Identifying AI Corporate Governance Principles That Should Be Prevalent in a Governance Framework for Business

Coovadia Husain, Marx Benjamin, and Ilse Botha

Abstract Artificial Intelligence (AI) is widely used in business to increase productivity and harness the benefits that could emerge from is use. However, with the increased use of AI in business there are number of risks that are brought to the fore. The task would be to develop sound AI corporate governance principles to reduce the AI risk. To the extent of literature search research in AI and corporate governance does not position AI principles that need to be included in any AI corporate governance framework from a South African perspective. Given the importance of AI corporate governance AI governance principles will be identified to be included in an AI governance framework. Through a documentary analysis of literature this study identifies eight broad themes with various corporate governance principles that need to be prevalent in an AI governance framework for South Africa. These eight broad themes include (1) Principle concerns, (2) Procedural governance mechanisms, (3) Overarching ethical concerns, (4) Reasons for creating AI governance frameworks, (5) AI applications and technology layer, (6) AI law regulation, (7) AI Society, (8) AI regulation and process layer. It is essential that business start considering these themes when developing an AI governance framework that will be implemented in business.

Keywords AI · AI Governance · Machine learning · Natural language processing · Corporate Governance

1 Introduction

Artificial intelligence (AI) is a broad field that encompasses computer science, psychology, philosophy, linguistics, and many other areas of study (Deloitte, 2018). "AI involves the analysis of big data to allow a machine to reason, learn and problem-solve" (Haenlein & Kaplan, 2019, p. 4). These faculties of

C. Husain (✉) · M. Benjamin · I. Botha
University of Johannesburg, Johannesburg, South Africa
e-mail: hcoovadia@uj.ac.za; benm@uj.ac.za; ilseb@uj.ac.za

T. Moloi, B. George (eds.), *Towards Digitally Transforming Accounting and Business Processes*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-46177-4_15

problem-solving have enticed businesses to harness the vast possibilities of AI (Mittelstadt et al., 2016, p. 1). Investments in AI to better processes are currently undertaken by many businesses globally (Crawford et al., 2019). According to Deloitte (2018), 37% of businesses using AI have invested more than US$five million into AI technologies. Deloitte (2018) adds that businesses that have made investments in AI are now seeing the benefits. They further stated that these companies have indicated that AI will enhance their operations and more importantly influence their decision making (Edelman, 2019). However, with the increased use of AI in business there are a number of risks that are brought to the fore. According to Boddington (2017), the task would be to develop sound AI corporate governance principles to reduce the AI risk.

2 Literature Review

Corporate governance is the alignment of all material stakeholders, including society, economics, individuals, and community to create value for all in a responsible manner (Institute of Directors South Africa, 2022). This view of corporate governance implies that wealth should be created for all stakeholders, but not to the detriment of any other stakeholder in the process. Boddington (2017) concurs with this understanding of corporate governance, in which it is stated that corporate governance essentially involves balancing the interests of a company's many stakeholders. Corporate governance is further described as the framework of rules, practices, and processes used to direct and manage a company (Haes & Grembergen, 2015). This view establishes the notion that corporate governance relies on the creation of certain processes and frameworks to guide a business in making the correct decisions for all material stakeholders. King IV™ defines corporate governance as "the exercise of ethical and effective leadership by the governing body towards the achievement of the governance outcomes of ethical culture, good performance, effective control, and legitimacy" (Institute of Directors South Africa, 2022, p. 15). This means that corporate governance is creating ethical value for all material stakeholders. It states that the "governing body should govern technology and information in a way that supports the organization setting and achieving its strategic objectives" (Institute of Directors South Africa, 2022, p. 18). This means that those that are charged with governance of an entity should create technological governance processes that provide direction to the entity, which could include aspects such as risk management, performance management, proactive monitoring of the technology, and ethical use of technology.

In addition to the above, AI corporate governance is defined as the process of defining policies and establishing accountability to guide the creation and deployment of AI systems in an organization (Mäntymäki et al., 2022). When done correctly, AI corporate governance empowers organizations to operate with agility and complete trust, rather than slowing them down. Moreover, the definition of AI corporate governance is noted to be the development of rules, practices, and

processes used to ensure that the organization's AI technology sustains and extends the organization's strategies and objectives (Abraham et al., 2019). Therefore, AI corporate governance can be defined as creating processes, rules, and practices for an entity that facilitates wealth creation for all stakeholders in an ethical manner.

However, current research by Butcher and Beridze (2019) and Schneider et al. (2020) in AI and corporate governance does not position AI principles that need to be included in any AI corporate governance framework from a South African perspective. Scholars including Boddington (2017) and Crawford et al. (2019) have outlined that each nation needs to develop its own set of AI ethical governance principles for AI and ethics. This is also echoed by the European Commission's High-Level Expert Group, the AI4People and the Institute of Electrical and Electronics Engineers (IEEE). Crawford et al. (2019) emphasise that most AI policies and statements are generated by the Global North, whereas the Global South is largely absent. Given the importance of AI corporate governance it would be essential that the core AI governance principles be identified that should be prevalent in an AI governance framework from a South African perspective. This creates an opportunity for the Global South, and more specifically, for South Africa, to determine which governance requirements are necessary for businesses to adopt, and in the process, determine the AI governance principles to create a conducive AI environment.

3 Objective

To bridge the gap between AI's potential and risks, stakeholders are asking for increased guidance on how to govern AI and manage the implications and risk of unintended outcomes (KPMG, 2021). Stakeholders realize that an AI governance framework can provide organizations with a much-needed mechanism to be proactive in governing, managing, and instilling trust in their technologies (KPMG, 2021). However, current research in AI and corporate governance does not position core AI principles that need to be included in any AI corporate governance framework form a South African perspective. The objective of this paper is to identify through literature the core AI principles that would need to prevalent in an AI governance framework for South African business.

4 Methods and Methodology

The method of this study is a documentary analysis. This method involves the systematic review or assessment of documents. Similarly, to other analytical methods in qualitative research, document analysis requires data to be examined and interpreted to draw meaning and create knowledge (Corbin & Strauss, 2008). Document analysis involves scanning, reading, and interpretation of documents.

This iterative process of scanning, reading, and interpretation results in a thematic analysis. Thematic analysis results in pattern recognition from the data (Fereday & Muir-Cochrane, 2006).

In this study, documentary analysis was used to review literature and identify core AI principles that need to be prevalent in an AI governance framework. The documentary analysis included peer-reviewed journals from the University of Johannesburg (UJ) databases, Google Scholar searches, publicly available records, legislation, white papers, reporting standards, and MIT open lab records. A documentary analysis of literature was undertaken to explain the AI governance processes and establish which should be prevalent in a Governance Framework for business.

5 Results

5.1 Core Principles for Corporate Governance Frameworks

Through the documentary analysis of literature five AI corporate governance themes were identified. These included principal concerns, procedural and structural mechanisms, overachieving ethical concerns, reasons for creating AI frameworks and four layers to create an intergrade AI governance framework. Each of these themes consisted of various principles. Each of these will be explored further to establish AI Governance Principles that should be prevalent in a corporate governance framework for business (Fig. 1).

5.2 Principles Concerns

There are many ethical principles that are brought to the fore when reviewing current literature. This is evident from the work undertaken by Jobin et al. (2019) in their meta-analysis of AI ethical principles. They identified key overlap of principles in various AI ethical documents, including principles such as transparency, justice, and non-maleficence. This was further echoed in a study by Fjeld et al. (2020), in which they refer to the principles of fairness and non-discrimination. These principles, noted in literature, were developed over a period of time, based on current needs and requirements (Fjeld et al., 2020). However, developing governance around these principles is not a straightforward process, but rather one that requires rigorous discussion and debates, as these principles form the backbone of good governance in AI. They allow complex constructs to be narrowed down into a few central themes (Hickok, 2021). Moreover, principles developed allow commitment to a wide range of shared values and can therefore result in influencing institutional decision-making processes (Whittlestone et al., 2019). In addition, they can address public concerns by clarifying the commitment of a business to good governance of AI (Whittlestone et al., 2019). Moreover, these principles created also allow and provide for an

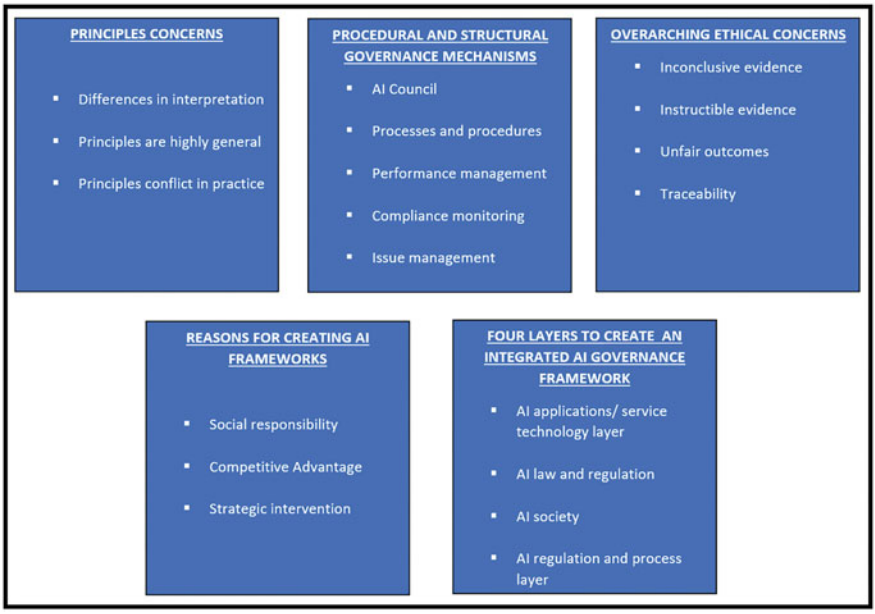

Fig. 1 AI coporate governance themes and its assoicated principles. Source: Authors deduction

informal basis of holding individuals accountable (Morley et al., 2021). The development and convergence of these principles is a key process in creating an AI environment with strong governance (Hickok, 2021).

However, Whittlestone et al. (2019) and Dancy (2004) argue that these core principles noted in the literature are not sufficient to address AI governance. This is due to the number of limitations that the principles bring to the fore that would need to be addressed (Whittlestone et al., 2019). These limitations include that different groups may interpret principles differently, that the principles are too general, and that principles conflict with that which takes place in real life. Whittlestone et al. (2019) and Dancy (2004) argue that no weight should be given to principles, while others (Whittlestone et al., 2019) hold that principles should be considered in conjunction with other facts that are brought to the fore. No matter the views of the scholars, they all agree that, if these principles are used, the limitations that come to the fore need to be addressed (Whittlestone et al., 2019). The discussion will now take a closer look at three limitations noted.

5.2.1 Differences in Interpretation

The terms used in AI governance are complex and can be ambiguous by their very nature (Whittlestone et al., 2019). As Clouser and Gert (1990) noted, this leads to various interpretations of the definitions of the words provided. In addition,

principles do not take into consideration the legitimate differences in values across various cultures and regions. For example, in bioethics, the term "justice' is not defined and thus this leaves it to the user, across regions and cultures, to determine what would be just and what would constitute unjust behaviour (Clouser & Gert, 1990).

Within the case of AI governance, this is no different. While all might agree to a certain principle in AI, for example, that the term 'fairness' is important, there are deep political and social differences about what constitutes fairness (Binns, 2017) In addition, based on the interpretation of principles, certain groups of people may put emphasis on a certain principle more than on others, leading to further concerns in interpretations (Winfield & Jirotka, 2018). The point now arises: how does one overcome this limitation of differences in interpretation? This can be achieved by ensuring that the governance framework created is more concise. This will require that each principle identified to be part of the governance framework is defined, limiting the interpretation differences (Winfield & Jirotka, 2018). Thus, as a core AI principle, it is noted that each term that is part of the AI governance framework should be defined.

5.2.2 Principles Are Highly General

Another concern in regard to governance of AI principles is that the principles are highly general in nature (Whittlestone et al., 2019). This means that they can serve as a kind of a checklist to be taken into consideration, however, from a practical point of view, they may not be useful due to the generalness of the principle (Beauchamp, 1995). Principles that are currently elaborated are very broad, for example, AI should be used for the common good, which everyone can agree on, but, from a practical perspective, what does that mean (Winfield & Jirotka, 2018)? An example of a narrower principle is not to use AI to develop autonomous weapons (Winfield & Jirotka, 2018). This is a much more specific principle but in essence is limited to a specific scenario in specific sectors. A balance between highly general and too specific principles is required in an AI governance framework. This can be achieved in an AI governance framework by each principle in a framework being divided into smaller concerns, creating more concise explanations. Hence, as a core AI principle, each principle identified should be divided into smaller concepts, creating more concise explanations.

5.2.3 Conflict in Practice

The last limitation of AI principles lies in the fact that a specific principle may conflict with what takes place in practice in the world of work (Winfield & Jirotka, 2018). For example, a current AI principle from the UK House of Lords AI Committee states that "it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual's life, unless it can generate a

full and satisfactory explanation for the decisions it will take" (Winfield & Jirotka, 2018, p. 10). From a pure principal perspective, this has good intentions, but when applied in practice it results in a conflict. For example, today there are algorithms that can diagnose medical conditions better and faster than a human doctor, however, there is no concrete explanation of the algorithm (Song et al., 2018). The benefit of using this algorithm in conjunction with a human doctor may save lives, but due to the principle mentioned above, such technology would not be deployed. Some developments may be so beneficial that trade-offs would need to be made and this is an example of such a case (Price, 2017). Using the blanket principle would stifle potential AI uses and thus cause a conflict between the principle and what takes place in practice. To overcome this limitation, any AI governance framework developed must be presented to persons in practice to identify if there is any conflict that may arise with application in practice. Accordingly, each item raised in the AI governance framework, with its explanations and definitions, should be provided to persons in practice for further insight.

From the above there are three key aspects described regarding principle concerns, namely that principles are general; they conflict in practise, and interpretation of principles could vary. Each of these aspects would need to be addressed to overcome the principle concerned. The next aspect is procedural and structural governance mechanisms.

5.3 Procedural and Structural Governance Mechanisms

When creating an AI governance framework, there are different governance mechanisms that come to the fore. These include structural mechanisms and procedural mechanisms (Schiff et al., 2020). The discussion will now look at these two different governance mechanisms and conclude on the mechanism suitable for this study to develop an AI governance framework.

5.3.1 Structural Mechanisms

This mechanism is established in the notion of creating an AI governance council rather than AI governance frameworks (Schneider et al., 2020). This council would handle complex AI-related questions, including the interrelation between model outputs, training data, regulatory, and business requirements (Reddy et al., 2020). This council would comprise of (i) roles and responsibilities, and (ii) the allocation of decision-making authority (Borgman et al., 2016). Using structural mechanisms within AI governance is not well researched and very sparse data is currently available to understand the mechanics around using this approach (Ho et al., 2019). In addition, this approach is counterintuitive to the objective of this study. Accordingly, given the sparse research in this area of governance and the direct contradiction to the objective of the study, structural AI mechanisms will not be

explored further. Rather, another form of governance mechanism will be employed. This brings the discussion to the next form of governance mechanism that is available, namely procedural governance mechanisms.

5.3.2 Procedural Governance Mechanisms

Procedural governance mechanisms refer to the creation of processes to ensure that AI systems operate correctly and efficiently (Schneider et al., 2020). It is entrenched in the notion that AI systems developed meet legal requirements, company requirements and policies with respect to explainability, fairness, accountability, security, and safety.

Given that this study's objective is to create an AI governance framework that is built on governance processes, this mechanism is perfectly suited to be adopted within this study. In addition, there is an abundance of research in this area. Thus, the study will employ aspects from this governance mechanism.

Included within the procedural governance mechanism approach are the following aspects: (iii) processes; (iv) procedures; (v) performance measurement; (vi) compliance monitoring; and (vii) issue management (Abraham et al., 2019). Each of these elements will be discussed in detail below.

Processes and procedures include the creation of standardized, documented methods and steps to follow to accomplish a specific task through use of AI (Zhang et al., 2020). The governance framework should include standardized points including different steps to follow to reach a specific outcome. This will form the basis of the creation of the AI governance framework, and this is noted as a core AI principle to develop the governance framework. The next aspect in the governance mechanism is compliance monitoring (Abraham et al., 2019). Compliance monitoring includes enforcing the conformity with any regulatory requirements, such as general data protection regulation (GDPR), the Protection of Personal Information Act (POPIA), or organizational requirements (Brundage et al., 2020). The AI governance framework to be created should include aspects of compliance monitoring when necessary and thus this is included as a core AI principle. Performance management is also included within this mechanism (Abraham et al., 2019). The AI governance framework to be created should include aspects to understand the performance management of an AI system. Ongoing health and performance checks of an AI system is an important aspect to help understand the functioning of an AI system. Thus, as a core AI principle, performance management is included. The last aspect is issue management, which refers to the identification and management of any AI issues (Abraham et al., 2019). The AI governance framework to be developed should include the notion of issue management as a core principle. This will be embedded in the procedures and processes created.

5.3.3 Relational Governance Mechanism

Relational governance mechanism is the facilitates the collaboration between stakeholders. This includes important aspects of training and communication. Training employees on AI is critical as stated Schneider et al. (2020). This training can take various forms of either training employees on how to use new AI technologies or it could take the form of training re-training employees who have been replaced or augmented by AI technology (Schneider et al., 2020). Communication is also another factor that is critical in relational governance. Effective communication allows employees fears regarding AI to be reduced. Thus, these aspects of communication and training should be embedded within an AI governance framework.

To sum up, the core AI principles that should inform the AI governance framework include aspects of:

- Processes and procedures (that is, creation of standardized, documented methods and steps to follow to accomplish a specific task using AI)
- performance measurement and, more specifically, ongoing health and performance checks of an AI system
- compliance monitoring, which relates to enforcing the conformity with any regulatory requirements and
- issue management, which relates to identification and management of any AI issues.

The discussion will now delve into overarching themes that have emerged when identifying AI governance and ethical concerns.

5.4 Overarching Ethical Concerns

According to Mittelstadt et al. (2016), when dealing with AI governance, there are different themes that may arise. Mittelstadt et al. (2016) separate these themes into two categories, namely epistemic and normative. Epistemic concerns are those which are inherent within the knowledge of AI, whereas normative concerns stem from the use of AI. Epistemic concerns, according to Mittelstadt et al. (2016), has two sub-categories, namely, inconclusive evidence, and inscrutable evidence. Normative concerns have two sub-categories, namely, unfair outcomes, and traceability. These different themes will influence the way in which AI governance is managed within an entity. The discussion will now delve deeper into each of these categories to provide insight into how this could impact the AI governance framework from a core principles perspective.

5.4.1 Epistemic Concerns

Inconclusive Evidence

The first theme identified is inconclusive evidence. Inconclusive evidence relates to conclusions drawn by algorithms using inferential statistics and/or ML techniques, which produce probable yet unavoidably uncertain knowledge (Mittelstadt et al., 2016). Recognizing this limitation is vital as one should always consider the risk of being incorrect and its relation to one's responsibilities (Miller & Record, 2013). Thus, a core AI principle that should be addressed in an AI governance framework is that AI systems can produce knowledge that is uncertain.

Inscrutable Evidence

The next theme is instructible evidence. It is reasonable to expect that when data is used as an input, there is a correlation between the data and the conclusions drawn, and that these correlations are accessible to scrutiny or critique (Mittelstadt et al., 2016). This is, however, not the case with ML, where there is a lack of knowledge as to the data points used and how these are interpreted (Mittelstadt et al., 2016). This creates practical and principles limitations (Mittelstadt et al., 2016). Thus, as a core AI principle, managing and understanding input data is essential, including monitoring the data on an ongoing basis, which must be prevalent in an AI governance framework.

5.4.2 Normative Concerns

Unfair Outcomes

Algorithms that produce well-argued, conclusive evidence could still be regarded as ethically inappropriate as the actions taken could be discriminatory (Mittelstadt et al., 2016). Moreover, the mere use of AI technology could result in unfair outcomes. This is a key aspect within the AI governance framework. Thus, as a core AI governance principle, the consequences of creating an AI system must be understood.

Traceability

The harm that could be caused by algorithms is difficult to detect and the cause is not easy to find. It is even more difficult to detect who should be held accountable (Mittelstadt et al., 2016). Accountability as an overarching theme is noted by

Mittelstadt et al. (2016). Therefore, as a core AI governance principle, accountability should be entrenched within the AI governance framework.

To sum up the above discussion, when developing the AI governance framework, it should be noted that AI can produce knowledge that is uncertain. In addition, AI systems inherently create practical and principles limitations due to uncertainty of input and output of data. It then coincides that input data is a key aspect in any AI solution. Moreover, the consequences of creating an AI system must be understood and, lastly, accountability should be entrenched within the governance framework. Reasons for creating AI governance frameworks will now be addressed in further detail.

5.5 Reasons for Creating AI Governance Frameworks

There are various reasons as to why a business would need or want to create an AI governance framework. Literature has identified four typologies for the creation of a governance framework (Schiff et al., 2020). These typologies, which will be described below, include social responsibility, competitive advantage, strategic planning, and strategic intervention.

5.5.1 Social Responsibility

The first reason for creating a governance framework is the motive of social responsibility (Schiff et al., 2020). This involves enhancing social benefit and removing harm. Many groups have created governance frameworks with this aspect in mind. These include the IEEE's Ethically Aligned Design, multi-stakeholder document, and the OECD principles. When creating the AI governance framework in this study it will be important to include aspects that promote social responsibility as this is one of the current motivations for creating AI governance frameworks globally. Thus, social responsibility will be noted as a core AI principle.

Competitive Advantage

The second reason is competitive advantage (Martinho-Truswell et al., 2018). This can take the form of economic and political advantage. This is prevalent in *China's New Generation of Artificial Intelligence Development Plan*, which describes AI as the new focus of international competition (State Council of China, 2017). Accordingly, business may create AI systems to gain a competitive advantage. It will be important to integrate competitive advantage within the AI governance framework. This could be achieved by not limiting innovation but rather promoting innovation with a social conscience. Thus, promoting innovation with a social conscience will be noted as a core AI principle.

Strategic Intervention

The next motivation for creating a governance framework includes strategic intervention (Schiff et al., 2020). This is based on the external environment of the entity. This lies in the notion that an entity would want to intervene in the environment, including legal, social, or socio-economic realms in which the business finds itself (Microsoft, 2018). For example, organizations may develop voluntary ethical AI frameworks to pre-empt regulations, thereby avoiding more restrictive laws being passed. Within this study, the AI governance framework should take into consideration strategic intervention. This will be in the form of the AI governance framework addressing, for example, socio-economic concerns and other concerns relating to the immediate environment in which the entity operates. Promoting strategic intervention with a social conscience will be noted as a core AI principle.

Motivation for Signalling Leadership

The last reason is signalling leadership (Schiff et al., 2020). Signalling leadership is an important aspect for any entity and this could be a driving force behind creating an AI governance framework. This would drive any entity to be a leader in the specific field. This typology is orthogonal to the ones above. Thus, the signalling leadership reason will inherently be embedded within the governance framework, once created.

The next section to be described is the discussion on the four layers to create an integrated AI governance framework.

5.6 Four Layers to Create an Integrated AI Governance Framework

There is limited literature regarding the creation of AI governance frameworks. However, from the literature that is available, it is noted that there are four layers in creating AI governance frameworks. The four layers include: (1) AI technology, services, and applications layer; (2) AI challenges layer; (3) AI regulation process layer; and (4) Collaborative AI governance layer (Wirtz et al., 2020). Each of these layers will be discussed in detail below.

5.6.1 AI Applications/Services and Technology Layer

This layer entails the gathering and processing of data to reach a specific result. Bataller and Harris (2016) have proposed that, as part of this layer within an AI governance framework, three aspects need to be considered, including identifying,

understanding, and actioning data to reach a conclusion. Identifying data would involve a process of sensing, which involves collecting data from the environment. This environment could include already known data sets or other sources, such as cameras, tactile sensors, microphones, etc. (Bataller & Harris, 2016). After receiving the data, the data needs to be further processed for understanding to take place. The algorithm needs to gather information to create a virtual knowledge base. This knowledge base will be analysed for patterns and other correlations (Bataller & Harris, 2016). This then leads to an actioning phase. The actions phase could take various forms, including, for example, the machine learning from the data or even humans actioning the outcome instructions from the algorithm (Bataller & Harris, 2016). Challenges to business may occur at any of these steps and thus developing an AI governance framework which governs each of these aspects (that is, identifying, understanding, and actioning) is essential (Wirtz et al., 2020). Therefore, AI applications/services and technology layer intervention will be noted as a core AI principle.

AI Challenges Layer

The next step in the fours layers is the AI challenges layer. The AI challenges layer consists of two aspects: AI law and regulation; and AI society.

AI Law and Regulation

AI law and regulation refers to standards, norms, and legislation that are established for various technologies. Some of the key aspects that fall under AI law and regulation include the governance of autonomous intelligence systems, responsibility and accountability, and privacy and safety (IEEE, 2017). Governance of autonomous intelligence systems refers to the black box effect (that is, the system takes decisions based on unknown information and does so without human intervention) (Bleicher, 2016). This could take many forms, including, but not limited to, autonomous cars making decisions when used on the road, or autonomous weapons being deployed (Heyns, 2014). Any governance framework that is developed should address the black box effect to eliminate actions that are taken by autonomous AI. The second point under AI law and regulation is responsibility and accountability. This is a key aspect that must come across in any AI governance framework. It addresses the point as to who will be held legally and otherwise responsible for the actions of an AI system (Helbing et al., 2017). Due to self-learning embedded within the AI systems, it could become a tricky situation to identify who will be held responsible. The last point under AI law and regulation is privacy and safety. This deals with securing human rights and individual data to unauthorised access, for example, accessing the location of the user via an application. Without explicit consent, the data obtained endangers the privacy of the individual (Coles, 2018).

Governance frameworks should include this aspect of privacy and safety within their ambit.

AI Society

The next aspect under the AI challenges layer is AI society. AI has shaped many different areas of daily life, including aspects such as transportation, education, surveillance, and public safety (Stone et al., 2016). Using these technologies, there is a concern that automation through AI could have far-reaching consequences for society (McGinnis, 2010; Scherer, 2016). When creating an AI governance framework, scholars have identified five aspects that should be included from an AI society perspective, including: workforce transformation, social acceptance, human interaction with machines, moral dilemmas, and rulemaking for humans.

In regard to workforce transformation, AI can have a huge impact on jobs, for example, in a study by Frey and Osborne (2017), which reviewed over 700 jobs it was noted that AI could replace 47% of the jobs. Addressing the impact of jobs within an AI governance framework will be key to ensure that AI does not impact society in a negative manner. The next aspect to discuss is social acceptance. For the use of AI to flourish and have a positive impact on society, social acceptance should be in place (Scherer, 2016). This can be achieved by business using governance frameworks that promote the beneficial use of AI (Scherer, 2016). The next aspect under AI society is human interaction with machines. Interaction with machines takes place every day, for example, a computer making decisions and humans acting on those decisions, or something more menial, such as speaking to an assistant like Siri or Google assistance. This adds a blur between human and machine as you are not easily able to distinguish between them. This aspect of the blur created between human and machine would need to be addressed in the AI governance framework. The next aspect to be discussed under AI society is moral dilemmas; moral dilemmas can occur when a machine must decide between two different options, with both having conflicting moral and ethical values. Rules can be written within the AI system to take a certain action but there is no certainty that those rules will remain the same once the system learns (Lin et al., 2008). An AI governance framework would need to address such aspects, as following the written rules at all costs could also negatively impact the outcome of the system. The last aspect for AI society is AI rulemaking for humans. There is no emotion or consciousness within an AI system, which is good for an AI system to reach a certain goal, but it may result in unintended consequences for humans. These unintended consequences would need to be addressed as part of the AI governance framework (Banerjee et al., 2017).

Based on the discussion above, the AI governance framework should include aspects of job transformation to eliminate negative effects. Regarding the blur between human and machine, the AI governance framework must address who is responsible and accountable for various actions undertaken. To overcome moral dilemmas, the AI governance framework should continuously monitor the various outcomes of the system and adjust the results accordingly, and lastly, unintended

consequences for humans should be addressed in the framework by undertaking a full risk assessment before the AI solution is implemented. These aspects must be included in an AI governance framework to ensure AI society is taken into consideration.

AI Regulation and Process Layer

The next layer in this four-step layer process is the AI regulation and process layer. Like König et al. (2010), the regulatory process proposed by Wirtz et al. (2020) comprises the concepts of framing, risk and benefit assessment, risk evaluation, and risk management. All these concepts will be discussed in detail below. The first principle in the AI regulation and process layer recommends framing (Wirtz et al., 2020). In this step, stakeholders need to be consulted to understand the problem and define the way in which the problem can be overcome. Through this process, risk, benefits, and costs must be taken into consideration.

The next principle in the AI regulation and process layer is evaluation of risk. Experts would need to collect data to perform an assessment of risk (Wirtz et al., 2020). This assessment of risk will guide the AI process. The risks and benefits need to be evaluated, including understanding the impact on various parties.

The next principle is risk management. This step takes all gathered information, processes the information, and chooses the way forward (Wirtz et al., 2020). The process is then evaluated on an ongoing basis to understand if any changes need to occur. The process should also consider the short to medium term, as consulting with stakeholders can be a long, drawn-out process (Wirtz et al., 2020).

Within the AI governance framework, the following core principles must be addressed for AI regulation and process, Firstly, various stakeholders must be included, through a consultative process, in identifying the AI problem and the manner in which the problem can be overcome. Next, the risks of AI should be evaluated, including the data collected. The risks and benefits should be evaluated to understand the impact on various parties. Lastly, risk management and ongoing monitoring should be embedded in the AI governance framework.

6 Conclusion

As AI is more widely used in business there are concerns that the corporate governance principles surrounding these technologies are not developed or implemented within business. Business is continuously using these technologies to make crucial business decisions that impact all stakeholders. The call to develop governance principles by stakeholder groups are more prevalent. The study identifies eight broad themes with various overarching corporate governance principles that need to be prevalent in an AI governance framework for South Africa as summarised in Fig. 2. These include; (1) Principle concerns, (2) Procedural governance

Fig. 2 Core AI Governance Principles Informing the Development of the AI Governance Framework. Source: Own deduction

mechanisms, (3) Overarching ethical concerns, (4) Reasons for creating AI governance frameworks, (5) AI applications and technology layer, (6) AI law regulation, (7) AI Society, (8) AI regulation and process layer. It is essential that business start considering these themes when developing an AI governance framework that will be implemented. Using the outcome of this study could lead to a self-regulated AI governance framework similar to one of the King IV™. There are several limitations also prevalent in the study. The study employs a documentary analysis approach and thus inherently there could be academic articles that were not considered in the study. In addition, the study is limited t o a certain time frame till 2022 and new studies post 2022 will not be considered.

References

Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*(1), 424–438.

Banerjee, S., Singh, P. K., & Bajpai, J. (2017). A comparative study on decision-making capability between human and artificial intelligence. In B. K. Panigrahi, M. N. Hoda, V. Sharma, & S. Goel (Eds.), *Advances in intelligent systems and computing. Nature inspired computing* (pp. 203–210). Springer.

Bataller, C., & Harris, J. (2016). *Turning artificial intelligence into business value, today*. Retrieved from https://www.accenture.com/t20160814T215045w/us-en/canmedia/Accenture/

Conversion-Assets/DotCom/Documents/Global/PDF/Technology_11/Accenture-Turning-ArtificialIntelligenceinto-Business-Value.pdf.
Beauchamp, T. L. (1995). Principalism and its alleged competitors. *Kennedy Institute of Ethics Journal, 5*(3), 181–198.
Binns, R. (2017). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research, 81*, 1–11. Retrieved from SSRN: https://ssrn.com/Abstract=3086546
Bleicher, A. (2016). *Demystifying the black box that is AI: Humans are increasingly entrusting our security, health and safety to "black box" intelligent machines*. Scientific American. Retrieved from https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/
Boddington, P. (2017). *Towards a code of ethics for Artificial Intelligence*. Springer.
Borgman, H., Heier, H., Bahli, B., & Boekamp, T. (2016. January). Dotting the I and crossing (out) the T in IT governance: New challenges for information governance. *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 4901–4909). IEEE.
Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... Anderlijung, M. (2020). *Toward trustworthy AI development: Mechanisms for supporting verifiable claims*. Retrieved from: https://arxiv.org/pdf/2004.07213.pdf.
Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? *The RUSI Journal, 164*(5–6), 88–96. https://doi.org/10.1080/03071847.2019.1694260
Clouser, K. D., & Gert, B. (1990). A critique of principalism. *The Journal of Medicine and Philosophy, 15*(2), 219–236.
Coles, T. (2018). How GDPR requirements affect AI and data collection. *ITPro Today*. Retrieved from https://www.itprotoday.com/compliance/how-gdpr-requirements-affect-ai-and-data-collection.
Corbin, J., & Strauss, A. (2008). *Qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). SAGE.
Crawford, K., Dobbe, R., West, S., Kak, A., Sánchez, A., Green, B., Raji, D., Kaziunas, E., McElroy, E., Fried, G., Schultz, J., Rankin, J., Whittaker, M., Richardson, R., & Dryer, T. (2019). *AI Now Report*. Retrieved from https://ainowinstitute.org.
Dancy, J. (2004). *Ethics without principles*. Clarendon Press.
Deloitte. (2018). *Artificial intelligence defined*. Retrieved from https://www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html
Edelman. (2019). *Edelman trust barometer global report*. Retrieved from Edelman: https://www.edelman.com/sites/g/files/aatuss191/files/2019-02/2019_Edelman _Trust_Barometer_Global_Report.pdf.
Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80–92.
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI.* Berkman Klein Center Research Publication No. 2020–1, Retrieved from SSRN: https://ssrn.com/Abstract=3518482.
Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019
Haenlein, M., & Kaplan, A. (2019). A brief history of Artificial Intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14. https://doi.org/10.1177/0008125619864925
Haes, S. D., & Grembergen, W. V. (2015). *Enterprise governance of information technology: Achieving alignment and value, featuring COBIT 5*. Springer.
Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoeven, J., Zicari, R., & Zwitter, A. (2017). *Will democracy survive big data and artificial intelligence?*

Scientific American. Retrieved from https://www.scientificamerican.com/article/will-democracy-survive-bigdata-and-artificial-intelligence/Hickok

Heyns, C. (2014). Report of the special rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns. *Human Rights Council of the United Nations General Assembly*. Retrieved from https://digitallibrary.un.org/record/771922/files/A_HRC _26_36-EN.pdf.

Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics, 1*(1), 41–47.

Ho, C. W. L., Soon, D., Caals, K., & Kapur, J. (2019). Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clinical Radiology, 74*(5), 329–337.

IEEE. (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. Version 2. *IEEE*. Retrieved from https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead_v2.pdf

IoDSA. (2022). *Institute of directors*. Retrieved from https://www.iodsa.co.za/page/king-iv

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

König, A., Kuiper, H. A., Marvin, H. J. P., Boon, P. E., Busk, L., Cnudde, F., et al. (2010). The Safe Foods framework for improved risk analysis of foods. *Food Control, 21*(12), 1566–1587. https://doi.org/10.1016/j.foodcont.2010.02.012

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

Lin, P., Bekey, G., & Abney, K. (2008). *Autonomous military robotics: Risk, ethics, and design*. California Polytechnic State University.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics,* 1–7.

Martinho-Truswell, E., Miller, H., Asare, I.N., Petheram, A., Stirling, R., Mont, C. G., & Martinez, C. (2018). *Mexico: Towards an AI strategy in Mexico: Harnessing the AI revolution. Technical Report*. Mexico City, Mexico: British Embassy in Mexico, Oxford Insights, & C Minds.

McGinnis, J. O. (2010). Accelerating AI. *Northwestern University Law Review, 104*(3), 1253–1269. https://doi.org/10.2139/ssrn.1593851

Microsoft. (2018). *The future computer: Artificial Intelligence and its role in society*. Microsoft.

Microsoft State Council of China. (2017). *China's new generation of artificial intelligence development plan*. Technical report 35. State Council of China, Beijing, China.

Miller, B., & Record, I. (2013). Justified belief in a digital age: On the epistemic implications of secret Internet technologies. *Episteme, 10*(2), 117–134.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 2053951716679679.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From what to how: An initial review of publicly available AIethics tools, methods and research to translate principles into practices. In L. Floridi (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence. Philosophical Studies Series 144*. Springer. https://doi.org/10.1007/978-3-030-81907-110

Price, N. W., II. (2017). Regulating black-box medicine. *Michigan Law Review, 116*, 421.

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association, 27*(3), 491–497.

Scherer, M. U. (2016). Regulating artificial intelligence systems: Risk, challenges, competencies, and strategies. *Harvard Journal of Law & Technology, 29*(2), 354–400. https://doi.org/10.2139/ssrn.2609777

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–158.

Schneider, J., Abraham, R.. & Meske, C. (2020). *AI Governance for Businesses*. Retrieved from arXivpreprint arXiv:2011.10672.

Song, M., Yang, Y., He, J., Yang, Z., Yu, S., Xie, Q., et al. (2018). Prognostication of chronic disorders of consciousness using brain functional networks and clinical characteristics. *eLife, 7*, e36173.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... Teller, A. (2016). *Artificial intelligence and life in 2030: One hundred year study on artificial intelligence: Report of the 2015–2016 Study Panel*.

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI*. Ethics, and Society, 195–200.

Winfield, F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences, 376*(2133), 1–13.

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration, 43*(9), 818–829.

Zhang, J. M., Harman, M., Ma, L., & Liu, Y. (2020). Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering, 48*(1), 1–36.