethical AI transformation. Therefore, integrating such a framework into business processes is essential for balancing innovation with ethical considerations and ensuring sustainable AI implementation.

To achieve the study's intention and answer the research question, the research was conducted in two phases. First, it identified prominent ethical AI issues highlighted by the literature. Secondly, a structured questionnaire was shared with business leaders to gather their insights, validate these issues, and determine necessary governance principles, ensuring alignment with the existing King IV Code on CG rather than contradicting it. The study concluded this phase with the development of an ethical AI CG framework, grounded in the practical insights of business leaders, thus enhancing its applicability and relevance.

The study contributes to knowledge by identifying five prominent ethical AI issues for businesses, namely transparency, machine bias, privacy, beneficial AI, and responsible AI. The innovative contribution lies in the development of a novel ethical AI CG framework specifically designed for South African businesses integrating AI technologies. The framework offers clear guidance on AI transparency, accountability, bias mitigation**,** and risk management. While tailored for South African businesses, the framework's principles can be adapted by businesses across Africa, allowing for contextual application. The study is timely because AI implementation in accounting and finance has been steering constant debate, particularly on issues pertaining to transparency, bias**,** and decision-making accountability. By intersecting AI CG with corporate accountability and ethical business practices, this research offers valuable insights to readers. It equips businesses to navigate the evolving legal landscape, align with existing CG codes, promote ethical AI usage, build stakeholder trust**,** and mitigate AI-related risks inherent in technologies.

The remainder of the paper is structured as follow: In Section 2, we present a literature review on AI in business processes, covering opportunities and risks. We also discuss why an ethical AI CG framework is vital, highlight prominent ethical AI concerns, and present the theoretical underpinning. Section 3 details the research methods, Section 4 presents the results and discussion of findings, and Section 5 concludes the study.

2. Literature review

AI has been defined in a variety of ways, often being associated with concepts such as systems that think and act like humans or operate rationally (de Almeida et al., 2021). Kalota (2024) describes AI as a computer system that has the capability to perform tasks that normally require human intelligence. Perifanis and Kitsios (2023), though, characterise AI as a broad range of leading-edge analytics, applications**,** and logic-based methods that mimic human behaviour, decision-making, learning**,** and problem-solving processes.

Similarly, the High-Level Expert Group of the European Commission (HLEG) has defined AI as systems that display intelligent behaviour by analysing their environment and taking actions autonomously to achieve specific goals (AI HLEG, 2019). This definition suggests that AI mechanisms must recognise patterns in their environment to create optimal solutions to problems. Collectively, these definitions highlight AI as man-made efforts that learn from interactions with humans, adapt to new challenges**,** and propose solutions to scenarios beyond previously encountered problems. The AI

HLEG (2019) noted further notes that AI systems can operate based on symbolic rules, learn numeric models**,** and adapt their behaviour to environmental feedback on their previous actions. Given the various definitions of AI and the challenges of establishing a universally applicable definition for regulation and governance, this study views AI as a rational system that perceives its environment through some sensors, collects and interprets data, processes the information, reasons the perceived data, then decides the best course of action and responds accordingly based on prompts. This definition builds on that of the HLEG.

2.1. *AI in business processes: Opportunities and risks, and the need for an ethical AI CG framework*

The fourth industrial revolution (4IR) is an era marked by transformative waves (Mohammed Shebeen et al., 2024). It is an era in which the distinction between the physical and digital realms blurs. This era sees smart manufacturing, the Internet of Things (IoT), AI and data analytics converging to reshape how businesses run (Mohammed Shebeen et al., 2024), while success in this era is dependent on a business's ability to use the digital tools to transcend traditional business paradigms and embrace adaptability as a core ethos (Mohammed Shebeen et al., 2024). Although technologies have been revolutionising business processes by enhancing efficiency, customer engagement and decision-making (Enholm et al., 2022), integrating them into business processes often introduces ethical dilemmas (de Almeida et al., 2021) such as lack of generalisability, automation bias, data security and privacy concerns, as well as autonomous decision-making.

Authors have stressed the lack of generalisability as a major concern with AI tools because they lack the ability to generalise results to many settings, groups and scenarios, hence limiting their applicability (Mathrani et al., 2021). Similarly, unlike humans, AI's ability is limited in applying learned experiences from one set of circumstances to another, needing businesses to continuously invest resources in retraining models, even for similar tasks (Chui et al., 2018). This rigidity poses a major challenge, as AI tools must be reconfigured on an ongoing basis to remain effective in evolving environments.

Furthermore, studies have raised issues about bias in AI. Bias can arise as systematic errors favouring certain outcomes (Mehrabi et al., 2021), or as unintended dependencies on specific attributes related to demographic groups (Fletcher et al., 2021), raising ethical and legal concerns. Roselli et al. (2019) categorised bias into three forms, namely bias in converting business intent into AI, bias from sample distribution in training data, and bias that is present in individual input samples. Crowe and Rodriguez (2024) have further identified computational, human, and systemic biases, arising from specific datasets and algorithms, cognitive biases in human thought, and those inherent in institutional structures, respectively. Since AIs are human-designed, their biases often reflect those of their developers (Matta et al., 2022), making humans ultimately responsible for mitigating AI-driven bias (Crowe & Rodriguez 2024).

Autonomous decision-making is another major concern in AI applications. Mittelstadt et al. (2016) argue that AI's complexity and adaptive nature make its decisions challenging for humans to interpret, leading to the prominent black-box concern (Hassija et al., 2024). This lack of transparency can deter organisational decision-making and complicate accountability for AI-driven harm (Leenes & Lucivero, 2014). In additional, data security and privacy pose significant risks in business (Alhitmi et al., 2024; Bao et al., 2022). AI's use of substantial quantities of consumer data for marketing increases the likelihood of breaches, unauthorised access and privacy violations, exposing businesses to legal and

reputational risks. These issues underscore the urgent need for an AI CG framework to ensure transparency, accountability, and ethical AI use (Crawford et al., 2019; Fessenko & Jasperse, 2024; Jobin et al., 2019; Laine et al., 2024).

Following the consensus on the need for ethical AI, we highlighted five (5) key ethical AI concerns comprising transparency, machine bias, data security and privacy, beneficial AI, and responsible AI that businesses must address in line with Jobin et al. (2019). The foundational study of Jobin et al. (2019) informed this decision, particularly because the research identified these principles as recurring themes across diverse geopolitical, institutional, and cultural contexts, reflecting a growing international consensus on the essential components of trustworthy AI, after they conducted a comprehensive analysis of 84 AI ethics guidelines from around the globe. These five elements were therefore chosen for their global relevance and for their particular significance in South Africa and other African nations, where AI's introduction is marked by challenges such as data governance, socio-economic inequality, and limited regulatory enforcement (Fokane, 2024; Ormond, 2023). Concerns like algorithmic transparency and bias mitigation are vital in preventing the reinforcement of historical injustices, while data security and privacy are central to protecting citizens' rights in regions with uneven digital infrastructure. Moreover, emphasising beneficial and responsible AI is crucial for ensuring that AI development supports inclusive growth and respects democratic values. Consequently, these elements serve as both ethical imperatives and practical guidelines for shaping AI governance in South Africa's evolving digital landscape. These five elements were employed to develop the Ethical AI CG framework for South African businesses. A brief explanation of the key elements is provided in Table 1.

2.2. *Linking stakeholder theory to ethical AI CG framework*

Stakeholder theory, propounded by Freeman (1984), emphasises that for long-term sustainability and ethical responsibility, businesses must consider the interests of all stakeholders, beyond just shareholders. Within the context of AI, this theory is applicable in that it provides a foundation for developing an ethical AI CG framework that aligns business practices with the diverse interests of all concerned parties, including employees, customers, regulators and the public at large. This theory advocates the inclusion of diverse views in decision-making processes, ensuring that AI systems are developed with a comprehensive understanding of their societal impact (Ademola, 2024).

AI-driven decisions impact numerous stakeholders, demanding transparency, fairness and accountability in governance mechanisms (Morante et al., 2024). Ethical AI CG framework must therefore incorporate stakeholder engagement to identify risks, address biases, and uphold privacy (Ejjami, 2024; Jobin et al., 2019; Morante et al., 2024). Prioritising stakeholder engagement in the development of the AI CG framework would ensure that technology aligns with societal values and ethical principles, fostering trust, minimising reputational risks and promoting compliance with evolving regulatory requirements (Bosse et al., 2023; Tan & Salo, 2023).

In additional, the CG gap between the global North and South (Jobin et al., 2019) further underscores the applicability of stakeholder theory. Given that Western bodies predominantly shape AI development and regulation, South Africa has an opportunity to design its own CG framework tailored to its local ethical concerns and socio-economic realities. Thus, stakeholder theory argues that AI governance should be context-specific, participatory, and aligned with diverse stakeholder needs to ensure responsible and beneficial AI deployment.

Table 1. Five key ethical AI concerns for businesses.

Ethical AI Concerns	Explanation	Sources
Transparency	To ensure transparency, AI systems must be understandable and explainable to stakeholders. To achieve this, businesses should disclose relevant information regarding AI operations, including decision-making processes, in a clear and easily interpretable manner to users to foster trust in the AI. Consequently, this will not just foster transparency but also ensure AI models' explainability, explicability, understandability, and interpretability, as well as disclosure of AI-generated outcomes.	(Bankins & Formosa, 2023; Jobin et al., 2019).
Machine Bias	AI can inherit biases from training datasets, resulting in unfair or discriminatory outcomes. To mitigate this concern, businesses must implement bias detection and mitigation strategies to ensure that AI models treat all stakeholders fairly. This will consequently ensure justice, fairness, inclusion, and non-discrimination to users.	(Fletcher et al., 2021; Mehrabi et al., 2021; Pasipamire & Muroyiwa, 2024)
Privacy	Since AI relies on large amounts of personal data, it often raises concerns about unauthorised access and data misuse. As a precaution, businesses must implement strong data protection and security measures and comply with privacy regulations and safeguard user information.	(Aledhari et al., 2018; Pasipamire & Muroyiwa, 2024)
Beneficial AI	For AI to be beneficial, it must serve human needs by enhancing productivity, decision-making, and societal well-being. An ethical AI CG framework should ensure that AI systems enhance business processes by contributing positively to the society through enhancing human dignity rather than imposing harm.	(Crawford et al., 2019; Raisch & Krakowski, 2020)
Responsible AI	Businesses must be accountable for the impact of AI, by ensuring compliance with ethical standards and regulations. Being responsible includes maintaining human oversight, risk assessments, continuous monitoring to prevent unintended consequences and detailing who is responsible for various actions of the AI system.	(Barredo Arrieta et al., 2020; Crawford et al., 2019; Jobin et al., 2019)

3. Research methods

The study utilised a sequential, exploratory mixed-methods approach, integrating qualitative analysis through PRISMA, with quantitative data gathered via a questionnaire, to develop a robust ethical AI CG framework. The questionnaire, focusing on business leaders in South Africa, was employed to gather business leaders' views of AI CG processes. The study conducted Cronbach's alpha (for internal consistency), Kaiser-Meyer-

Olkin (KMO) sampling adequacy, Bartlett's Test of Sphericity, and the Kolmogorov-Smirnov test for robustness checks, followed by a final factor analysis. This approach ensured an in-depth grasp of AI CG issues that led to the framework's development.

3.1. *Phase one: PRISMA Review*

The study conducted a systematic search using Google Scholar, a database with over 160 million scholarly articles and EBSCOhost (with African AI governance insights), covering the period 2000–2022. Our search began from the year 2000 owing to the "AI winter" experienced in the years prior to then. AI winter refers to a time of declining interest and investment in AI owing to the enormous disparity between what humans expected of these technologies and their actual capabilities (Hirsch-Kreinsen, 2024). During such periods, AI advancements failed in real-world applications, causing discouragement among investors, governments, and researchers (Floridi, 2020; Markelius et al., 2024). The Boolean search strings included the five key ethical AI terms highlighted previously. Titles and abstracts were screened to identify potential articles for inclusion and exclusion according to the set criteria. An initial identification yielded 334 papers, 262 of which were excluded for lack of relevance to ethical concerns, 25 were inaccessible, and only 47 could be retrieved from the various sources. This formed the articles that the research finally considered in building the CG constructs for the ethical AI concerns raised. A PRISMA diagram illustrates this selection process in Figure 1.

3.2. *Questionnaire (Phase two)*

3.2.1. *Development and design*

The authors designed the questionnaire to capture insights on transparency, machine bias, privacy, beneficial AI, and responsible AI. These constructs were formulated into questions and positioned on a Likert scale. This comprised:

i. A background section to assess the business leaders' AI knowledge.
ii. Ethical concerns divided into three sections, namely true/false questions, a five-point Likert scale, and open-ended questions.

We employed a mixed-methods approach to data analysis, combining both quantitative and qualitative techniques. The background section, which assessed business leaders' AI knowledge, was analysed employing descriptive statistics comprising frequencies and percentages to summarise self-reported familiarity with AI concepts. For the ethical concerns section, we divided the analysis according to question format. True/false responses were examined using frequency analysis to assess participants' factual understanding of ethical AI principles. We analysed responses from the five-point Likert scale by employing descriptive statistics (means, medians, and standard deviations) to gauge the level of agreement or concern regarding the Ethical AI issues. Open-ended responses were subjected to thematic analysis, allowing for the identification of recurring themes and insights that complemented the quantitative data. This approach ensured a comprehensive understanding of both the awareness and perceptions of ethical AI among business leaders.

Prior to administering the instrument, the authors sought and obtained ethical approval with clearance code SAREC20210610/01, ensuring compliance with the research ethics guidelines.

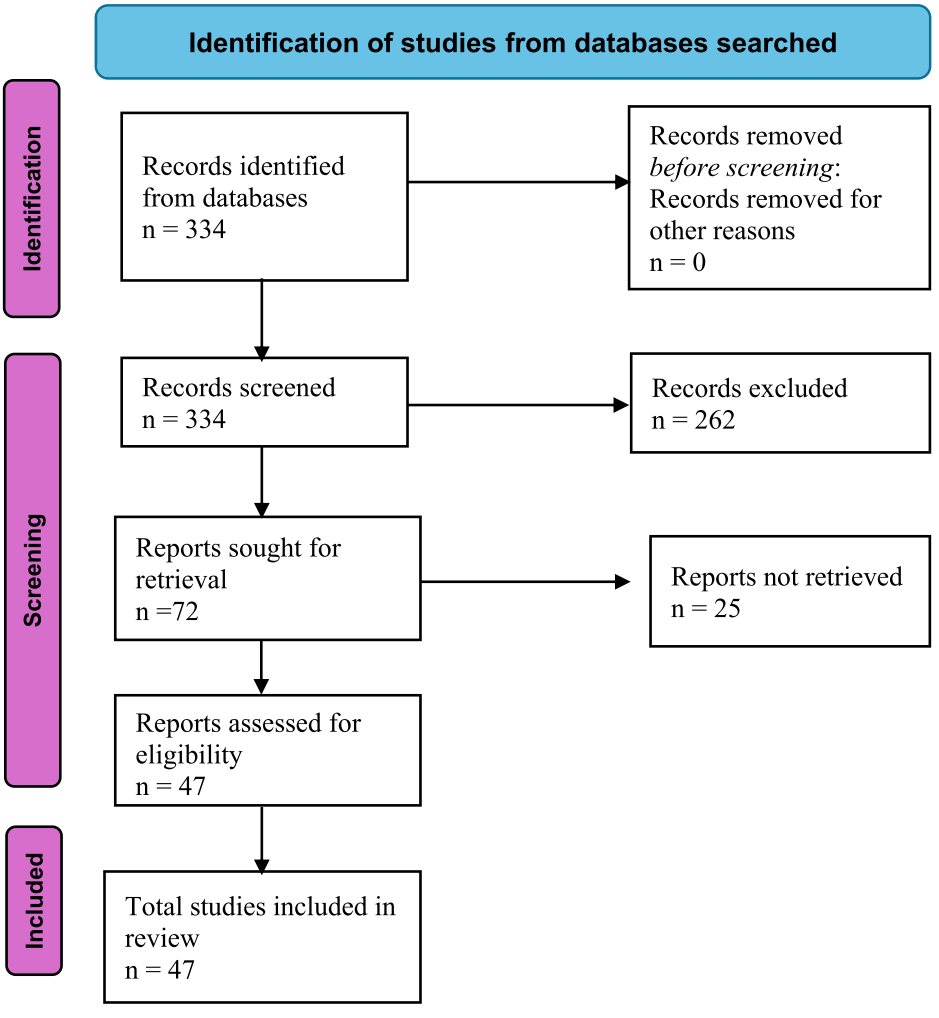


Figure 1. Flowchart on PRISMA approach.

3.2.2. *Pilot study*

A pilot study was undertaken with eight experts comprising audit partners, professors, and senior managers, who refined the questionnaire for clarity, particularly on transparency and machine bias. Additional context on instrument purpose, respondent qualification, and an anonymity statement were provided, with a 'not disclosed' option added for demographic questions like gender to ensure inclusivity.

3.2.3. *Population and sample*

Using a purposive judgemental sampling technique, the population targeted business leaders who were also Chartered Accountants (CAs) in South Africa. This group was selected owing to their strategic involvement in corporate decision-making, professional training in ethics and governance (South African Institute of Chartered Accountants (SAICA), 2022), and recognised influence across industries. CAs usually have advanced

expertise in business strategy, auditing, financial management, and taxation, putting them at the centre of business operations. Furthermore, they are constantly trained and developed professionally, under prominent bodies like the South African Institute of Chartered Accountants (SAICA) and the International Federation of Accountants (IFAC), which promote ethical behaviour, accountability, and good governance for all members. Moreover, these individuals frequently occupy senior leadership positions such as chief financial officers (CFOs), chief executive officers (CEOs), and board members. As a result, they are valued consultants, and their insights influence organisational policies and procedures. Their involvement is therefore critical for understanding how ethical AI CG can be effectively implemented in the business context.

The final sample comprised of 144 professional accountants across diverse business sectors trained in accounting, taxation, financial management, auditing, strategy, governance, ethics, and AI. Of these, only 111 responses were received, yielding a response rate of 77%.

4. Results and discussion

In this section, we present the results from our analysis and discuss the findings. We start first with discussion of findings from the PRISMA and then we proceeded to findings obtained from the survey instrument.

4.1. *Phase one: Findings from PRISMA*

The five prominent ethical concerns, as previously highlighted, based on the PRISMA approach, are transparency, machine bias, privacy, beneficial AI, and responsible AI which are also in line with the foundational study of Jobin et al. (2019). These constructs were expanded into structured processes essential for an ethical AI CG implementation. The core findings are as follows:

- to enhance transparency, businesses must focus on explainability, auditing, taking corrective measures, and managing technical limitation;
- to mitigate machine bias, businesses must identify bias types, implement mitigation mechanisms, and establish continuous monitoring and feedback strategies;
- to address privacy concerns, privacy-by-design principles, data control measures and bias transparency are needed;
- to ensure beneficial AI, businesses need to integrate social impact assessments, encourage inclusive decision-making through stakeholder participation, and promote positive contributions to the society, and
- to promote responsible AI deployment, businesses must ensure that technologies remain accountable and aligned with ethical standards to foster an environment that prioritises transparency.

The detailed governance processes which are based on the foundational study of Jobin et al. (2019) are summarised in Appendix A.

4.2. *Phase two: Analysis of survey instrument*

The findings from the demographic data showed that 109 respondents disclosed their gender, with 56% being female and 44% male. The mean age was 34 years (SD = 5.268), ranging

from 25 to 49 years, with a high majority of the participants being professionals working outside academia (64%); 49 held senior positions like financial managers, audit managers, managing directors, chief financial officers (CFOs) and senior lecturers.

4.2.1. *Robustness checks*

The robustness of the data was confirmed through the Cronbach's alpha, the KMO measure of sampling adequacy, and Bartlett's Test of Sphericity. All tests indicated robustness of data and suitability for factor analysis. The Cronbach's alpha, ranging from 0.923 to 0.971, confirmed high internal consistency; KMO values, ranging between 0.872 to 0.937, indicated sample adequacy, while Bartlett's Test ($p < 0.001$) confirmed the data's strong correlation. These checks ensured the questionnaire's reliability and validity (see Table 2).

4.2.2. *Normality testing*

The Kolmogorov-Smirnov (KS) test for normality was employed, but the test indicated non-normality ($p < 0.01$) for all items (see Table 3). However, since our sample size exceeded 30 ($N \geq 30$), the central limit theorem (CLT) was applied and this justified proceeding with factor analysis, as argued by del Barrio et al. (2020) and Uttley (2019), that the means of large samples tend to approximate normality as sample size increases.

Table 2. Cronbach's alpha, KMO, and Bartlett's Test of Sphericity.

		Transparency	Machine Bias	Privacy	Beneficial AI	Responsible AI
No. of Items		13	19	10	16	5
Cronbach's alpha		0.959	0.971	0.931	0.964	0.923
KMO		0.913	0.937	0.918	0.913	0.872
Barlett's Test	Approx. Chi-sq	1503.02	2269.53	1017.89	1017.89	1017.89
	df(sig)	78***	171***	45***	45***	45***

***significant level at 1%.
Source: Authors' computation from SPSS output.

Table 3. Kolmogorov-Smirnov.

Kolmogorov-Smirnov	Statistic	df	Sig.
Transparency: disclosure and monitoring	0.199	109	0.000
Transparency: independent observers	0.213	109	0.000
Machine bias	0.178	109	0.000
Factor 1: Privacy risk management	0.233	109	0.000
Factor 2: Privacy limiting data	0.140	109	0.000
Factor 1: Beneficial AI: human-centred	0.208	109	0.000
Factor 2: Beneficial AI: health, safety and rights	0.243	109	0.000
Responsible AI	0.214	109	0.000

Source: Authors' computation from SPSS output.

4.3. *Initial eigenvalues for possible factors in the data*

The initial eigenvalues presented in Table 4 help determine the number of underlying components that explain the variance in each key ethical AI concern. According to the Kaiser (1974) criterion, an eigenvalue > 1 is an indication that a component significantly contributes to explaining variance in the data, and such a component should be retained. Based on this criterion, this study considered one or more components. For instance, "Transparency", and "Privacy" both had two components with eigen values > 1, suggesting that these concerns are multifaceted and multiple aspects are required to fully understand their complexity. For transparency, monitoring and disclosure, and independent observers were key aspects deemed important while for privacy, data limitation and risk management were deemed important aspects. "Beneficial AI" also had two components comprising human-centred, and health and safety rights having eigen values > 1, indicating multiple dimensions are also required to fully capture its complexity. However, "Machine Bias" and "Responsible AI" had only one component each with an eigenvalue > 1, indicating that a single dominant factor sufficiently explains most of the variance in those areas. These interpretations validate the multidimensional nature of some ethical concerns while confirming the unidimensional importance of others, reinforcing their integration into the AI CG framework. Table 4 presents the pattern matrix.

4.4. *Summary of factor analysis*

The results of the factor analysis are presented in Table 5. Under transparency, two factors were loaded. First is "disclosure and monitoring", comprising items B1-6, 11, and 13, while the second is "independent observers", comprising items B7-9. Items B10 and B12 were removed to eliminate redundancy. "Machine Bias" and "Responsible AI" loaded a single factor, with machine bias explaining 67% of variance having all CG processes retained, while "Responsible AI" explained 77.17% variance, reinforcing its critical role in AI CG. "Privacy" loaded two factors from privacy risk management (D4-10) and privacy data limitation (D1-3), revealing strong internal consistency (Cronbach's alpha 0.963, 0.70). Similarly, two factors loaded for "Beneficial AI", comprising human-centred AI (factor items E2, E8-16; Cronbach's alpha: 0.950) and health, safety, and rights (E1, E3-7; Cronbach's alpha: 0.933) accounting for 73% variance. These results validate submissions of previous studies on the components of CG processes and ethical AI considerations (Barredo Arrieta et al., 2020; Costello, 2019).

4.5. *Discussion of overall findings*

The findings of our investigation reinforce the significance of the five key elements highlighted for developing an ethical AI CG framework (See Figure 2 for the framework). These are discussed in detail in the following sections.

4.5.1. *Transparency*

Our findings indicate that South African business leaders considered in this study emphasise business policies, disclosure, and monitoring as critical components that will enhance AI transparency, aligning with the information disclosure content for public listed companies, a requirement under the International Financial Reporting Standards (IFRS) and the Johannesburg Stock Exchange (JSE) practices. This builds on the study of Ananny and Crawford (2018).

Table 4. Pattern matrix.

Transparency				Machine Bias				Privacy				Beneficial AI				Responsible AI			
Com	Total	% of Variance	Cum %	Com	Total	% of Variance	Cum %	Com	Total	% of Variance	Cum %	Com	Total	% of Variance	Cum %	Com	Total	% of Variance	Cum %
1	8.794	67.646	67.646	1	12.807	67.407	67.407	1	6.694	66.942	66.942	1	10.524	65.776	65.776	1	3.859	77.171	77.171
2	1.159	8.913	76.559	2	0.926	4.876	72.283	2	1.2	12.001	78.943	2	1.293	8.084	73.86	2	0.456	9.126	86.296
3	0.713	5.488	82.047	3	0.776	4.085	76.368	3	0.53	5.303	84.247	3	0.807	5.042	78.902	3	0.31	6.2	92.496
4	0.623	4.794	86.842	4	0.709	3.729	80.097	4	0.393	3.925	88.172	4	0.625	3.905	82.807	4	0.229	4.584	97.08
5	0.333	2.56	89.401	5	0.67	3.527	83.624	5	0.328	3.283	91.456	5	0.488	3.048	85.856	5	0.146	2.92	100
6	0.268	2.061	91.463	6	0.482	2.535	86.158	6	0.271	2.708	94.164	6	0.405	2.533	88.388				
7	0.246	1.892	93.355	7	0.413	2.176	88.335	7	0.21	2.101	96.265	7	0.397	2.48	90.869				
8	0.217	1.672	95.027	8	0.37	1.947	90.282	8	0.164	1.644	97.909	8	0.287	1.795	92.663				
9	0.197	1.513	96.539	9	0.327	1.722	92.004	9	0.118	1.177	99.086	9	0.242	1.511	94.144				
10	0.158	1.216	97.755	10	0.27	1.421	93.425	10	0.091	0.914	100	10	0.202	1.262	95.437				
11	0.131	1.007	98.762	11	0.235	1.239	94.664					11	0.185	1.159	96.596				
12	0.117	0.903	99.665	12	0.232	1.223	95.887					12	0.149	0.93	97.525				
13	0.044	0.335	100	13	0.194	1.021	96.908					13	0.121	0.759	98.284				
				14	0.139	0.732	97.639					14	0.121	0.755	99.039				
				15	0.128	0.674	98.313					15	0.107	0.667	99.706				
				16	0.111	0.584	98.897					16	0.047	0.294	100				
				17	0.091	0.48	99.377												
				18	0.079	0.418	99.796												
				19	0.039	0.204	100												

Note: Com = Component; Cum % = Cumulative Percentage.
Source: Authors' computation from SPSS output.
The pattern matrix was used to identify governance processes for each factor. The highest loadings on each governance process were then determined per factor identified. Any governance processes with a loading greater than 0.4 were accepted (Sakamoto et al., 2006), while any with more than one factor were renamed.

Table 5. Factor analysis result.

Transparency			Machine Bias		Privacy			Beneficial AI			Responsible AI	
	Factor 1	Factor 2		Factor 1		Factor 1	Factor 2		Factor 1	Factor 2		Factor 1
B5	0.978	−0.14	C8	0.895	D6	0.951	−0.072	E16	0.976	0.149	F4	0.944
B4	0.972	−0.139	C7	0.873	D9	0.921	−0.022	E14	0.885	0.008	F2	0.861
B6	0.877	−0.013	C6	0.865	D8	0.91	−0.03	E13	0.841	−0.009	F3	0.837
B3	0.85	0.045	C5	0.852	D5	0.896	0.013	E15	0.795	0.078	F1	0.835
B13	0.805	0.139	C15	0.838	D4	0.858	0.052	E12	0.612	−0.313	F5	0.748
B11	0.723	0.226	C3	0.838	D10	0.84	0.032	E11	0.589	−0.334		
B1	0.701	0.097	C12	0.837	D7	0.806	0.106	E8	0.584	−0.348		
B2	0.596	0.282	C14	0.833	D1	−0.103	0.699	E10	0.529	−0.319		
B10	0.457	0.425	C4	0.831	D2	0.207	0.698	E9	0.503	−0.41		
B8	−0.061	1.008	C9	0.829	D3	0.286	0.548	E2	0.406	−0.401		
B7	0.055	0.887	C17	0.822				E5	−0.06	-0.967		
B9	0.039	0.876	C16	0.815				E4	−0.081	-0.916		
B12	0.408	0.529	C10	0.814				E6	0.035	-0.838		
			C13	0.802				E3	0.254	−0.644		
			C19	0.783				E7	0.414	−0.518		
			C11	0.768				E1	0.382	−0.459		
			C2	0.727								
			C1	0.697								
			C18	0.629								

Source: Authors' computation from SPSS output.
Factors removed from transparency-B10: maintain a comprehensive understanding of previous iterations of the AI system, including their functionality and the reasons for changes, to effectively audit the system; B12: disclose technical limitations of the AI system to stakeholders.

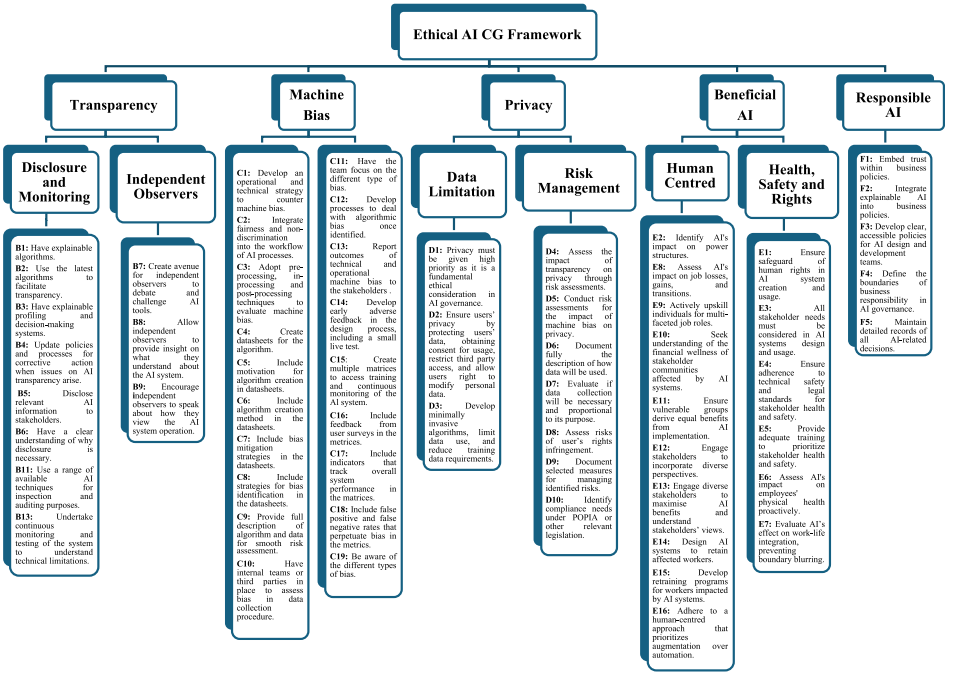


Figure 2. Ethical AI CG framework.
Source: Authors' compilation.

To achieve transparency in ethical AI for CG implementation, concerned business leaders need to develop sound policies for AI CG and ensure adequate monitoring and strict compliance with the policies. Likewise, there must be adherence to full disclosure of relevant information. However, on the role of independent observers in enhancing AI transparency, our findings contradict the previous arguments of Ananny and Crawford (2018) and Resnick et al. (2000). Probable reasons for this disparity include technical knowledge gaps, limited access to proprietary data (Hassija et al., 2024), AI complexity surrounding the black-box effect, organisational resistance, lack of standardised transparency measures, evolving AI technologies, weak regulatory backing, and resource constraints limiting the role of independent observers (Hassija et al., 2024). These issues might also be some of the factors contributing to the ongoing corporate scandals facing South Africa (Holtzblatt et al., 2020), such as the Steinhoff scandal, the African Bank Investments Ltd (ABIL) failure, and Experian data breaches.

Therefore, in line with stakeholder theory, it is important to adopt a stakeholder-centric approach that incorporates diverse perspectives, including independent observers, to enhance AI CG, identify ethical risks, and establish transparency norms (Ademola, 2024; Morante et al., 2024).

4.5.2. *Machine bias*

The findings support prior literature of D'Alessandro et al. (2017), Madaio et al. (2020) and Silberg and Manyika (2019), indicating that robust CG processes are essential in addressing machine bias. The results suggest that many business leaders considered in this

study may have limited exposure to, or understanding of, machine bias and how it can be effectively managed. Notably, most respondents preferred to address machine bias through stakeholder reporting rather than integrating fairness and non-discrimination mechanisms directly into AI systems.

This preference reflects a common trend in South African corporate practice, particularly the risk management approach, where businesses often report risks to stakeholders as part of good governance. Probably, they do this to comply with a section in the King IV Code that mandates the board to establish risk committees (Institute of Directors South Africa (IoDSA), 2022).

From the lens of stakeholder theory, an ethical AI CG framework must proactively incorporate mechanisms to detect oversight and machine bias in real time. This is vital for upholding ethical standards and building trust in AI systems. As Ejjami (2024) also argues, effective AI governance must prioritise continuous oversight and proactive bias detection to ensure fair and accountable outcomes.

4.5.3. *Privacy*

The findings indicate that business leaders prioritise risk management, including compliance with privacy legislation, process documentation, and risk assessments, probably owing to the influence of regulations like POPIA and the King IV Code, which mandates risk assessments as part of sound CG (IoDSA, 2022). Our findings indicate that data limitation is less emphasised (limiting data alone explains only 12% of the broader concept of privacy), likely because of concerns about AI's black-box effect. It is, however, important to understand that an ethical AI CG should balance privacy concerns with regulatory compliance.

Our findings lend support to the stakeholders' clamour for designing AI systems that address privacy concerns by aligning ethical expectations with regulatory requirements (Bosse et al., 2023). The findings aligns with Tan and Salo's (2023) emphasis on the need for AI tools to foster stakeholder value, enhance relationships, and uphold rights and responsibilities within the ecosystem.

4.5.4. *Beneficial AI*

The findings here suggest that for an ethical AI CG framework to be truly beneficial, it must prioritise the development of systems that deliver substantial benefits to all stakeholder groups (Ademola, 2024; Tan & Salo, 2023). For instance, a beneficial AI system should enhance business efficiency and support employee well-being, as well as contribute to broader societal goals such as environmental sustainability and equitable access. This requires adopting a human-centred, stakeholder-oriented approach, consistent with stakeholder proposition. This emphasises the need to account for the interests and well-being of all individuals affected by business decisions.

Furthermore, a human-centred AI CG model must evaluate AI technologies' impact on stakeholders' health, safety, and rights, ensuring that systems not only provide improved services but also uphold fairness, dignity, and user autonomy. This confirms the submission of Ademola (2024) that the diverse views of stakeholders must be considered in decision-making processes, as this will ensure the AI systems are developed with a comprehensive understanding of their societal impact. It also complements the submission of Tan and Salo (2023) that ethically deployed AI can foster sustainable stakeholder relationships, strengthen rights of ownership, and promote mutual

accountability within digital ecosystems. Ultimately, an AI system is deemed beneficial when it actively enhances human welfare and identifies and mitigates potential harms across stakeholder groups.

4.5.5. *Responsible AI*

The findings here reveal that South African business leaders strongly support embedding responsible AI within the CG framework through the development of clear policies and procedures that reflect a business's understanding of its responsibilities. This aligns with existing research such as Barredo Arrieta et al. (2020) that emphasises structured CG as fundamental to responsible AI. By incorporating these CG processes, the traceability principle that focuses on the ability to track decisions and actions within AI systems is addressed, reinforcing accountability across the AI development (Mittelstadt et al., 2016). This approach aligns with the stakeholder position, that calls for a nuanced understanding of the relationships among concerned groups. In addition, ensuring traceability and responsibility helps map these complex relationships, thus strengthening the ethical foundation of AI CG. When businesses adopt a stakeholder-centric mindset, they are more likely to engage diverse opinions in the design and implementation of AI systems, leading to more transparent, explainable, and unbiased AI solutions.

Furthermore, for an AI system to be truly responsible, it must be subject to proper risk management and continuous monitoring. This proactive approach not only mitigates potential harms but also supports ethical decision-making, strengthens accountability structures, and builds stakeholder confidence. Ultimately, embedding them in ensures that all governance processes relating to responsible AI are integral to the overall CG framework, promoting sustainable and trustworthy AI adoption.

5. Conclusions, contribution of the study, and limitations

This study developed an ethical AI CG framework tailored to South African businesses, employing a mixed-methods approach that combines qualitative data from PRISMA with quantitative data from a questionnaire to guide the adoption and implementation of responsible AI technology. The framework comprised five key elements essential for ethical AI CG, namely transparency, machine bias mitigation, privacy protection, beneficial AI, and responsible AI. Transparency was found to rely on clear policies, continuous monitoring, and disclosure. Independent observers' role was also captured, although respondents viewed it as less effective, probably because of technical limitations, lack of access to proprietary data, and organisational resistance.

For machine bias and privacy concerns, technical and operational strategies were emphasised, alongside inclusive stakeholder engagement and transparency in AI design. It is also essential for systems to be adaptable, ethical, and aligned with regulatory standards. For AI to be viewed as beneficial, it must prioritise stakeholder health, operational efficiency, and societal well-being. The authors emphasise compliance with existing frameworks like the King IV code and POPIA. By tackling "*how"* AI ethics can be implemented, the study offers practical solutions through ethical AI CG integration into business processes, mitigating the possibility of blue washing and fostering stakeholder trust.

This research makes several significant contributions to knowledge. By incorporating the views and recommendations of South African business leaders, the study offers unique insights into what constitutes ethical AI practices. The resulting ethical AI CG framework (Figure 2) addresses a critical gap and fulfils the overarching objective of developing a

practical and context-specific framework for South African businesses adopting AI technologies.

Theoretically, the study advances the discourse on ethical AI by anchoring its CG approach in stakeholder theory. It demonstrates how stakeholder interests can and should be integrated into governance processes to ensure ethical, inclusive, and accountable AI deployment. Practically, the framework complements South Africa's existing King IV Code on CG principles, but in addition provides actionable guidance that can be adapted by businesses across the continent.

Although this study developed a valuable ethical AI CG framework, it is not without limitations. First, context bound, the study focuses only on South Africa; therefore, caution must be exercised in applying its conclusions to other jurisdictions. In addition, several interpretive insights, especially those pertaining to machine bias, privacy, data limitations, and the perceived ineffectiveness of independent observers were based on reasoned assumptions of the authors rather than empirical validation from literature. Future research is urged to further investigate these findings to confirm or refine the explanations offered. The study employed the questionnaire, which may have limitations in accurately capturing nuanced opinions. Future studies are encouraged to use alternative methodologies, like interviews or case studies, to enrich future investigations. Finally, while factor analysis was used to interpret the quantitative data, this method has inherent limitations that may constrain the scope of interpretation. Despite these limitations, the findings contribute significantly to the existing body of knowledge and offer a foundation for further exploration of ethical AI CG.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID iD

Husain Coovadia http://orcid.org/0000-0003-3870-0072

References

Ademola, O. E. (2024). Detailing the Stakeholder Theory of Management in the AI World: A Position Paper on Ethical Decision-Making. *Advances in Multidisciplinary & Scientific Research Journal Publications*, *10*(1), 1–8. https://doi.org/10.22624/aims/sij/v10n1p1

AI HLEG - High-Level Expert Group on Artificial Intelligence. (2019). A definition of AI: Main capabilities and scientific disciplines. In *European Commission*. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Aledhari, M., Pierro, M. Di, Hefeida, M., & Saeed, F. (2018). A Deep Learning-Based Data Minimization Algorithm for Fast and Secure Transfer of Big Genomic Datasets. *IEEE Transactions on Big Data*, *7*(2), 271–284. https://doi.org/10.1109/tbdata.2018.2805687

Alhitmi, H. K., Mardiah, A., Al-Sulaiti, K. I., & Abbas, J. (2024). Data security and privacy concerns of AI-driven marketing in the context of economics and business field: an exploration into possible solutions. *Cogent Business and Management*, *11*(1), 2393743. https://doi.org/10.1080/23311975.2024.2393743

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media and Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bankins, S., & Formosa, P. (2023). The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. *Journal of Business Ethics*, *185*(4), 725–740. https://doi.org/10.1007/s10551-023-05339-7

Bao, Y., Li, W., Ye, Y., & Zhang, Q. (2022). Ethical Disputes of AI Surveillance: Case Study of Amazon. *Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022), 648*. https://doi.org/10.2991/aebmr.k.220307.220

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Blanchard, A., Thomas, C., & Taddeo, M. (2024). Ethical governance of artificial intelligence for defence: normative tradeoffs for principle to practice guidance. *AI and Society*, 1–14. https://doi.org/10.1007/s00146-024-01866-7

Bosse, D., Thompson, S., & Ekman, P. (2023). In consilium apparatus: Artificial intelligence, stakeholder reciprocity, and firm performance. *Journal of Business Research*, *155*, 113402. https://doi.org/10.1016/j.jbusres.2022.113402

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert Systems*, *41*(7), e13406. https://doi.org/10.1111/exsy.13406

Chui, M., Manyika, J., & Miremadi, M. (2018). What AI can and can't do (yet) for your business. *McKinsey Quarterly*, *2018*(1), 96–108. https://www.mckinsey.de/~/media/McKinsey/BusinessFunctions/McKinseyAnalytics/OurInsights/WhatAIcanandcan'tdoyetforyourbusiness/What-AI-can-and-cant-do-yet-for-your-business.pdf

Costello, K. W. (2019). Essay on climate apocalypse and a carbon tax. *Electricity Journal*, *32*(10), 106669. https://doi.org/10.1016/j.tej.2019.106669

Crawford, Kate, Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Sánchez, A. N., Raji, D., Rankin, J. L., Rashida Richardson, J., Schultz, A., West, S. M., & Whittaker, M. (2019). AI Now 2019 Report. In *AI Now*. https://ainowinstitute.org/AI_Now_2019_Report.html

Crowe, B., & Rodriguez, J. A. (2024). Identifying and addressing bias in artificial intelligence. *Jama Network Open*, *7*(8), 2425955. https://doi.org/10.1001/jamanetworkopen.2024.25955

D'Alessandro, B., O'Neil, C., & Lagatta, T. (2017). Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification. *Big Data*, *5*(2), 120–134. https://doi.org/10.1089/big.2016.0048

de Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. (2021). Artificial Intelligence Regulation: a framework for governance. *Ethics and Information Technology*, *23*(3), 505–525. https://doi.org/10.1007/s10676-021-09593-z

del Barrio, E., Inouzhe, H., & Matrán, C. (2020). On approximate validation of models: a Kolmogorov–Smirnov-based approach. *Test*, *29*(4), 938–965. https://doi.org/10.1007/s11749-019-00691-1

Ejjami, R. (2024). The Ethical Artificial Intelligence Framework Theory (EAIFT): A New Paradigm for Embedding Ethical Reasoning in AI Systems. *International Journal for Multidisciplinary Research*, *6*(5), 1–15. https://doi.org/10.36948/ijfmr.2024.v06i05.28231

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*, *24*(5), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

Fessenko, D. S., & Jasperse, A. (2024). Ethics at the heart of AI regulation. *AI and Ethics*, *5*, 3387–3398. https://doi.org/10.1007/S43681-024-00562-Y

Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2021). Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Frontiers in Artificial Intelligence*, *3*, 561802. https://doi.org/10.3389/frai.2020.561802

Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. In *Philosophy & Technology* (Vol. 32, pp. 185–193). Springer Netherlands. https://doi.org/https://doi.org/10.1007s13347-019-00354-x

Floridi, L. (2020). AI and Its New Winter: from Myths to Realities. In *Philosophy and Technology* (Vol. 33, Issue 1, pp. 1–3). Springer. https://doi.org/10.1007/s13347-020-00396-6

Fokane, T. (2024). *Towards Ethical AI Regulation in Africa*. Collaboration on International ICT Policy for East and Southern Africa (CIPESA). https://cipesa.org/2024/03/towards-ethical-ai-regulation-in-africa/

Freeman, R. E. (1984). Stakeholder management: Framework and philosophy. In *Strategic Management* (pp. 52–82). https://doi.org/10.1017/cbo9781139192675.006

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hirsch-Kreinsen, H. (2024). Artificial intelligence: a "promising technology." *AI and Society*, *39*(4), 1641–1652. https://doi.org/10.1007/s00146-023-01629-w

Holtzblatt, M. A., Foltin, C., & Tschakert, N. (2020). Learning from ethical violations in public accounting: A South African audit scandal and a firm's transformation. *Issues in Accounting Education*, *35*(2), 37–63. https://doi.org/10.2308/issues-19-062

IoDSA. (2022). *Understanding King IV and what it is intended to achieve*. Sandton: The Institute of Directors in South Africa. https://www.iodsa.co.za/news/389613/Understanding-King-IV-and-what-it-is-intended-to-achieve.htm

Jobin, A., Ienca, M., & Vayena, E. (2019). *Artificial Intelligence: the global landscape of ethics guidelines*. [Online]. http://arxiv.org/abs/1906.11668

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kalota, F. (2024). A Primer on Generative Artificial Intelligence. *Education Sciences, 14*(2), 172. https://doi.org/10.3390/educsci14020172

KPMG. (2021). *The shape of AI governance to come*. Retrieved from https://assets.kpmg/content/dam/kpmg/xx/pdf/2021/01/the-shape-of-ai-governance-to-come.pdf

Laine, J., Minkkinen, M., & Mäntymäki, M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. *Information and Management*, *61*(5), 103969. https://doi.org/10.1016/j.im.2024.103969

Leenes, R., & Lucivero, F. (2014). Laws on robots, laws by robots, laws in robots: Regulating robot behaviour by design. *Law, Innovation and Technology*, *6*(2), 193–220. https://doi.org/10.5235/17579961.6.2.193

Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). Co-Designing checklists to understand organizational challenges and opportunities around fairness in AI. *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. https://doi.org/10.1145/3313831.3376445

Markelius, A., Wright, C., Kuiper, J., Delille, N., & Kuo, Y.-T. (2024). The mechanisms of AI hype and its planetary and social costs. *AI and Ethics*, *4*(3), 727–742. https://doi.org/10.1007/s43681-024-00461-2

Martin, K. D., Kim, J. J., Palmatier, R. W., Steinhoff, L., Stewart, D. W., Walker, B. A., Wang, Y., & Weaven, S. K. (2020). Data privacy in retail. *Journal of Retailing*, *96*(4), 474–489. https://doi.org/10.1016/j.jretai.2020.08.003

Mathrani, A., Susnjak, T., Ramaswami, G., & Barczak, A. (2021). Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. *Computers and Education Open*, *2*, 100060. https://doi.org/10.1016/j.caeo.2021.100060

Matta, V., Bansal, G., Akakpo, F., Christian, S., Jain, S., Poggemann, D., Rousseau, J., & Ward, E. (2022). Diverse perspectives on bias in AI. *Journal of Information Technology Case and Application Research*, *24*(2), 135–143. https://doi.org/10.1080/15228053.2022.2095776

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, *3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Mohammed Shebeen, T., Shanthi, R., & Mathiyarasan, M. (2024). Navigating Industry 4.0: Skill Development Strategies for Empowering MSMEs in the Digital Age. In R. El Khoury (Ed.), *Anticipating Future Business Trends: Navigating Artificial Intelligence Innovations, Studies in Systems, Decision and Control* (Vol. 536, pp. 453–462). Cham: Springer. https://doi.org/10.1007/978-3-031-63402-4_38

Morante, G., Viloria-Nunez, C., Florez-Hamburger, J., & Capdevilla-Molinares, H. (2024). Proposal of an Ethical and Social Responsibility Framework for Sustainable Value Generation in AI. 2024 *IEEE Technology and Engineering Management Society, TEMSCON LATAM* 2024. https://doi.org/10.1109/TEMSCONLATAM61834.2024.10717855

Ormond, E. (2023). AI in South Africa comes with special dilemmas – plus the usual risks. *Daily Maverick*, 8 February 2023. https://www.dailymaverick.co.za/article/2023-02-08-artificial-intelligence-in-south-africa-comes-with-special-dilemmas-plus-the-usual-risks/

Pasipamire, N., & Muroyiwa, A. (2024). Navigating algorithm bias in AI: ensuring fairness and trust in Africa. *Frontiers in Research Metrics and Analytics*, *9*, 1486600. https://doi.org/10.3389/frma.2024.1486600

Perifanis, N.-A., & Kitsios, F. (2023). Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review. *Information*, *14*(2), 85. https://doi.org/10.3390/info14020085

Raisch, S., & Krakowski, S. (2020). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review*, 1–48. https://doi.org/10.5465/2018.0072

Resnick, P., Zeckhauser, R., Friedman, E., & Kuwabara, K. (2000). Reputation systems. *Communications of the ACM*, *43*(12), 45–48. https://doi.org/10.1145/355112.355122

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. *In Companion Proceedings of the 2019 World Wide Web Conference*, 10. https://doi.org/10.1145/3308560.3317590

SAICA. (2022). *What is a Chartered Accountant CA(SA)*. The South African Institute of Chartered Accountants (SAICA). [Online]. https://www.saica.org.za/become-a-member/how-to-become-a-member/chartered-accountant-casa/what-is-a-chartered-accountant-casa

Sakamoto, T., Van Nguyen, N., Ohno, H., Ishitsuka, N., & Yokozawa, M. (2006). Spatio-temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers. *Remote Sensing of Environment*, *100*(1), 1–16. https://doi.org/10.1016/j.rse.2005.09.007

Silberg, J., & Manyika, J. (2019). *Tackling bias in artificial intelligence (and in humans)*. McKinsey Global Institute. [Online]. https://www.mckinsey.com/featured-insights/artificial

Tan, T. M., & Salo, J. (2023). Ethical Marketing in the Blockchain-Based Sharing Economy: Theoretical Integration and Guiding Insights. *Journal of Business Ethics*, *183*(4), 1113–1140. https://doi.org/10.1007/s10551-021-05015-8

Turlapati, V. R., Vichitra, P., Raval, N., Khaja Mohinuddeen, J., & Mishra, B. R. (2024). Ethical Implications of Artificial Intelligence in Business Decision-making: A Framework for Responsible AI Adoption. *Journal of Informatics Education and Research*, *4*(1), 530–537. https://doi.org/10.52783/jier.v4i1.597

Uttley, J. (2019). Power Analysis, Sample Size, and Assessment of Statistical Assumptions—Improving the Evidential Value of Lighting Research. In *LEUKOS 15*(2–3), 143–162. https://doi.org/10.1080/15502724.2018.1533851

Appendix A

Explanation about transparency included in PRISMA

Unexplainable AI	Auditing and Inspection	Corrective Action	Harmful Transparency	Seeing over Understanding	Temporal limitations	Technical Limitations
Transparency Deloitte (2017) and Turek (2018) state that to overcome transparency, the construct of unexplainable AI should be addressed. Unexplainable AI, also termed black-box nature, refers to where users are highly reliant on the decision making of an AI, with feeling of inadequacy and lack of competence in their use of the AI owing to little or no understanding of how it functions (Bankins & Formosa, 2023). Therefore, to	According to Deloitte (2018) and Ryngaert, and Taylor (2020), business should perform regular auditing and inspection of its AI solutions to overcome transparency concerns. Thus, as a process, business should use a range of available AI techniques for inspection and auditing purposes.	Burt (2019) notes that since organisations have long been aware of the transparency paradox, they must update their methods to address AI corrective action. This is crucial because transparency of an algorithm does not necessarily translate to corrective action being taken (Hood, 2006).	Ananny and Crawford (2018) submit that if information is shared with stakeholders without understanding of why the information is being revealed, it could be harmful to the business. Moreover, where too much algorithm is revealed, detractors could use the information to harm the business or even steal proprietary investments and trade secrets.	Transparency can also result in seeing over understanding. Learning about complex systems means not simply being able to look inside systems or take them apart. According to Resnick et al., (2000:46), it means "dynamically interacting with [the systems] in order to understand how they behave in relation to their environments". Thus, as a process, business should have independent observers debate and challenge AI tools from a	Temporal limitations are brought to the fore because of the ever-changing nature of technology and data used in AI (Ananny & Crawford, 2018). Owing to the rapid development of technology and constant movements in data, a temporal limitation is caused (Ananny & Crawford, 2018). Thus, to overcome the temporal limitation, it is proposed as a governance process that inspection or auditing of the system would	There are technical limitations to transparency that would need to be disclosed (Burrell, 2016; Crain, 2018). Businesses must disclose technical limitations to stakeholders, as creators of these AI systems often do not understand the logic behind the systems' decisions. Thus, businesses should disclose the technical limitations of the AI system to stakeholders. They should also undertake

(*Continued*)

Continued.

Unexplainable AI	Auditing and Inspection	Corrective Action	Harmful Transparency	Seeing over Understanding	Temporal limitations	Technical Limitations
overcome unexplainable AI, technical methods like inspections can be used (Lang, Bernier, & Knoppers, 2022; Mulligan, & Elaluf-Calderwood, 2021).			Thus, as a process, business should only disclose relevant AI information regarding transparency to stakeholders. They must also state clearly why the disclosure of relevant AI information is necessary (Crain, 2018; Schudson, 2015; Stohl et al., 2016).	transparency perspective. The independent observers provide insight into how the AI system operates. They are a neutral person or an entity with the necessary knowledge to observe but with no direct interest or links to the matter under consideration, thus, ensuring no bias.	require a comprehensive understanding of previous iterations of the system, how they operated, and why they changed (Minsky, 2007).	continuous monitoring and testing of the system to understand technical limitations.

Source: Authors' own contribution

From the above, it can be deduced that transparency in AI governance involves addressing key issues such as unexplainable AI (i.e., black-box effect), auditing, harmful transparency, temporal, and technical limitations. To overcome unexplainable AI, businesses should adopt inspection methods and develop explainable algorithms. Similarly, regular auditing and inspection as well as corrective actions to ensure accountability, should be considered. In addition, careful disclosure of AI information is crucial to avoid harmful transparency risks like security breaches. Temporal limitations would need comprehensive auditing that considers previous system iterations, while technical limitations require continuous monitoring and testing, because both system developers and users may not fully understand the AI decisions. These measures would collectively foster responsible and transparent AI governance.

Explanation about machine bias included in PRISMA

Types of AI bias	Strategy to overcome bias in business	Design and feedback	Monitoring and training
Machine Bias First, business needs to be aware of the different forms of bias that could present themselves when using AI. These include: Historical bias, Representation bias, Measurement bias, Evaluation bias, Aggregation bias, Simpson's Paradox, Sampling bias, Temporal bias, Popularity bias, Algorithmic bias, Funding bias, Data manipulation, Data collection, ML bias (Fletcher et al., 2021; Mehrabi et al., 2021; Pasipamire & Muroyiwa, 2024).	Business should create a de-biasing strategy, which includes technical and operational components to overcome machine bias. This should include the following aspects: • Integrate fairness and non-discrimination into the workflow of AI processes; • Adopt pre-processing, in-processing, and post-processing techniques to evaluate machine bias; • Create datasheets for the algorithm. • Include the motivation for creating the algorithm in the datasheets; • Include the creation method of the algorithm in the datasheets; • Include methods employed for reducing bias in the datasheets; • Include methods for identifying machine bias in the datasheets; • Assess the risk of bias present by obtaining a full understanding of the algorithm and data; • Determine if the collection of data is biased though creating internal teams or using third parties; • Have internal teams or third parties focus on the different types of bias; • Develop processes to deal with algorithmic bias once it has been identified, and • Report to stakeholders on the	Embedding early adverse feedback in the design process is a key aspect within design and feedback, including test data (Google, 2020). Thus, businesses should develop early adverse feedback in the design process, including a small live test.	Burt (2020), Google (2021) and Microsoft (2021) note that multiple matrices are important to access training and continuous monitoring of the AI system. Google (2021) goes further by recommending that matrices should include feedback from user surveys, quantities that track overall system performance, short- and long-term product health, as well as false positive and false negative rates which cut across different sub-groups. Thus, businesses should create multiple matrices to access training and continuous monitoring of the AI system.

(*Continued*)

Continued.

Types of AI bias	Strategy to overcome bias in business	Design and feedback	Monitoring and training
	technical and operational machine bias strategy outcomes (d'Alessandro, O'Neil & LaGatta, 2017; Kamishima et al., 2012; Kantarci, 2021; Madaio et al., 2020; Mitchell et al., 2019; Silberg & Manyika, 2019).		

Source: Authors' own contribution

It is observed that to address machine bias in AI governance, businesses must first recognise the various types of bias existing on AI technology including historical, representation, and algorithmic biases, most of which are inherent to the system or are sometimes human induced. Thus, having a comprehensive de-biasing strategy is important, and so is incorporating fairness, evaluation techniques, keeping updated datasheets, documenting algorithm creation and bias mitigation methods, and stakeholder reporting. Developing an effective design process should embed early adverse feedback mechanisms, which will include live testing. Furthermore, continuous monitoring and training must be considered, with businesses employing multiple measures to assess system performance and user feedback, ensuring ongoing refinement and reduced bias in AI systems.

Explanation about privacy included in PRISMA

Design	Large data sets	Transparency and bias resulting in privacy concerns
Privacy Business needs to create or purchase technology that embodies privacy by design (Norwegian Data Protection Authority Report, 2018; Pagallo, 2021)	Data minimisation should be adopted by business (Aledhari et al., 2018; Norwegian Data Protection Authority Report, 2018, OECD, 2018; Organisation for Economic Co-operation and Development (OECD), 2018). Business should seek to minimise the data that are required for training and limit the use of data to only relevant data and create algorithms with the least amount of invasiveness to data subjects (Aledhari et al., 2018).	Business should conduct data protection impact risk assessments to overcome the possibility of transparency and bias resulting in privacy concerns (Aledhari et al., 2018; Institute of Electrical & Electronic Engineers (IEEE), 2017). Thus, as a process, business should perform a risk assessment to understand the impact of transparency and machine bias on privacy; document a full description of how the data will be used; assess whether the use of the data collected is necessary and proportional to its purpose; undertake an assessment to understand the risk of

(*Continued*)

Continued.

Design	Large data sets	Transparency and bias resulting in privacy concerns
		infringement of people's rights; document the measures selected for managing the identified risk, and identify if POPIA or any other legislation needs to be met using AI.

Source: Authors' own contribution

To safeguard privacy in AI governance, it is crucial for organisations to adopt privacy-by-design principles and practice data minimisation, limit data collection and usage to what is strictly necessary. Also, it is advisable to develop algorithms that are minimally invasive, while measures for data protection impact assessments are deemed critical. These assessments, we believe, can foster risks of transparency and bias on privacy, document data usage, and assess necessity and proportionality. Besides, businesses should outline measures that will assist to minimise risks, ensure compliance with policies and regulations like the POPIA, and uphold individuals' rights in AI systems.

Explanation about beneficial AI included in PRISMA

Human Rights	Health and Safety	Financial Wellbeing	Intellectual Wellness
Beneficial AI			
Businesses should uphold and respect human rights when creating and using AI systems (Partnership on AI, 2019; Plastino & Purdy, 2018; Raisch & Krakowski, 2021). Thus, as a business process, when creating an AI system, business should follow a human-centred approach, that involves augmentation rather than automation.	Businesses should prioritise the health and safety of all stakeholders by maintaining adherence to technical safety and legal standards, including providing adequate training (Crawford et al., 2019; Partnership on AI, 2019, Raisch & Krakowski, 2021). Thus, they proactively understand the impact of AI on the physical health of employees and proactively assess work- life integration (because AI has the potential to blur the boundaries between the two dimensions).	Businesses should fully understand the impact of AI on job losses, gains, and transitions, including actively upskilling individuals to obtain multi- faceted job profiles (Crawford et al., 2019; Partnership on AI, 2019); understanding financial wellness of communities affected by AI; pursuing a strategy that ensures that vulnerable communities and groups derive equal benefit from the introduction of AI, and engaging a wide range of stakeholders to ensure that maximum number of people's views are understood regarding the AI system and maximum number of people benefit from the AI system.	Businesses should be aware of the impact of AI on workers' cognitive abilities (IEEE, 2019; Partnership on AI, 2019). Cognitive capabilities used require that organisations should consider how AI adoption affects workers' capacity to learn, think critically, solve problems, and make decisions. Thus, as a business process, they should actively upskill individuals so that they obtain multi-faceted job profiles and design systems to retain workers who are affected by AI.

Source: Authors' own contribution

Here, we noted that it is important for businesses to ensure that AI systems uphold human rights, promote health and safety, support financial wellbeing, and foster intellectual wellness. There is need to consider adopting a stakeholder-centric approach, upskilling workers, safeguarding job transitions, maintaining work-life balance, and enhancing cognitive abilities through thoughtful system design.

Explanation about responsible AI included in PRISMA

Transparency relating to responsible AI	Responsible AI environment
Responsible AI To create a responsible AI environment within an entity, the concepts of transparency, coupled with explainable AI and trust, must be entrenched in organisational policies (Barredo-Arrieta et al., 2020; Crawford et al., 2019). There is need to create policies that are clear and accessible to design and development teams, including all those who are involved in the creation of the AI tools.	Business would need to create policies that are clear and accessible to design and development teams, including all those who are involved in the creation of the AI tools (IBM, 2020). Thus, as a process, business should have a clear understanding of where the responsibility for the business starts and ends and have a detailed record of all decisions regarding the AI tools.

Source: Authors' own contribution
Here, we noted that to foster responsible AI environment, organisations need to integrate transparency, explainable AI, and trust into their organisational policies. Also, clear and accessible policies should be provided as this will serve as a good guide for the design and development teams, thus ensuring accountability and maintaining detailed records of all decisions regarding the AI tools.

Appendix B

Articles reviewed by PRISMA approach

Article	Database	Construct that informed the CG process	Article Details
1	EBSCOHOST	Large Data Sets	Aledhari, M., Di Pierro, M., Hefeida, M. and Saeed, F. (2018). A deep learning-based data minimization algorithm for fast and secure transfer of big genomic datasets. *IEEE Transactions on Big Data*, *7*(2), 271–184.
2	EBSCOHOST	Temporal Limitations; Seeing over Understanding; Harmful Transparency	Ananny, M. and Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1109/TBDATA.2018.2805687
3	EBSCOHOST	Privacy by Design	Bang, J., Kim, S., Nam, J.W. and Yang, D.G. (2021, August). Ethical chatbot design for reducing negative effects of biased data and unethical conversations. In 2021 *International Conference on Platform Technology and Service (PlatCon)* (pp. 1–5). IEEE.
4	EBSCOHOST	Responsible AI Environment; Transparency Relating to Responsible AI	Barredo-Arrieta, A., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020). Explainable

(*Continued*)

Continued.

Article	Database	Construct that informed the CG process	Article Details
			artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
5	EBSCOHOST	Technical Limitations	Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1),.2053951715622512.
6	EBSCOHOST	Monitoring and Training; Corrective Action	Burt, A. (2019). The AI transparency paradox. *Harvard Business Review*, *13*.
7	EBSCOHOST	Strategy to Overcome Bias in Business	Chandra, R., Sanjaya, K., Aravind, A.R., Abbas, A.R. and Gulrukh, R. (2023). Algorithmic Fairness and Bias in Machine Learning Systems. In *E3S Web of Conferences* (Vol. 399, p. 04036). EDP Sciences.
8	EBSCOHOST	Health and Safety	Chatterjee, S. (2020). AI strategy of India: Policy framework, adoption challenges and actions for government. *Transforming Government: People, Process and Policy*, *14*(5), 757–775.
9	Google Scholar	Responsible AI Environment	Cheatham, B., Javanmardian, K. and Samandari, H. (2019). Confronting the risks of artificial intelligence. *McKinsey Quarterly, 2*(38), 1–9.
10	EBSCOHOST	Technical Limitations	Crain, M., 2018. The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88–104.
11	Google Scholar	Financial Wellbeing; Intellectual Wellness; Transparency Relating to Responsible AI; Health and Safety	Crawford, K., Dobbe, R., Myers West, S., Kak, A., Dryer, T., Fried, G., Kaziunas, E., Mathur, V., Richardson, R., Schultz, J., Schwartz, O. and Whittaker, M. (2018). *AI Now Report 2018* (pp. 1–62). New York: AI Now Institute at New York University.
12	EBSCOHOST	Strategy to Overcome Bias in Business	d'Alessandro, B., O'Neil, C. and LaGatta, T. (2017). Conscientious classification: A data scientist's guide to discrimination-aware classification. *Big Data*, *5*(2), 120–134.
13	EBSCOHOST	Strategy to Overcome Bias in Business	Giovanola, B. and Tiribelli, S. (2023). Beyond bias and discrimination: Redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & Society*, *38*(2), 549–563.
14	Google Scholar	Monitoring and Training; Financial Wellbeing;	Pichai, S (2018). Artificial Intelligence at Google: Our Principles. Retrieved from

(*Continued*)

Continued.

Article	Database	Construct that informed the CG process	Article Details
		Intellectual Wellness; Transparency Relating to Responsible AI; Design and Feedback; Corrective Action	https://blog.google/technology/ai/ai-principles/
15	EBSCOHOST	Corrective Action	Hood, C. and Heald, D. (2006). *Transparency: The key to better governance?* (Vol. 135). Oxford: Oxford University Press for The British Academy.
16	Google Scholar	Responsible AI Environment	Engstrom, D.F. and Ho, D.E. (2020). Algorithmic accountability in the administrative state. *Yale Journal on Regulation*, *37*, 800.
17	EBSCOHOST	Transparency and Bias resulting in Privacy Concerns	Shahriari, K. and Shahriari, M. (2017). IEEE standard review – ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In 2017 *Canada International Humanitarian Technology Conference* (*IHTC*) (pp. 197–201). IEEE
18	EBSCOHOST	Intellectual Wellness	IEEE. (2019). Ethically Aligned Design. [Online]. Retrieved from https://ethicsinaction.ieee.org
19	EBSCOHOST	Strategy to Overcome Bias in Business	Kamishima, T., Akaho, S., Asoh, H. and Sakuma, J. (2012). Fairness-aware classifier with prejudice remover regularizer. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part II, 23* (pp. 35-50). Berlin Heidelberg: Springer.
20	EBSCOHOST	Strategy to Overcome Bias in Business	Kantarci, A. (2021). Bias in AI: What it is, types & examples, how & tools to fix it. Estonia: AIMultiple.
21	Google Scholar	Responsible AI Environment	Katyal, V., Liepold, C., & Iyengar, S. (2020, June 25). Artificial Intelligence and Ethics: An Emerging Area of Board Oversight Responsibility. Harvard Law School Forum on Corporate Governance. Retrieved from https://corpgov.law.harvard.edu/2020/06/25/artificial-intelligence-and-ethics-an-emerging-area-of-board-oversight-responsibility/
22	EBSCOHOST	Unexplainable AI	Lang, M., Bernier, A. and Knoppers, B.M. (2022). Artificial intelligence in cardiovascular imaging: "Unexplainable" legal and ethical challenges? *Canadian Journal of Cardiology, 38*(2), 225–233.

(*Continued*)

Continued.

Article	Database	Construct that informed the CG process	Article Details
23	EBSCOHOST	Responsible AI Environment; Privacy by Design	Lu, Q., Zhu, L., Xu, X. and Whittle, J. (2023). Responsible-AI-by-design: A pattern collection for designing responsible artificial intelligence systems. *IEEE Software*, *40*(3), 63–71. https://doi.org/10.1109/MS.2022.3233582
24	Google Scholar	Strategy to Overcome Bias in Business	Madaio, M.A., Stark, L., Wortman Vaughan, J. and Wallach, H. (2020, April). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14).
25	Google Scholar	Monitoring and Training	Microsoft. (2021). Artificial Intelligence at Google: Our Principles. Retrieved from https://www.microsoft.com/en-us/ai/our-approach?activetab=pivot1%3a primaryr5
26	EBSCOHOST	Temporal Limitations	Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind.* Simon and Schuster.
27	Google Scholar	Strategy to Overcome Bias in Business	Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T. (2019, January). Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220–229).
28	EBSCOHOST	Unexplainable AI	Mulligan, C. and Elaluf-Calderwood, S. (2022). AI ethics: A framework for measuring embodied carbon in AI systems. *AI and Ethics*, 1–13.
29	EBSCOHOST	Monitoring and Training Transparency and Bias Resulting in Privacy Concerns; Privacy by Design	Lohne, K. (2012). The Norwegian Data Inspectorate: Between Governance and Resistance. *Surveillance & Society, 10*(2). Accessed at SSRN: https://ssrn.com/abstract=3173168
30	Google Scholar	Large Data Sets	Höchtl, J., Parycek, P. and Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 147–169.
31	Google Scholar	Transparency and Bias Resulting in Privacy Concerns; Privacy by Design	Pagallo, U., Aurucci, P., Casanovas, P., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Schafer, B. and Valcke, P. (2019). AI 4People-on good AI governance: 14 priority

(*Continued*)

Continued.

Article	Database	Construct that informed the CG process	Article Details
			actions, a SMART model of governance, and a regulatory toolbox. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3486508
32	Google Scholar	Financial Wellbeing; Intellectual Wellness; Human Rights	Available from: https://www.partnershiponai.org/workforce-wellbeing/
33	EBSCOHOST	Intellectual Wellness; Human Rights	Plastino, E. and Purdy, M. (2018). Game changing value from Artificial Intelligence: Eight strategies. *Strategy & Leadership*, *46*(1), 16–22.
34	Google Scholar	Transparency Relating to Responsible AI	Wang, Y., Xiong, M. and Olya, H. (2020, January). Toward an understanding of responsible artificial intelligence practices. In *Proceedings of the 53rd Hawaii international conference on system sciences* (pp. 4962–4971).
35	EBSCOHOST	Health and Safety; Human Rights	Raisch, S. and Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.
36	EBSCOHOST	Seeing over Understanding	Resnick, M., Berg, R. and Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences*, *9*(1), 7–30.
37	Google Scholar	Auditing and Inspection; Unexplainable AI	Roovers, R. (2019). Transparency and responsibility in artificial intelligence. A call for explainable AI. [Accessed on 25 December]. p. 2022
38	Google Scholar	Responsible AI Environment	*MIT Sloan Management Review*. (2020). How AI Is Transforming the Organization. MIT Press.
39	EBSCOHOST	Intellectual Wellness	Russell, S. (2022, March). Provably beneficial artificial intelligence. *In Proceedings of the 27th International Conference on Intelligent User Interfaces* (p. 3).
40	EBSCOHOST	Corrective Action; Auditing and inspection	Ryngaert, C. and Taylor, M. (2020). The GDPR as global data protection regulation?
41	EBSCOHOST	Harmful Transparency	Schudson, M. (2015). *The rise of the right to know: Politics and the culture of transparency, 1945–1975*. Cambridge, MA: Belknap Publishing.
42	EBSCOHOST	Strategy to Overcome Bias in Business	Silberg, J. and Manyika, J. (2019). Tackling bias in artificial intelligence

(*Continued*)

Continued.

Article	Database	Construct that informed the CG process	Article Details
			(and in humans). *In McKinsey. Notes from the AI frontier*. 1–8.
43	EBSCOHOST	Responsible AI Environment	Stahl, B.C. (2023). Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems. *Scientific Reports*, *13*(1), 7586.
44	Google Scholar	Human Rights	Stix, C. (2023). Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: Multifaceted Framework for Governance. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Eindhoven University of Technology.
45	EBSCOHOST	Harmful Transparency	Stohl, C., Stohl, M. and Leonardi, P.M. (2016). Digital age| managing opacity: Information visibility and the paradox of transparency in the digital age. *International Journal of Communication*, *10*, 15.
46	EBSCOHOST	Unexplainable AI	Turek, Gunning, Vorm, Wang, et al. (2021, November 15). DARPA's Explainable AI (XAI) program: A retrospective. *TechRxiv*.
47	EBSCOHOST	Privacy By Design	Umbrello, S. (2019). Beneficial artificial intelligence coordination by means of a value sensitive design approach. *Big Data and Cognitive Computing*, *3*(1), 5.

Appendix C

A copy of the questionnaire will be made available upon request from the corresponding author.