Artificial intelligence governance: Ethical considerations and implications for social responsibility

By Mark Anthony Camilleri[1][2][3]

Suggested Citation: Camilleri, M.A. (2023). Artificial intelligence governance: Ethical considerations and implications for social responsibility, *Expert Systems*, https://doi.org/10.1111/exsy.13406

Abstract

A number of articles are increasingly raising awareness on the different uses of artificial intelligence (AI) technologies for customers and businesses. Many authors discuss about their benefits and possible challenges. However, for the time being, there is still limited research focused on AI principles and regulatory guidelines for the developers of expert systems like machine learning (ML) and/or deep learning (DL) technologies. This research addresses this knowledge gap in the academic literature. The objectives of this contribution are threefold: (i) It describes AI governance frameworks that were put forward by technology conglomerates, policy makers and by intergovernmental organizations, (ii) It sheds light on the extant literature on "AI governance" as well as on the intersection of "AI" and "corporate social responsibility" (CSR), (iii) It identifies key dimensions of AI governance, and elaborates about the promotion of accountability and transparency; explainability, interpretability and reproducibility; fairness and inclusiveness; privacy and safety of end users, as well as on the prevention of risks and of cyber security issues from AI systems. This research implies that all those who are involved in the research, development and maintenance of AI systems, have social and ethical responsibilities to bear toward their consumers as well as to other stakeholders in society.

Keywords: AI governance; AI explainability; AI fairness; AI accountability; AI transparency; AI risks.

[1] Department of Corporate Communication, Faculty of Media and Knowledge Sciences, University of Malta, Malta. Mark.A.Camilleri@um.edu.mt

[2] Medill School of Journalism, Media and Integrated Marketing Communications, Northwestern University, Evanston, Illinois, United States of America.

[3] The Business School, University of Edinburgh, Edinburgh, Scotland, United Kingdom.

1. Introduction

Artificial intelligence (AI) is related to those technologies that simulate human intelligence, as they can emulate decision-making processes and behaviors. Most of them can resolve complicated tasks in an independent manner or with minimal interventions (LeCun et al., 2015; Zhang & Lu, 2021; Zhang et al., 2023). AI is concerned with expert systems that rely on natural language processing (Carvalho et al., 2019), speech recognition (Narwani et al., 2022) and/or machine vision (Silva et al., 2022) to continuously learn through the acquisition of new data (Berente et al., 2021).

The benefits of AI are already being felt across a wide range of businesses (Dwivedi et al., 2021). Various researchers already confirmed that AI applications can automate repetitive tasks including data entry, invoice processing, online customer services, among others (Ribeiro et al., 2021). These expert systems are characterized by their quick data analytical capabilities as they can optimize workflows in different contexts, make complex decisions faster and more accurately than humans, leading to increased efficiencies and productivity levels in various industries (Javaid et al., 2021; Ng et al., 2021; Wamba-Taguimdje et al., 2020).

AI-powered chatbots and virtual assistants can provide customer centered personalized recommendations round the clock (24/7) (Camilleri & Troise, 2023; Selamat & Windasari, 2021). Today's businesses can obtain deep insights from the data they gather through online interactions with customers and employees. Some of them are utilizing natural language processing technologies that are capable of understanding the languages and jargons used in different businesses and industries (Wu et al., 2022). Others rely on AI expert systems to extract information from complex documents and data, automate business processes and workflows, drive effective and accurate decisions in a flexible manner on premises and across a hybrid cloud (Sachan et al., 2020; Weber et al., 2022).

Hence, employees can dedicate more time to higher value work. For example, IBM Watson services clients within different service industries (Magistretti et al., 2019; Strickland, 2019). IBM's AI solutions provide personalized responses to customer inquiries. Its customers include Lufthansa, GlaxoSmithKline (GSK) and Ernst Young (EY), among others.

Currently, there are a number of academic articles that describe the use of AI for business applications (Janiesch et al., 2021; Matytsin et al., 2023; Minkkinen et al., 2022; Mullins et al., 2021; Pai & Chandra, 2022; Raisch & Krakowski, 2021). Most of them have even outlined their strengths and weaknesses (Dauvergne, 2022; Huang & Rust, 2020). Very often researchers discuss about how the advancements of AI are raising serious concerns among the businesses themselves and their stakeholders including the governments, academia and civil societies, regarding the risk of possible harm associated with the use intelligent, learning technologies (Galaz et al., 2021; John-Mathews et al., 2022).

Recently, during a United States (US) Senate hearing, OpenAI CEO Sam Altman, one of the developers behind ChatGPT, raised awareness about the opportunities and challenges of using AI. Mr Altman also warned senators that it could spread disinformation, influence people and even interfere with elections, among other perils. Hence, he urged policymakers to enact regulation for AI governance (CNBC, 2023). A few commentators argue that AI is not always deployed in a responsible manner, and/or is not managed properly (Butcher & Beridze, 2019; Erdélyi & Goldsmith, 2022; McBride et al., 2022; Minkkinen et al. 2023).

This research raises awareness on the importance of AI governance in an age where more individuals and organizations are utilizing AI systems for different applications. Today, online users can easily access conversational technologies like generative pre-trained transformers (GPT). Some businesses are already availing themselves from facial recognition technologies. Arguably, these disruptive AI technologies may be used in an irresponsible manner and/or for malicious purposes. Hence, their adoption could raise serious concerns of different stakeholders in society. Various governments and international organizations are stepping in with their commitment to protect their citizens and the businesses' interests. As a result, several regulatory authorities are outlining governance principles and guidelines that are intended to support practitioners in the development of AI, ML and DL technologies, with the aim to mitigate and reduce the risks associated with them. AI governance is intended to minimize risks including the violations of privacy, misuse of personal information, bias, discrimination, and the like.

For the time being, there are limited contributions that are focused on AI governance frameworks that provide substantive (outcome-based) and reflexive (process-based) guidelines to practitioners who are developing AI innovations. This research addresses this knowledge gap. Specifically, its objectives are threefold: (i) To shed light on the latest developments in terms of regulatory instruments, rules and principles on AI governance that apply to practitioners who are creating, testing and implementing AI models; (ii) To describe the findings from a rigorous review of high impact articles focused on "AI governance" and on the intersection of "AI" and "corporate social responsibility" (CSR), and (iii) To raise awareness about the importance and timeliness of formalizing responsible AI governance protocols to ensure that ML and DL systems are reliable, dependable and safe for business and society at large. This contribution puts forward an AI governance framework that is intended to promote accountable, transparent, explainable interpretable reproducible, fair, inclusive and secure AI solutions. It clarifies the meanings of these essential elements of AI governance that are meant to prevent unnecessary risks and occurrences from affecting any parties. In conclusion, it discusses managerial implications for AI practitioners and policymakers.

The following section describes different governance frameworks and regulatory guidelines focused on responsible AI, ML and DL technologies. Then, the methodology section clarifies how the data is captured from high impact sources. It explains that the researcher relied on a rigorous systematic review of articles about AI governance. Afterwards, it identifies different aspects of AI governance and presents a discursive argumentation on the best practices that are intended for AI practitioners and for the developers of autonomous learning technologies. In conclusion, it presents future research avenues.

2. Background

Many companies are increasingly relying on AI algorithms, prior to making strategic decisions (Janiesch et al., 2021; Rąb-Kettler & Lehnervp, 2019). The automated technologies are helping them in their organizations' performance. AI innovations can interact with online users through two-way communications (Camilleri & Troise, 2023). Their dialogue formats enable them to respond to questions (Thorp, 2023), to admit their mistakes (Barrot, 2023), and to even reject requests (Crawford & Paglen, 2021), if they are not recognized as appropriate.

Several companies are using ML/DL algorithms for business process automation (BPA), fraud prevention, malware detection, spam filtering, as well as for the predictive maintenance of recommender systems, among other purposes (Engel et al., 2022; Romao et al. 2020). Such technologies are also useful for customer relationship management (CRM) systems as they can scrutinize email content and prompt business practitioners to respond to the most important messages.

Advanced systems are equipped to provide fast and effective responses to customers. Other ML/DL applications are related to business intelligence (BI) and analytics, as algorithms can be used to identify important information in datasets, and reveal patterns, trends, cycles and anomalies from the big data as well as from small data (Carvalho et al., 2019). ML/DL may also be used in human resources information databases to identify the best candidates for an open position, and for other business purposes.

DL algorithms enable computers and their artificial neural networks to collect and process data like a human brain. They can complex patterns in texts, images, audio and video, and can provide reliable insights and predictions into the future (Buhmann & Fieseler, 2023; Lin & Huang, 2020). The deep-learning architectures include deep belief networks, deep neural networks, deep reinforcement learning, convolutional neural networks, recurrent neural networks, and transformers are applied in various fields including for bioinformatics, computer vision, machine translation, material inspection, natural language processing, and speech recognition, among other areas (LeCun et al., 2015). Frequently, DL algorithms are yielding significant results that are similar to (and in some cases, are even surpassing) the human experts' performance.

Such developments call for AI governance (Erdélyi & Goldsmith, 2022; Filgueiras, 2022; Gonzalez et al., 2020; Mullins et al., 2021; Papagiannidis et al., 2023) and corporate responsibility (Dauvergne, 2022; Du & Xie, 2021; Frank, 2021, Li & Li, 2021) to ensure that humanity can easily access and benefit from AI systems, in a protected, safe and secure environment. Every person who is involved in the advancement of AI ought to be accountable for their innovations (Raji et al., 2020). AI models are incrementally improving the accuracy of their algorithms and are offering human-like performance for service businesses in areas like automated decision-making systems and recommender systems, among other solutions.

However, in reality, there are still a number of challenges for business practitioners to keep up the pace with the latest developments in the real world and with the proliferation of rich data that could/could not be uploaded into AI systems (for various reasons). What is/was learned by AI systems may quickly become obsolete or outdated and could no longer be applicable in the near future (Thorp, 2023). For example, when ChatGPT is asked about the reliability of its responses, it answered in the following manner:

"AI language model, ChatGPT strives to provide accurate and helpful responses based on the input it received. However, it's important to note that my responses are generated algorithmically, and therefore, there may be instances where my responses are not completely accurate or may not fully address the question being asked. Additionally, my responses are based on the data available to me at the time of my training and knowledge cutoff date, so any new developments or information that have emerged since then, may not be reflected in my responses. It's always a good idea to verify information obtained from any source, including AI language models like ChatGPT, with other credible sources to ensure its accuracy and reliability" (ChatGPT, 20th March 2023).

When it was questioned about its cutoff date, ChatGPT indicated that any events, developments or information that occurred after September 2021 were not reflected in its responses. Its automated decisions may be incorrect, or not factual, as their algorithms are not always up to date.

For the time being, practitioners can avail themselves of certain technologies to help implement guardrails at every stage of the AI lifecycle. They could include preventative security policies and practices. They may track all activities of AI models, check that relevant controls are in place, during data collection and instrumental processing, to minimize risks to their customers (like service businesses operating call centers) (Galaz et al., 2021; Hickok, 2022). They can also articulate ethical principles and introduce transparent reporting to share information on their AI governance and about responsible ML/DL initiatives (Minkkinen et al., 2022; Sætra, 2021).

Renieris et al. (2022) indicated that organizations should tie their responsible AI efforts to their CSR strategies. They implied that core ideas behind responsible AI, such as bias prevention, transparency and fairness, are already aligned with fundamental principles of CSR. For example, the International Standards Organization's social responsibility standard (ISO 26000) commends that organizations ought to be accountable and transparent to their stakeholders. Its non-binding principles encourage them to engage in ethical behaviors, respect the rule of law, respect international norms of behavior and to respect human rights (Camilleri, 2019). This argumentation is also related to the organizations' social license to operate paradigm (Camilleri, 2017), as they are expected to justify corporate decisions and behaviors including about responsible AI governance with stakeholders including policy makers, among others.

Table 1 features a summary of the most popular AI principles and guidelines that are meant to support practitioners who are developing, testing and using AI designs and applications.

Table 1. Regulatory principles and guidelines for artificial intelligence governance

Institution / Organization / Business	Entity	Responsibility dimensions
Policymakers	European Union (EU)'s Artificial Intelligence Act	Accuracy; Clear and adequate information; Detailed documentation; High quality datasets that reduce risks and discrimination; Human oversight measures; Logging of activities to trace any tampering of data; Robustness; Security.
	Singaporean government's National AI Strategy	Explainable; Fair; Reproducibility; Robustness; Transparent.
	United States' AI Bill of Rights	Algorithmic discrimination protection; Data privacy; Human alternatives consideration and fallback; Notice and explanation; Safe and effective systems;
Organization	Institute of Electrical and Electronics Engineers (IEEE)'s AI Ethics and Governance Standards	Addressing ethical issues during design; Child-friendly digital services framework; Ongoing evaluations on the impacts of automated systems on human well-being; Data privacy process; Ontological standards for ethically-driven automation systems and robotics; Transparency of autonomous systems; Transparent employer data governance.
	OECD's AI Principles	Accountability, transparency and explainability; Fairness and human-centered values; Inclusive growth, sustainable development and well-being of humans; Robustness, safety and security.
Businesses	Microsoft's Responsible AI (Principles)	Accountability and transparency; Fairness; Inclusiveness; Privacy, safety and security; Reliability and safety.
	IBM's AI Governance	Explainability; Fairness; Privacy; Robustness; Transparency.

2.1 EU’s Artificial Intelligence Act

The European Union (EU) put forward its proposed AI regulatory framework that is referred to “The Artificial Intelligence Act (AI Act)” in April 2021. This document introduced AI principles and a legal framework for its member states. It specifies that its objectives are: (i) to increase the safety and security of AI systems, as they have to respect relevant legislation on fundamental rights and should reflect EU values; (ii) facilitate investment in automated systems; (iii) reinforce responsible AI governance through regulations and principles; (iv) to create a trustworthy and safe eco-system for the development of AI systems.

The EU Commission developed a risk-based approach pyramid that identifies four levels of risk: (i) minimal risk, (ii) limited risk, (iii) high-risk, and (iv) unacceptable risk. It reported that end-users should be informed that they are interacting with AI, to enable them to make an informed decision as to, whether they should continue with their engagement with the machine or not.

The EU proposed that public authorities are entrusted to monitor the developments of AI products once they are launched in the market. It requests AI developers to continue appraising the quality and assurance of AI systems, and to undertake risk management assessments as they are expected to report any serious incidents and malfunctioning in them.

2.2 Singapore’s National AI strategy

On the 25 May 2022, the Singaporean government launched A.I. Verify, an AI Governance Testing Framework and Toolkit for companies that may want to prove that their AI systems are responsible and trustworthy. Google, Meta and Microsoft among other businesses have already adopted the Singaporean framework, to confirm their AI governance credentials. In sum, the guiding principles suggest that AI systems ought to be human centric and their modus operation should be explainable, transparent and fair.

Subsequently, the model framework integrated additional considerations like robustness (to conform with IBM's AI governance principles) and reproducibility, in order to increase its relevance and usability. Singaporean's framework also specified that AI developers and users ought to engage in interactions and communications with a wide array of stakeholders (again, this is consistent with IBM and Microsoft's transparency principles).

2.3 The AI Bill of Rights

In October 2022, American policy makers released a document that identified five principles that are meant to guide practitioners in the development and utilization of AI designs. Their "AI Bill of Rights" is a voluntary guideline that is intended to protect the interests of American citizens who will be using AI innovations. This document raises awareness on why AI's automated systems ought to be safe and effective for their users. It clarifies that AI designers, developers, and deployers have to ensure that they are committed to safeguard their users' data privacy and to protect them from algorithmic discriminations. Their automated systems are expected to explain how and when AI is being used and should provide clear information on how it works. Users ought to be in a position to opt out, when they want, and to communicate with a human customer service agent to resolve queries or to find solutions to their problems.

2.4 IEEE' AI Ethics and Governance Standards

On the 17th January 2023, IEEE introduced free access to AI Ethics and Governance standards. Currently, IEEE Standards Association (IEEE SA) provides free access to its global socio-technical standards to guide practitioners to engage in trustworthy AI innovations. The standards advocate the importance of transparency (of autonomous systems and of employer data governance) as well as of data privacy. They address ethical issues of robotic and of other AI systems. In addition, one of IEEE standards is focused on evaluating the effects of autonomous and intelligent technologies on all citizens, including children. In fact, they make reference to the United Nations Convention on the Rights of the Child.

2.5 OECD's AI Principles

As of May 2019, OECD has started raising awareness about its principles that guide practitioners in the creation of innovative and trustworthy AI systems. OECD's AI principles urge practitioners to respect human rights and democratic values, in all stages of their research and development. Its standards promote accountability, transparency and explainability; robustness, security and safety; fairness and human centered values; as well as sustainable development and inclusive growth. OECD specifies that all AI actors are expected to ensure that all of their procedures can be traced, to reduce the risks of AI systems. It implies that everyone ought to be accountable for their actions. OECD has also dedicated a standard to transparent reporting and disclosures of AI processes.

2.6 IBM's AI governance

IBM dedicated a website to explain its guiding values and governance principles related to AI applications and processes. It clarified that its foundational properties of its AI ethics rest on five pillars: (i) explainability (AI designs that deliver seamless experiences); (ii) fairness (AI designs that assist humans in making fairer choices); (iii) robustness (AI designs that are employed to make crucial decisions); (iv) transparency (AI designs that reinforce trust through disclosures); and, (v) privacy (AI designs that prioritize and safeguard consumers' privacy and data rights).

The multinational technology corporation recognized the importance of articulating governance policies based on principles, regulations and legislation, which are supporting it in its AI strategic management and operations. IBM uses technology to implement guardrails at each stage of the AI/ML lifecycle, during data collection and in its instrumenting processes. It is also transparent with its reporting of AI activities for the benefit of different stakeholders. Its AI governance framework is intended to operationalize AI with confidence through lifecycle governance, to manage risk and reputation, to strengthen regulatory compliance and to meet stakeholder demands.

2.7 Microsoft's Responsible AI

Similarly, Microsoft's AI systems provide valid solutions for the problems they are designed to solve. include capabilities that support informed human oversight and control. Its AI products are customized to be consistent with the designs ideas of practitioners and are congruent with their values and principles.

The company's corporate website suggests that its AI governance is based on responsible dimensions, including accountability, transparency, reliability and safety, privacy and security, fairness and inclusiveness. Microsoft assures its customers that it regularly evaluates operational factors of its AI systems, to determine whether they are performing reliably and safely. Its AI systems are subject to ongoing monitoring, and evaluation processes to manage and maintain extant systems, to improve them over time, troubleshoot problematic issues and to identify new uses. It methodically quantifies the risks to minimize the time to remediation of predictable or known failures and to avoid mistakes that may result in any harm to human beings.

Moreover, the technology giant confirms that is committed to protect the privacy of their users. It adds that its secure features increase the reliability of data and protect personal data from being disseminated with other users. It makes specific reference to AI security aspects including to data origin and lineage, internal and external data usage; data corruption considerations, anomaly detections, changes in the data that might indicate that there may be users who are trying to acquire data.

Microsoft's Framework for Responsible AI underlines that its systems are intended to treat diverse people in a fair manner, by reducing existing stereotypes, cultural denigration, under-representation and bias. It reported that its AI products are designed to provide a similar quality of service for various demographic segments in society including to marginalized groups in order to minimize disparities among different people. It suggests that its AI systems are inclusive as they empower everyone around the world, making sure that no one is left out. It clarifies that members of minority communities are involved in the research, development and testing of AI designs and solutions.

The technology company is accountable with its customers and partners about how its AI systems are impacting the world, in order to make informed choices. Microsoft posits that it is transparent with stakeholders as it is clear about the strengths and limitations of its AI systems. This is often referred to interpretability or intelligent-ability, as AI is in a position to generate or to manipulate content including visual, verbal or vocal communications.

3. Methodology

A systematic literature review (SLR) approach was used to scrutinize articles focused on AI governance and CSR. This rigorous methodology ensures that the findings of this research are rigorous and trustworthy, as other scholars can follow the procedures that are clearly outlined in this SLR (Camilleri et al., 2023). Therefore, they could easily replicate and validate the results reported in this paper.

The following search query was inserted through Scopus: TITLE-ABS-KEY ("artificial intelligence governance") AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019)). It sought to investigate articles published in English, through journals that included the keywords “artificial intelligence governance” in their title, abstract or keywords. The results indicated that there were thirteen (13) articles through Scopus, that featured the specified keywords. Twelve (12) of these publications were also indexed through Web of Science’s (WOS) Core Collections including in Arts & Humanities Citation Index (A&HCI), Emerging Sources Citation Index (ESCI), Science Citation Index (SCIE) and/or in Social Science Citation Index (SSCI).

The SLR reported that the most prominent (top ten) keywords that were used by researchers who investigated AI governance were Artificial Intelligence, Artificial Intelligence Governance, AI Governance, AI, AI Ethics, Artificial Intelligence Ethic, Artificial Intelligence Systems, Ethical Technology, Decision Making and AI Systems.

Table 2 sheds light on the articles published though both Scopus and WOS outlets, since 2019. It appraises the authors, outlines their research objectives and describe the methodology they used to capture the data.

Table 2. List of articles focused on artificial intelligence governance

Authors	**Date**	**Journal**	**Indexed through**	**Research objective(s)**	**Methodology**
Antonov	2022	Revista CIDOB d'Afers Internacionals	Scopus and WOS ESCI	This paper investigates AI governance in the European Union (EU) context.	Discursive
Erdélyi & Goldsmith	2022	Government Information Quarterly	Scopus and WOS SSCI	This research describes international AI governance frameworks and regulatory structures that are supporting the development of responsible AI practices.	Discursive
Filgueiras	2022	Ain Shams Engineering Journal	Scopus and WOS ESCI	The article uses an institutional theory perspective to explore the design of AI systems that affect decision-making processes in the public sector.	Review
Fosch-Villaronga et al.	2022	Computer Law & Security Review	Scopus and WOS SSCI	This paper investigates the biases of algorithmic-based AI systems in healthcare-related applications.	Discursive
Gonzalez et al.	2020	Ain Shams Engineering Journal	Scopus and WOS SCIE	This article discusses about the concepts of artificial intelligence, and governance of smart cities.	Case study
Gonzalez et al.	2022	AI and Society	Scopus and WOS ESCI	The research relies on bibliographic approach to explore the decision-making processes and policy formulation through AI systems.	Review
Hickok	2022	AI and Society	Scopus and WOS ESCI	This research explains how public entities use AI in procurement systems.	Discursive

Koniakou	2023	Information Systems Frontiers	Scopus and WOS SCIE	This research provides an overview on the developments in AI governance. The researchers elaborate on how AI developments ought to consider human rights and ethical principles.	Discursive
Minkkinen et al.	2023	Information Systems Frontiers	Scopus and WOS SCIE	This research explores the technological frames technology-centered ecosystems and responsible AI (RAI).	Document analysis
Mullins et al.	2021	Patterns	Scopus and WOS ESCI	This paper provides an overview of the use of AI in insurance applications. The authors elaborate about AI ethics in the financial services industry.	Discursive
Papagiannidis et al.	2023	Information Systems Frontiers	Scopus and WOS SCIE	This research investigates AI governance. It promotes the development of robust AI applications that are intended to mitigate their negative effects, in the context of the energy industry sector.	Qualitative (interviews)
Schneider et al.	2022	Information Systems Management	Scopus and WOS SCIE	This research explores the governance of AI programs, and machine learning systems. The researchers clarify how, who, what governs AI technologies.	Review

(Sorted in alphabetical order, as of 31st March 2023)

Table 2 clearly indicates that most articles (75%) that were captured through this review involved secondary research methodologies as they were discursive contributions and/or featured literature reviews. This finding suggests that, for the time being, there are few researchers who carried out primary research activities focused on AI governance.

Another bibliographic study (through Scopus and WOS repositories) sought to explore articles that included “artificial intelligence” and “corporate social responsibility”, as follows: TITLE-ABS-KEY ("artificial intelligence" AND "corporate social responsibility") AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019)).

In this case, the results reported that there were thirty-six (36) articles indexed in Scopus. However, fourteen (14) articles were excluded as they were not focused on AI or on corporate social responsibility (CSR) paradigms. Alternatively, the discarded publications were not published in one of WOS’s Core Collections (in addition to Scopus). Table 3 features all (22) articles on the intersection of artificial intelligence and CSR. These contributions were published through both Scopus and WOS journals, between January 2019 and March 2023.

This bibliographic study indicates that the most popular keywords on the intersection of AI and CSR were: Corporate Social Responsibility, Artificial Intelligence, Sustainability, Machine Learning, Sustainable Development, Business Ethics, Corporate Governance, Ethics, Health Care and Human Resource.

Table 3. List of articles focused on artificial intelligence governance

Authors	Date	Source title	Indexed through	Research objective(s)	Methodology
Abina et al.	2022	Sustainability	Scopus, WOS SCIE and WOS SSCI	This paper describes sustainability and leadership competency models. The researchers elaborate the use of systems that detect the individuals' digital and soft skills.	Discursive
Aitken et al.	2021	Technology in Society	Scopus and WOS SSCI	This paper investigates socially responsible data intensive innovation within the private sector.	Qualitative (Focus groups)
Alnamrouti et al.	2022	Sustainability (Switzerland)	Scopus, WOS SCIE and WOS SSCI	This study sheds light on the effects of corporate social responsibility and of organizational learning on the sustainable performance of non-governmental organizations (NGOs).	Quantitative (survey)
Broer	2022	Social Science and Medicine	Scopus and WOS SCIE and WOS SSCI	This research is focused on one of Facebook's AI program that is intended to safeguard the wellbeing of its subscribers.	Qualitative (content analysis)
Buhmann & Fieseler	2023	Business Ethics Quarterly	Scopus and WOS SSCI	This research explores how and to what extent corporate governance structures are related to ethical AI frameworks.	Discursive
Carvalho et al.	2019	Communications of the Association for Information Systems	Scopus and WOS ESCI	This research explains that IBM's Natural Language Understanding (NLU), can resolve data-analytics problems.	Sentiment analysis
Damoah	2021	Journal of Cleaner Production	Scopus, WOS SCIE AND	This research investigates the use of drones in a healthcare supply chain (HSC).	Qualitative (semi-structured interviews)

			WOS SSCI		
Dauvergne	2022	Review of International Political Economy	Scopus and WOS SSCI	This article reports that CSR disclosures are not revealing the disadvantages of AI.	Discursive
Du & Xie	2021	Journal of Business Research	Scopus and WOS SSCI	This paper evaluates ethical issues related to AI. The researchers elaborate about ethical AI practices and on socially responsible behaviors.	Conceptual (Discursive)
Du et al.	2022	Journal of Business Ethics	Scopus and WOS SSCI	This article links CSR perspectives with AI governance.	Discursive
Frank	2021	Journal of Cleaner Production	Scopus, WOS SCIE AND WOS SSCI	This study explores consumer evaluations about AI products for environmental sustainability.	Quantitative (hierarchical linear modeling)
Krkač	2019	Social Responsibility Journal	Scopus and WOS ESCI	This research discusses about AI versus Human CSR and CSI.	Discursive
Li et al.	2021	Production and Operations Management	Scopus and WOS SCIE	This research examines the effects on AI on CSR and idiosyncratic risk.	Quantitative observations
Lin & Huang	2020	Discrete Dynamics in Nature and Society	Scopus and WOS SCIE	This research investigates the use of deep learning for forecasting accuracy of financial data.	Quantitative (regression)
Magas & Kiritsis	2022	International Journal of Production Research	Scopis and SCIE	This paper outlines opportunities and challenges related to data sharing through the Industry Commons Ecosystem (ICE).	Discursive
Matytsin et al.	2023	Humanities and Social Sciences Communications	Scopus, WOS SSCI and A&HCI	The research is focused on the use of AI among enterprises.	Quantitative (regression)

McBride et al.	2022	Managerial Finance	Scopus and WOS ESCI	This paper explores AI, corporate governance and socially responsible investing options in the financial markets.	Literature review
Minkkinen et al.	2022	AI and Society	Scopus and WOS ESCI	This research examines the use of AI for ESG investment analyses.	Qualitative (semi-structured interviews)
Pai & Chandra	2022	Pacific Asia Journal of the Association for Information Systems	Scopus and WOS ESCI	This research investigates the use of AI for CSR purposes.	Quantitative
Rab-Kettler & Lehnervp	2019	Management Systems in Production Engineering	Scopus and WOS ESCI	This paper explores socioeconomic and technological changes. The researchers discuss on their implications on human resources management and on talent attraction.	Discursive
Sætra	2021	Sustainability (Switzerland)	Scopus, WOS SCIE and WOS SSCI	This research explores the environmental, social, and governance (ESG) impacts of AI.	Discursive
Saurabh et al.	2022	Journal of Information, Communication and Ethics in Society	Scopus and WOS ESCI	This research links CSR and ethics with AI-led digital transformation.	Qualitative (interviews)

(Sorted in alphabetical order, as of 31st March 2023)

Again, Table 3 confirms that most articles (41%) that were featured in this SLR exercise involved secondary research methodologies. The majority of researchers who sought to explore the link between CSR and AI, have yielded discursive, theoretical, and/or conceptual contributions. In many cases, they presented a critical review of the academic literature.

4. Artificial intelligence governance

The term “artificial intelligence governance” or “AI governance” integrates the notions of “AI” and “corporate governance”. AI governance is based on formal rules (including legislative acts and binding regulations) as well as on voluntary principles that are intended to guide practitioners in their research, development and maintenance of AI systems (Butcher & Beridze, 2019; Gonzalez et al., 2020). Essentially, it represents a regulatory framework that can support AI practitioners in their strategy formulation and in day-to-day operations (Erdélyi & Goldsmith, 2022; Mullins et al., 2021; Schneider et al., 2022). The rationale behind responsible AI governance is to ensure that automated systems including ML/DL technologies, are supporting individuals and organizations in achieving their long terms objectives, whist safeguarding the interests of all stakeholders (Corea et al., 2023; Hickok et al., 2022).

AI governance requires that the organizational leaders comply with relevant legislation, hard laws and regulations (Mäntymäki et al., 2022). Moreover, they are expected to follow ethical norms, values and standards (Koniakou, 2023). Practitioners ought to be trustworthy, diligent and accountable in how they handle their intellectual capital and other resources including their information technologies, finances as well as members of staff, in order to overcome challenges, minimize uncertainties, risks and any negative repercussions (E.g. decreased human oversight in decision making, among others) (Agbese et al., 2023; Smuha, 2019).

Procedural governance mechanisms ought to be in place to ensure that AI technologies and ML/DL models are operating in a responsible manner. Figure 1 features some of the key elements that are required for the responsible governance of artificial intelligence. The following principles are aimed to provide guidelines for the modus operandi of AI practitioners (including ML/DL developers).

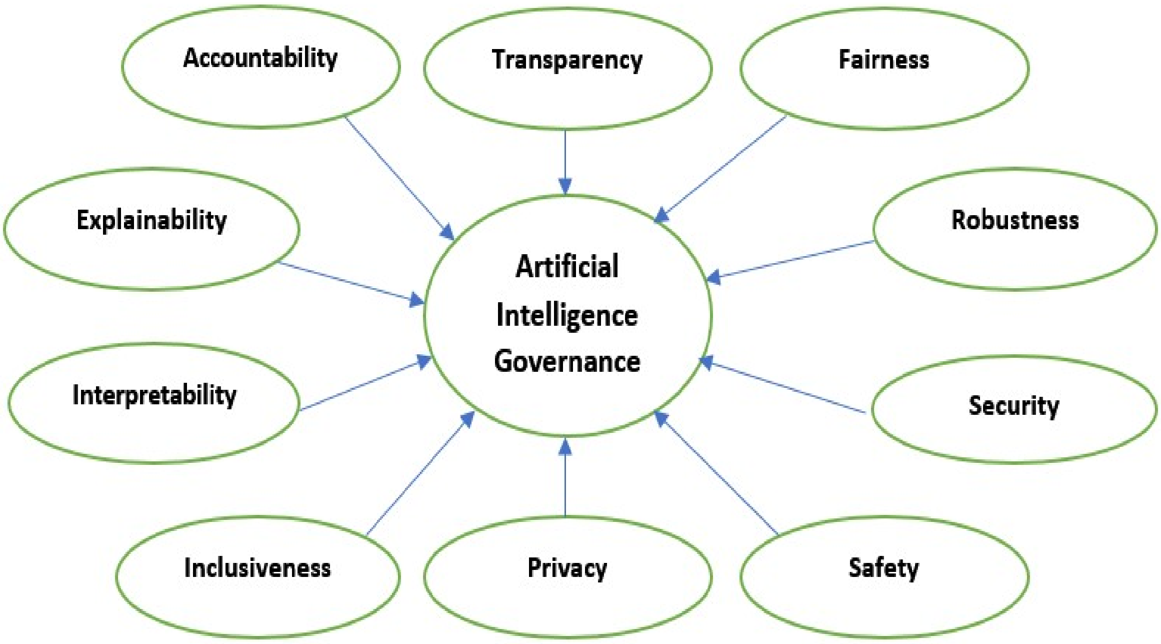

Figure 1. A Responsible Artificial Intelligence Governance Framework

4.1 Accountability and transparency

"Accountability" refers to the stakeholders' expectations about the proper functioning of AI systems, in all stages, including in the design, creation, testing or deployment, in accordance with relevant regulatory frameworks. It is imperative that AI developers are held accountable for the smooth operation of AI systems throughout their lifecycle (Raji et al., 2020). Stakeholders expect them to be accountable by keeping a track record of their AI development processes (Mäntymäki et al., 2022).

The transparency notion refers to the extent to which end-users could be in a position to understand how AI systems work (Andrada et al., 2020; Hollanek, 2020). AI transparency is associated with the degree of comprehension about algorithmic models in terms of "simulatability" (an understanding of

AI functioning), “decomposability” (related to how individual components work), and algorithmic transparency (this is associated to the algorithms’ visibility).

In reality, it is difficult to understand how AI systems, including deep learning models and their neural networks are learning (as they acquire, process and store data) during training phases. They are often considered as black box models. It may prove hard to algorithmically translate derived concepts into human-understandable terms, even though developers may use certain jargon to explain their models’ attributes and features. Many legislators are striving in their endeavors to pressurize AI actors to describe the algorithms they use in automated decision-making, yet the publication of algorithms is useless if outsiders cannot access the data of the AI model.

4.2 Explainability and interpretability

Explainability is the concept that sheds light on how AI models work, in a way that is comprehensible to a human being. Arguably, the explainabilty of AI systems could improve their transparency, trustworthiness and accountability. At the same time, it can reduce bias and unfairness. The explainability of artificial intelligence systems could clarify how they reached their decisions (Arya et al., 2019; Keller & Drake, 2021). For instance, AI could explain how and why autonomous cars decide to stop or to slow down when there are pedestrians or other vehicles in front of them.

Explainable AI systems might improve consumer trust and may enable engineers to develop other AI models, as they are in a position to track provenance of every process, to ensure reproducibility, and to enable checks and balances (Schneider et al., 2022). Similarly, interpretability refers to the level of accuracy of machine learning programs in terms of linking the causes to the effects (John-Mathews, 2022).

4.3 Fairness and inclusiveness

The responsible AI's fairness dimension refers to the practitioners' attempts to correct algorithmic biases that may possibly (voluntarily or involuntarily) be included in their automation processes (Bellamy et al., 2019; Mäntymäki, et al., 2022). AI systems can be affected by their developers' biases that could include preferences or antipathies toward specific demographic variables like genders, age groups and ethnicities, among others (Madaio et al., 2020). Currently, there is no universal definition on AI fairness.

However, recently many multinational corporations have developed instruments that are intended to detect bias and to reduce it as much as possible (John-Mathews et al., 2022). In many cases, AI systems are learning from the data that is fed to them. If the data are skewed and/or if they comprise implicit bias into them, they may result in inappropriate outputs.

Fair AI systems rely on unbiased data (Wu et al., 2020). For this reason, many companies including Facebook, Google, IBM and Microsoft, among others are striving in their endeavors to involve members of staff hailing from diverse backgrounds. These technology conglomerates are trying to become as inclusive and as culturally aware as possible in order to minimize bias from affecting their AI processes. Previous research reported that AI's bias may result in inequality, discrimination and in the loss of jobs (Butcher & Beridze, 2019).

4.4 Privacy and safety for consumers

Consumers are increasingly concerned about the privacy of their data. They have a right to control who has access to their personal information. The data that is collected or used by third parties, without the authorization or voluntary consent of individuals, would result in the violations of their privacy (Zhu et al., 2020; Wu et al., 2022).

AI-enabled products, including dialogue systems like chatbots and virtual assistants, as well as digital assistants (e.g. like Siri, Alexa or Cortana), and/or wearable technologies such as smart watches and sensorial smart socks, among others, are increasingly capturing and storing large quantities of consumer information. The benefits that are delivering these interactive technologies may be offset by a number of challenges. The technology businesses who developed these products are responsible to protect their consumers' personal data (Rodríguez-Barroso et al., 2020). Their devices are capable of holding a wide variety of information on their users. They are continuously gathering textual, visual, audio, verbal, and other sensory data from consumers. In many cases, the customers are not aware that they are sharing personal information to them.

For example, facial recognition technologies are increasingly being used in different contexts. They may be used by individuals to access websites and social media, in a secure manner and to even authorize their payments through banking and financial services applications. Employers may rely on such systems to track and monitor their employees' attendance. Marketers can utilize such technologies to target digital advertisements to specific customers. Police and security departments may use them for their surveillance systems and to investigate criminal cases. The adoption of these technologies has often raised concerns about privacy and security issues. According to several data privacy laws that have been enacted in different jurisdictions, organizations are bound to inform users that they are gathering and storing their biometric data. The businesses that employ such technologies are not authorized to use their consumers' data without their consent.

Companies are expected to communicate about their data privacy policies with their target audiences (Wong, 2020). They have to reassure consumers that the consented data they collect from them is protected and are bound to inform them that they may use their information to improve their customized services to them. The technology giants can reward their consumers to share sensitive information. They could offer them improved personalized services among other incentives, in return for their data. In addition, consumers may be allowed to access their own information and could be provided with more control (or other reasonable options) on how to manage their personal details.

4.5 The security and robustness of AI systems

AI algorithms are vulnerable to cyberattacks by malicious actors. Therefore, it is in the interest of AI developers to secure their automated systems and to ensure that they are robust enough against any risks and attempts to hack them (Gehr et al., 2018; Li et al., 2020).

The accessibility to AI models ought to be continuously monitored at all times during their development and deployment (Bertino et al., 2021). There may be instances when AI models could encounter incidental adversities, leading to the corruption of data. Alternatively, they might encounter intentional adversities when they experience sabotage from hackers. In both cases, the AI model will be compromised and can result in system malfunctions (Papagiannidis et al., 2023).

AI models have to prevent such contingent issues from happening. Their developers' responsibilities are to improve the robustness of their automated systems, and to make them as secure of possible, to reduce the chances of threats, including by inadvertent irregularities, information leakages, as well as by privacy violations like data breaches, contamination and poisoning by malicious actors (Agbese et al., 2023; Hamon et al., 2020).

AI developers should have preventive policies and measures related to the monitoring and control of their data. They ought to invest in security technologies including authentication and/or access systems with encryption software as well as firewalls for their protection against cyberattacks. Routine testing can increase data protection, improve security levels and minimize the risks of incidents.

5. Conclusions

This review indicates that more academics as well as practitioners, are increasingly devoting their attention to AI as they elaborate about its potential uses, as well as on its opportunities and threats. It reported that its proponents are raising awareness on the benefits of AI systems for individuals as well as for organizations. At the same time, it suggests that a number of scholars and other stakeholders including policy makers, are raising their concerns about its possible perils (e.g. Berente et al., 2021; Gonzalez et al., 2020; Zhang & Lu, 2021).

Many researchers identified some of the risks of AI (Li et al., 2021; Magas & Kiritsis, 2022). In many cases, they warned that AI could disseminate misinformation, foster prejudice, bias and discrimination, raise privacy concerns, and could lead to the loss of jobs (Butcher & Beridze, 2019). A few commentators argue about the “singularity” or the moment where machine learning technologies could even surpass human intelligence (Huang & Rust, 2022). They predict that a critical shift could occur if humans are no longer in a position to control AI anymore.

In this light, this article sought to explore the governance of AI. It sheds light on substantive regulations, as well as on reflexive principles and guidelines, that are intended at practitioners who are researching, testing, developing and implementing AI models. It clearly explains how institutions, non-governmental organizations and technology conglomerates are introducing protocols (including self-regulations) to prevent contingencies from even happening due to inappropriate AI governance.

Debatably, the voluntary or involuntary mishandling of automated systems can expose practitioners to operational disruptions and to significant risks including to their corporate image and reputation (Watts & Adriano, 2021). The nature of AI requires practitioners to develop guardrails to ensure that their algorithms work as they should (Bauer, 2022). It is imperative that businesses comply with relevant legislations and to follow ethical practices (Buhmann & Fieseler, 2023). Ultimately, it is in their interest to operate their company in a responsible manner, and to implement AI governance procedures. This way they can minimize unnecessary risks and safeguard the well-being of all stakeholders.

This contribution has addressed its underlying research objectives. Firstly, it raised awareness on AI governance frameworks that were developed by policy makers and other organizations, including by the businesses themselves. Secondly, it scrutinized the extant academic literature focused on AI governance and on the intersection of AI and CSR. Thirdly, it discussed about essential elements for the promotion of socially responsible behaviors and ethical dispositions of AI developers. In conclusion it put forward an AI governance conceptual model for practitioners.

This research made reference to regulatory instruments that are intended to govern AI expert systems. It reported that, at the moment there are a few jurisdictions that have formalized their AI policies and governance frameworks. Hence, this article urges laggard governments to plan, organize, design and implement regulatory instruments that ensure that individuals and entities are safe when they utilize AI systems for personal benefit, educational and/or for commercial purposes.

Arguably, one has to bear in mind that, in many cases, policy makers have to face a “pacing problem” as the proliferation of innovation is much quicker than legislation. As a result, governments tend to be reactive in the implementation of regulatory interventions relating to innovations. They may be unwilling to hold back the development of disruptive technologies from their societies. Notwithstanding, they may face criticism by a wide array of stakeholders in this regard, as they may have conflicting objectives and expectations.

The governments’ policy is to regulate business and industry to establish technical, safety and quality standards as well as to monitor their compliance. Yet, they may consider introducing different forms of regulation other than the traditional “command and control” mechanisms. They may opt for performance-based and/or market-based incentive approaches, co-regulation and self-regulation schemes, among others (Hepburn, 2009), in order to foster technological innovations.

This research has shown that a number of technology giants, including IBM and Microsoft, among others, are anticipating the regulatory interventions of different governments where they operate their businesses. It reported that they are communicating about their responsible AI governance initiatives as they share information on their policies and practices that are meant to certify, explain and audit their AI developments. Evidently, these companies, among others, are voluntarily self-regulating themselves as they promote accountability, fairness, privacy and robust AI systems. These two organizations, in particular, are raising awareness about their AI governance frameworks to increase their CSR credentials with stakeholders.

Likewise, AI developers who work for other businesses, are expected to forge relationships with external stakeholders including with policy makers as well as with actors including individuals and organizations who share similar interests in AI. Innovative clusters and network developments may result in better AI systems and can also decrease the chances of possible risks. Indeed, practitioners can be in better position if they cooperate with stakeholders for the development of trustworthy AI and if they increase their human capacity to improve the quality of their intellectual properties (Camilleri et al., 2023). This way, they can enhance their competitiveness and growth prospects (Troise & Camilleri, 2021). Arguably, it is in their interest to continuously engage with internal stakeholders (and employees), and to educate them about AI governance dimensions, that are intended to promote accountable, transparent, explainable interpretable reproducible, fair, inclusive and secure AI solutions. Hence, they could maximize AI benefits, minimize their risks as well as associated costs.

5.1 Future research directions

Academic colleagues are invited to raise more awareness on AI governance mechanisms as well as on verification and monitoring instruments. They can investigate what, how, when and where protocols could be used to protect and safeguard individuals and entities from possible risks and dangers of AI.

The "what" question involves the identification of AI research and development processes that require regulatory or quasi regulatory instruments (in the absence of relevant legislation) and/or necessitate revisions in existing statutory frameworks.

The "how" question is related to the substance and form of AI regulations, in terms of their completeness, relevance, and accuracy. This argumentation is synonymous with the true and fair view concept applied in the accounting standards of financial statements.

The “when” question is concerned with the timeliness of the regulatory intervention. Policy makers ought to ensure that stringent rules do not hinder or delay the advancement of technological innovations.

The “where” question is meant to identify the context where mandatory regulations or the introduction of soft laws, including non-legally binding principles and guidelines are/are not required.

Future researchers are expected to investigate further these four questions in more depth and breadth. This research indicated that most contributions on AI governance were discursive in nature and/or involved literature reviews. Hence, there is scope for academic colleagues to conduct primary research activities and to utilize different research designs, methodologies and sampling frames to better understand the implications of planning, organizing, implementing and monitoring AI governance frameworks, in diverse contexts.

References

Abina A., Batkovič T., Cestnik B., Kikaj A., Kovačič Lukman R., Kurbus M., Zidanšek A. (2022). Decision Support Concept for Improvement of Sustainability-Related Competences, Sustainability (Switzerland), 14(14), 8539.

Agbese, M., Alanen, H. K., Antikainen, J., Erika, H., Isomaki, H., Jantunen, M., ... & Vakkuri, V. (2023). Governance in Ethical and Trustworthy AI Systems: Extension of the ECCOLA Method for AI Ethics Governance Using GARP. e-Informatica Software Engineering Journal, 17(1), 230101.

Aitken M., Ng M., Horsfall D., Coopamootoo K.P.L., van Moorsel A., Elliott K. (2021). In pursuit of socially-minded data-intensive innovation in banking: A focus group study of public expectations of digital innovation in banking, Technology in Society, https://doi.org/10.1016/j.techsoc.2021.101666

Alnamrouti, A., Rjoub, H., & Ozgit, H. (2022). Do strategic human resources and artificial intelligence help to make organisations more sustainable? evidence from non-governmental organisations. Sustainability, 14(12), 7327.

Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. AI & Society, https://doi.org/10.1007/s00146-021-01326-6

Arya, V., Bellamy, R. K., Chen, P. Y., Dhurandhar, A., Hind, M., Hoffman, S. C., ... & Zhang, Y. (2019). One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. Cornell University, Ithaca, NY, USA, https://doi.org/10.48550/arXiv.1909.03012

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, https://doi.org/10.1016/j.asw.2023.100745

Bauer, J. M. (2022). Toward new guardrails for the information society. Telecommunications Policy, 46(5), https://doi.org/10.1016/j.telpol.2022.102350

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM Journal of Research and Development, 63(4/5), 4-1.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS quarterly, 45(3), 1433-1450.

Bertino, E., Kantarcioglu, M., Akcora, C. G., Samtani, S., Mittal, S., & Gupta, M. (2021, April). AI for Security and Security for AI. In Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy (pp. 333-334), ACM Digital Library, New York, USA, https://doi.org/10.1145/3422337.3450357

Broer, T. (2022). The Googlization of Health: Invasiveness and corporate responsibility in media discourses on Facebook's algorithmic programme for suicide prevention. Social Science & Medicine, 306, https://doi.org/10.1016/j.socscimed.2022.115131

Buhmann, A., & Fieseler, C. (2023). Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. Business Ethics Quarterly, 33(1), 146-179.

Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally?. The RUSI Journal, 164(5-6), 88-96.

Camilleri, M. A., & Troise, C. (2023). Live support by chatbots with artificial intelligence: A future research agenda. Service Business, 17, 61–80.

Camilleri, M. A. (2017). Corporate sustainability, social responsibility and environmental management. Cham, Switzerland: Springer Nature, Cham, Switzerland.

Camilleri, M. A. (2019). Measuring the corporate managers' attitudes towards ISO's social responsibility standard. Total Quality Management & Business Excellence, 30(13-14), 1549-1561.

Camilleri, M. A., Troise, C., Strazzullo, S., & Bresciani, S. (2023). Creating shared value through open innovation approaches: Opportunities and challenges for corporate sustainability. Business Strategy and the Environment, https://doi.org/10.1002/bse.3377

Carvalho, A., Levitt, A., Levitt, S., Khaddam, E., & Benamati, J. (2019). Off-the-shelf artificial intelligence technologies for sentiment and emotion analysis: a tutorial on using IBM natural language processing. Communications of the Association for Information Systems, 44, https://doi.org/10.17705/1CAIS.04443

CNBC (2023). OpenAI CEO Sam Altman says he’s a ‘little bit scared’ of A.I., https://www.cnbc.com/2023/03/20/openai-ceo-sam-altman-says-hes-a-little-bit-scared-of-ai.html

Corea, F., Fossa, F., Loreggia, A., Quintarelli, S., & Sapienza, S. (2022). A principle-based approach to AI: the case for European Union and Italy. AI & Society, 38, 521–535

Crawford, K., & Paglen, T. (2021). Excavating AI: The politics of images in machine learning training sets. AI & Society, 36(4), 1105-1116.

Damoah, I. S., Ayakwah, A., & Tingbani, I. (2021). Artificial intelligence (AI)-enhanced medical drones in the healthcare supply chain (HSC) for sustainability development: A case study. Journal of Cleaner Production, 328, https://doi.org/10.1016/j.jclepro.2021.129598

Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. Review of International Political Economy, 29(3), 696-718.

Du, S., El Akremi, A., & Jia, M. (2022). Quantitative Research on Corporate Social Responsibility: A Quest for Relevance and Rigor in a Quickly Evolving, Turbulent World. Journal of Business Ethics, https://doi.org/10.1007/s10551-022-05297-6

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. Journal of Business Research, 129, 961-974.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994.

Engel, C., Ebel, P., & Leimeister, J. M. (2022). Cognitive automation. Electronic Markets, 32(1), 339-350.

Erdélyi, O.J. & Goldsmith, J. (2022). Regulating Artificial Intelligence: Proposal for a Global Solution, Government Information Quarterly 39 (4), 1-13.

EU (2021). Regulation of the European Parliament and the council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, European Commission, Brussels, Belgium, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Filgueiras, F. (2022). New Pythias of public administration: ambiguity and choice in AI systems as challenges for governance. AI & Society, 37(4), 1473-1486.

Fosch-Villaronga, E., Drukarch, H., Khanna, P., Verhoef, T., & Custers, B. (2022). Accounting for diversity in AI for medicine. Computer Law & Security Review, 47, https://doi.org/10.1016/j.clsr.2022.105735

Frank, B. (2021). Artificial intelligence-enabled environmental sustainability of products: Marketing benefits and their variation by consumer, location, and product types. Journal of Cleaner Production, 285, https://doi.org/10.1016/j.jclepro.2020.125242

Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., ... & Levy, K. (2021). Artificial intelligence, systemic risks, and sustainability. Technology in Society, 67, https://doi.org/10.1016/j.techsoc.2021.101741

Gehr, T., Mirman, M., Drachsler-Cohen, D., Tsankov, P., Chaudhuri, S., & Vechev, M. (2018). Ai2: Safety and robustness certification of neural networks with abstract interpretation. In 2018 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.

Gonzalez, R. A., Ferro, R. E., & Liberona, D. (2020). Government and governance in intelligent cities, smart transportation study case in Bogotá Colombia. Ain Shams Engineering Journal, 11(1), 25-34.

Hamon, R., Junklewitz, H., & Sanchez, I. (2020). Robustness and explainability of artificial intelligence. Publications Office of the European Union, Brussels, Belgium.

Hepburn, G. (2009). Alternatives to traditional regulation, Organization for Economic Cooperation and Development, Paris, France, https://www.oecd.org/gov/regulatory-policy/42245468.pdf

Hickok, M. (2022). Public procurement of artificial intelligence systems: new risks and future proofing. AI & Society, https://doi.org/10.1007/s00146-022-01572-2

Hollanek, T. (2020). AI transparency: a matter of reconciling design with critique. AI & Society, https://doi.org/10.1007/s00146-020-01110-y

Huang, M. H., & Rust, R. T. (2022). AI as customer. Journal of Service Management, 33(2), 210-220.

IBM (2022). Introducing IBM AI Governance: IBM AI Governance is a new, one-stop solution built on IBM Cloud Pak® for Data. Armonk, New York, United States of America, https://www.ibm.com/cloud/blog/announcements/introducing-ibm-ai-governance

IEEE (2023). IEEE Introduces New Program for Free Access to AI Ethics and Governance Standards, Institute of Electrical and Electronics Engineers, New York, United States of America, https://standards.ieee.org/news/get-program-ai-ethics/

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695.

Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2021). Substantial capabilities of robotics in enhancing industry 4.0 implementation. Cognitive Robotics, 1, 58-75.

John-Mathews, J. M. (2022). Some critical and ethical perspectives on the empirical turn of AI interpretability. Technological Forecasting and Social Change, 174, 121209.

John-Mathews, J. M., Cardon, D., & Balagué, C. (2022). From reality to world. A critical perspective on AI fairness. Journal of Business Ethics, 178(4), 945-959.

Keller, P., & Drake, A. (2021). Exclusivity and Paternalism in the public governance of explainable AI. Computer Law & Security Review, 40, https://doi.org/10.1016/j.clsr.2020.105490

Koniakou, V. (2023). From the "rush to ethics" to the "race for governance" in Artificial Intelligence. Information Systems Frontiers, 25(1), 71-102.

Krkač, K. (2019). Corporate social irresponsibility: humans vs artificial intelligence. Social Responsibility Journal, 15(6), 786-802.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Li, W., Su, Z., Li, R., Zhang, K., & Wang, Y. (2020). Blockchain-based data security for artificial intelligence applications in 6G networks. IEEE Network, 34(6), 31-37.

Li, G., Li, N., & Sethi, S. P. (2021). Does CSR reduce idiosyncratic risk? Roles of operational efficiency and AI innovation. Production and Operations Management, 30(7), 2027-2045.

Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). 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, ACM Digital Library, New York, USA, and DOI: https://dl.acm.org/doi/pdf/10.1145/3313831.3376445

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. AI and Ethics, 2(4), 603-609.

Magas, M., & Kiritsis, D. (2022). Industry Commons: an ecosystem approach to horizontal enablers for sustainable cross-domain industrial innovation (a positioning paper). International Journal of Production Research, 60(2), 479-492.

Magistretti, S., Dell'Era, C., & Petruzzelli, A. M. (2019). How intelligent is Watson? Enabling digital transformation through artificial intelligence. Business Horizons, 62(6), 819-829.

Matytsin, D. E., Dzedik, V. A., Makeeva, G. A., & Boldyreva, S. B. (2023). "Smart" outsourcing in support of the humanization of entrepreneurship in the artificial intelligence economy. Humanities and Social Sciences Communications, 10(1), 1-8.

McBride, R., Dastan, A., & Mehrabinia, P. (2022). How AI affects the future relationship between corporate governance and financial markets: A note on impact capitalism. Managerial Finance, 48(8), 1240-1249.

Microsoft (2023). Responsible and trusted AI, Redmont, United States of America, https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/best-practices/trusted-ai

Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2022). What about investors? ESG analyses as tools for ethics-based AI auditing. AI & society, https://doi.org/10.1007/s00146-022-01415-0

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2023). Co-shaping an ecosystem for responsible AI: five types of expectation work in response to a Technological Frame. Information Systems Frontiers, 25(1), 103-121.

Mullins, M., Holland, C. P., & Cunneen, M. (2021). Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market. Patterns, 2(10), https://doi.org/10.1016/j.patter.2021.100362

Narwani, K., Lin, H., Pirbhulal, S., & Hassan, M. (2022). Towards AI-Enabled Approach for Urdu Text Recognition: A Legacy for Urdu Image Apprehension. IEEE Access, https://doi.org/10.1109/ACCESS.2022.3203426

Ng, K. K., Chen, C. H., Lee, C. K., Jiao, J. R., & Yang, Z. X. (2021). A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives. Advanced Engineering Informatics, 47, 101246.

OECD (2019). Recommendation of the Council on Artificial Intelligence, Organization for Economic Cooperation and Development, Paris, France, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Pai, V., & Chandra, S. (2022). Exploring Factors Influencing Organizational Adoption of Artificial Intelligence (AI) in Corporate Social Responsibility (CSR) Initiatives. Pacific Asia Journal of the Association for Information Systems, 14(5), 4.

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. Information Systems Frontiers, 25(1), 123-141.

Rąb-Kettler, K., & Lehnervp, B. (2019). Recruitment in the times of machine learning. Management Systems in Production Engineering, 27(2), 105-109.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 33-44), ACM Digital Library, New York, USA, https://doi.org/10.1145/3351095.3372873

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review, 46(1), 192-210.

Renieris, E.M., Kiron, D., & Mills, S. (2022). Should Organizations Link Responsible AI and Corporate Social Responsibility? It’s Complicated. MIT Sloan, https://sloanreview.mit.edu/article/should-organizations-link-responsible-ai-and-corporate-social-responsibility-its-complicated/

Ribeiro, J., Lima, R., Eckhardt, T., & Paiva, S. (2021). Robotic process automation and artificial intelligence in industry 4.0–a literature review. Procedia Computer Science, 181, 51-58.

Rodríguez-Barroso, N., Stipcich, G., Jiménez-López, D., Ruiz-Millán, J. A., Martínez-Cámara, E., González-Seco, G., ... & Herrera, F. (2020). Federated Learning and Differential Privacy: Software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. Information Fusion, 64, 270-292.

Romao, M., Costa, J., & Costa, C. J. (2019, June). Robotic process automation: A case study in the banking industry. In 2019 14th Iberian Conference on information systems and technologies (CISTI) (pp. 1-6). IEEE.

Sachan, S., Yang, J. B., Xu, D. L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. Expert Systems with Applications, 144, https://doi.org/10.1016/j.eswa.2019.113100

Sætra, H. S. (2021). A Framework for Evaluating and Disclosing the ESG Related Impacts of AI with the SDGs. Sustainability, 13(15), 8503.

Saurabh, K., Arora, R., Rani, N., Mishra, D., & Ramkumar, M. (2022). AI led ethical digital transformation: Framework, research and managerial implications. Journal of Information, Communication and Ethics in Society, 20(2), 229-256.

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2022). Artificial intelligence governance for businesses. Information Systems Management, https://doi.org/10.48550/arXiv.2011.10672

Silva, R. L., Canciglieri Junior, O., & Rudek, M. (2022). A road map for planning-deploying machine vision artifacts in the context of industry 4.0. Journal of Industrial and Production Engineering, 39(3), 167-180.

Smart Nation (2019). National Artificial Intelligence Strategy: Advancing our smart nation journey. Smart Nation and Digital Government Office, Singapore, https://www.smartnation.gov.sg/initiatives/artificial-intelligence/

Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. Computer Law Review International, 20(4), 97-106.

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum, 56(4), 24-31.

Thorp, H. H. (2023). ChatGPT is fun, but not an author. Science, 379(6630), 313-313.

Troise, C., & Camilleri, M. A. (2021). The use of digital media for marketing, CSR communication and stakeholder engagement. In Strategic corporate communication in the digital age. Emerald Publishing Limited, Bingley, United Kingdom.

Wamba-Taguimdje, S. L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. Business Process Management Journal, 26(7), 1893-1924.

Watts, J., & Adriano, A. (2021). Uncovering the sources of machine-learning mistakes in advertising: Contextual bias in the evaluation of semantic relatedness. Journal of Advertising, 50(1), 26-38.

Weber, M., Beutter, M., Weking, J., Böhm, M., & Krcmar, H. (2022). AI Startup Business Models: Key Characteristics and Directions for Entrepreneurship Research. Business & Information Systems Engineering, 64(1), 91-109.

WhiteHouse (2022). Blueprint for an AI Bill of Rights: Making automated systems work for the American people. The White House Washington DC, United States of America, https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Wong, P. H. (2020). Cultural differences as excuses? Human rights and cultural values in global ethics and governance of AI. Philosophy & Technology, 33(4), 705-715.

Wu, W., Huang, T., & Gong, K. (2020). Ethical principles and governance technology development of AI in China. Engineering, 6(3), 302-309.

Wu, L., Dodoo, N. A., Wen, T. J., & Ke, L. (2022). Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing. International Journal of Advertising, 41(4), 685-702.

Zhang, B., Zhu, J., & Su, H. (2023). Toward the third generation artificial intelligence. Science China Information Sciences, 66(2), 1-19.

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, https://doi.org/10.1016/j.jii.2021.100224

Zhu, T., Ye, D., Wang, W., Zhou, W., & Philip, S. Y. (2020). More than privacy: Applying differential privacy in key areas of artificial intelligence. IEEE Transactions on Knowledge and Data Engineering, 34(6), 2824-2843.