AI Ethics in Action: A Circular Model for Transparency, Accountability, and Inclusivity

Seyedeh Asieh Hosseini Tabaghdehi[1#], Ozlem Ayaz[2]

[1] Senior Lecturer in Strategy and Business Economics, Brunel University London, Brunel Business School, Uxbridge, UB8 3PH, United Kingdom
E-mail: Asieh.Tabaghdehi@brunel.ac.uk
Orcid ID: https://orcid.org/0000-0002-6650-766x
 # Coresponding Author
Dr. Seydeh Asieh Hosseini Tabaghdehi Bio: Dr Tabaghdehi is a senior Lecturer in strategy and Business Economy with research interests in responsible innovation, AI ethics, digital transformation, and inclusive leadership. Her interdisciplinary work explores the intersection of emerging technologies with ethical data governance, stakeholder engagement, AI auditing, and social impact. With a background in both academia and industry, she brings a critical and practical perspective to the role of digital technologies in shaping equitable futures. Dr Tabaghdehi was a Co-Investigator on a UKRI/ESRC-funded project on the ethical implications of SMEs' digital footprint data. She is actively involved in knowledge exchange initiatives and thought leadership forums, advocating for ethical, human-centred approaches to responsible AI and technology adoption. She is particularly committed to advancing inclusive and transparent frameworks for the ethical use of AI in business and organizational contexts and has published in top journals, including the *Journal of Business Research* and *International Journal of Hospitality Management,* among others.

[2] Senior Lecturer in Strategic Management, Brunel University London, Brunel Business School, Uxbridge, UB8 3PH, United Kingdom
Tel: +44(0)1895 267623
E-mail: Ozlem.Ayaz@brunel.ac.uk
Orcid ID: https://orcid.org/0000-0002-2836-6317
Dr Ozlem Ayaz Bio: Dr. Ozlem Ayaz is a Senior Lecturer in Strategic Management at Brunel University London. She holds a PhD in Management and Organisations from Bahçeşehir University and brings over a decade of combined industry and academic experience. Prior to joining academia, she held senior roles at prominent Turkish corporations, including Koç Holding and Barem Research, where she specialised in financial auditing, strategic planning, and quantitative analysis. Her industry insights continue to shape her research, which focuses on strategic management, digital transformation in SMEs, ethical leadership, and inclusive workplace practices. Dr. Ayaz was Principal Investigator on a UKRI/ESRC-funded project exploring the digital footprints of SMEs and has published in leading journals such as the Journal of Business Research, International Journal of Management Reviews, and Supply Chain Management: An International Journal.

Abstract:

Purpose: Drawing upon a circular model proposition, Kantian ethics is employed to explore how ethical considerations within AI translate into concrete actions that prioritize transparency, privacy, inclusivity, and equality. Additionally, agency theory is applied to understand the relevance of fairness in the interactions between agents, principals, and algorithmic systems, particularly in the creation of value through digital platforms.

Design/methodology/approach: A review of literature on ethical concerns within the AI ecosystem is conducted, proposing a unifying ethical principle and standards. The circular model for ethics in action is then developed, emphasizing the responsible use of AI and its role in capturing and creating social value, ultimately contributing to sustainable organizational outcomes. The model also highlights key drivers that shape the ethical framing of AI, as well as the influence of the institutional context on its adoption and effectiveness.

Findings: Responsible use of AI positively affects organisational performance and digital ecosystem via the psychological mechanism of ethical identity. Ethical standard and regulation are the global requirement for the AI ecosystem that is required for achieving a sustainable digital society.

Research limitations/implications: This study contributes to a comprehensive understanding of the responsible use of AI's and its practical and theoretical implications for organisations in the current digital ecosystem.

Lack of global understanding, awareness and implementation of ethical practice in AI ecosystem is not yet developed and practice. Future researchers can design a cross border ethical framework to overcome these limitations. Organizations targeting to increase responsible digital interactions can benefit from maintaining ethical principles through responsible labours, leaders and all stakeholders involved in the ecosystem.

Practical Implication: This study offers practical guidance for businesses, policymakers, and AI practitioners on the ethical use of AI. It emphasizes the need for robust data governance, a "human-first" approach focusing on privacy and accountability, and alignment with ethical standards. Given AI's global reach, international cooperation and standard-setting are essential to navigate diverse regulatory and cultural contexts. The paper also highlights the importance of ethics education for AI developers and practitioners. Investing in training that integrates technical skills with ethical awareness will help build a responsible AI workforce capable of addressing societal impacts and maintaining public trust.

Social Implication: This study underscores the urgent need for responsible AI adoption, highlighting risks such as bias, lack of transparency, and privacy concerns. As AI reshapes work, decision-making, and governance, its social impact grows- potentially deepening inequalities if left unchecked. The study calls for explainable, fair, and inclusive AI systems guided by ethical frameworks that respect human dignity. A "human-first" approach ensures AI supports- not replaces- human agency. By fostering transparency, accountability, and cultural sensitivity, organizations can build public trust, empower diverse communities, and contribute to a more equitable digital future. Ethical leadership and inclusive design are essential to avoid reinforcing systemic harms.

Keywords: Artificial Intelligence, digital ethics, social inclusivity, fairness, accountability, responsible digital transformation.

1. Introduction

In recent years, Artificial Intelligence (AI) deployment emerged as a major transformative force in every business by facilitating the development of intelligent products and services, the creation of innovative business models, and the optimization of workflows for greater efficiency (Chowdhury et al., 2022; Ghosh et al., 2021). The integration of AI into organizational practices is rapidly transforming the global digital economy and driving innovation by enhancing operational efficiencies and creating opportunities for workforce development (Henry-Nickie and Sun, 2019; Theres and Strohmeier, 2023; Meijerink and Keegan, 2019). As AI continues to influence managerial decisions, its ethical governance becomes crucial for promoting trust, safeguarding individual rights and mitigating risks associated with algorithmic bias, data privacy breaches, and other unintended consequences (Mujtaba and Mahapatra, 2019; Kramer et al., 2014), which threaten organizational credibility and stakeholder trust (Hagendorf, 2020; Schneier, 2018). Policymakers and organizations have increasingly emphasised the need to address issues such as inclusivity, accountability, and the structural risks associated with AI systems (Askell et al., 2019). Nevertheless, comprehensive, accessible and practical guidelines that combine both 'Ethical AI 'and 'Responsible AI' in ways that promote individual and societal well-being remain elusive (Tsamados et al., 2021). So, it is essential to delineate the distinct concepts of 'Ethical AI' and 'Responsible AI,' which serve as foundational frameworks guiding how AI should be developed, deployed, and governed within organizations.

Ethical AI involves aligning with universal moral standards that prioritize fairness, transparency, and respect for individual autonomy (Floridi et al., 2021). By contrast, responsible AI, translates the ethical principles into actionable organizational frameworks, ensuring accountability, stakeholder trust and risk mitigation throughout the AI ecosystem (Dwivedi et al., 2019; Arrieta et al., 2020). The lack of responsible AI and algorithm fairness created significant challenges and disruption for many businesses and organisations (Faraj and Leonardi, 2022, Makridakis, 2017).

While ethical AI embodies the commitment to 'doing what is morally right', responsible AI operationalises these values into organizational structures and processes (Askell et al., 2019), ensuring alignment with both ethical principles and managerial objectives. Therefore, there is critical need for organizations to adopt a dual-focus approach that integrates ethical AI and responsible AI, effectively addressing the growing complexities of AI governance. However,

how can this be efficiently achieved across various industries and countries. Hence, there may be an urgent need for a comprehensive circular framework linking these principles to theoretical constructs such as Kantian ethics and Agency theory to address the ethical issues surrounding AI ecosystem. Where, Kantian ethics emphasizes fairness, transparency, and respect for stakeholders' rights, while agency theory explores the conflicts of interest that arise between organizational principals (e.g., shareholders or owners) and agents (e.g., managers or employees) (Bowie, 2002; Bosse & Phillips, 2016). Kantianism emphasises that some rules related to individuals' privacy and people's rights to decision making must be followed and respected (L'Etang, 1992). By applying this dual theoretical lens, the study demonstrates how ethical principles can inform Responsible AI practices, ensuring organisational alignment with societal values while addressing broader ethical concerns. Agency theory, for instance, highlights the principal-agent problem, wherein conflicting interests between principals and agents may obstruct responsible AI deployment (Bowie & Freeman, 1992). Principals may prioritise profit optimisation, while agents may focus on personal or positional goals, exacerbating ethical risks. Responsible AI offers a mediating framework to balance these conflicting interests, enabling inclusive, transparent decision-making processes that align ethical responsibilities with organisational objectives.

Through this dual theoretical lens, the study addresses three critical gaps in the literature. First, it integrates Kantian ethics and Agency theory to establish a robust conceptual foundation for understanding both the ethical and managerial implications of AI. While Kantian ethics underscores the necessity of fairness, transparency, and respect for individual rights in AI governance (L'Etang, 1992; Bowie, 2002), Agency theory highlights the principal-agent problem, where misaligned incentives between decision-makers (principals) and AI system developers or operators (agents) can lead to ethical risks, including bias, accountability gaps, and regulatory evasion (Eisenhardt, 1989; Bosse & Phillips, 2016; Martin, 2019). This theoretical integration extends existing scholarship by demonstrating that AI ethics cannot be addressed solely through normative principles but must be reinforced by structural governance mechanisms.

Second, this study advances existing literature by demonstrating how Ethical AI principles can be operationalized into Responsible AI practices, bridging the gap between abstract ethical commitments and concrete implementation. While prior research has identified ethical risks in AI, there remains a significant gap in translating these concerns into enforceable organizational

policies (Floridi et al., 2018; Jobin et al., 2019). By applying Kantian ethics, which stresses universal moral duties and respect for individual autonomy, the study illustrates how AI governance frameworks can ensure inclusivity, fairness, and sustainability (Mittelstadt, 2019; Dignum, 2021). This approach aligns with emerging regulatory efforts, such as the EU AI Act, which seeks to embed ethical principles into AI compliance structures (Veale & Borgesius, 2021).

Finally, the study explores the role of Responsible AI as a mediator of principal-agent conflicts, providing a structured pathway for ethical AI governance that balances diverse stakeholder interests with long-term value creation. While Agency theory has been widely applied in corporate governance to examine power asymmetries and accountability mechanisms, its relevance to AI oversight remains under-explored (Jensen & Meckling, 1976). By integrating agency theory within AI ethics discourse, this research contributes to ongoing discussions on how regulatory frameworks and organizational oversight mechanisms can mitigate the risks associated with AI-driven decision-making. Furthermore, it highlights the necessity of adaptive governance models that evolve alongside AI advancements to prevent regulatory lag (Calo, 2021). By linking these ethical and managerial perspectives, this study not only enhances theoretical understanding but also provides actionable insights for policymakers and organizations aiming to develop AI governance strategies that uphold ethical integrity while fostering innovation and trust. The findings underscore the need for a multidimensional approach that integrates ethical reasoning with institutional accountability, reinforcing the call for interdisciplinary research at the intersection of AI ethics, law, and corporate governance.

Therefore, we investigate further the key questions, including: i) How can Kantian ethics and agency theory inform the development of a Responsible AI framework for ethical AI governance? ii) What are the key organizational challenges in implementing Ethical AI principles within managerial practices? iii) How can Responsible AI mediate principal-agent conflicts and ensure alignment between ethical responsibilities and organizational goals? iv) What role does Responsible AI play in fostering inclusivity, fairness, and long-term sustainability in organizational contexts? This study aims to advance understanding of how Ethical AI and Responsible AI can be integrated into organizational practices, contributing to both managerial and practical applications to ensure transparency, fairness, and inclusivity in AI systems while advocating for socially responsible practices that facilitate long-term value creation and sustainability in the digital era.

2. Literature review:

The Role of AI in Business Transformation and Business Value Creation: Technologies like AI, big data, cloud computing, robotics, 3D printing, and blockchain significantly impact industrial businesses, fostering growth (Ghosh et al., 2021). AI now acts as a facilitator, aiming to create opportunities for intelligent products, services, and innovative business models that enhance both value creation and capture (Chowdhury et al., 2022; Kohtamäki et al., 2019). The propositional framework of AI as an emerging technology and a technological facilitator and driver of business growth were demonstrated in Figure 1.

<< Figure 1 goes here>>

A 2022 global AI survey highlights that AI has significant impact in marketing and sales. Tech giant companies like GAFAM (Google, Amazon, Facebook, Apple, Microsoft) lead in AI integration, leveraging cloud computing to surpass expectations while heavily investing in advancing AI in everyday applications like Siri, Alexa, Cortana, and the rise of self-driving cars (Kumar et al., 2021). Over the past two decades, GAFAM's significant investment of $71 billion in 2017, along with active mergers and acquisitions, underscores their strategic deployment of AI and reveals its impact on industries, user behaviours, and societal norms. Despite its benefits, AI is prone to algorithmic biases, notably seen in facial recognition technology, revealing biases from training data, analytical models, and sociocultural influences (Gautier and Lamesch, 2021; Varian, 2021; Akter et al., 2021; Van Noorden, 2020). Current AI development underscores the importance of addressing these biases to ensure the fairness, accuracy, and ethical use of AI technologies.

The Ethical Concerns of AI Bias in Current Smart Society: The emergence of a smart society, integrating advanced technologies like AI, raises ethical concerns, particularly regarding algorithm bias and fairness, transparency and predictability. Exploring the need for ethics in AI and managing biases in algorithms is essential. Despite AI rapid progress, many nations are unprepared for the potential consequences on a global scale. It is imperative to maintain awareness of evolving AI developments, tracking trends and interactions for societal adaptability (Foresti et al., 2020; Calp and Butuner, 2022; Naicker and Van Der Merwe, 2018; Gruetzemacher and Whittlestone, 2022). The key challenge with AI bias arises from the historical biases by relying on past data (Roselli et al., 2019). As AI operates with parallel cognitive architectures, generating solutions beyond human comprehension, ensuring safety

assurances, especially for AI with superhuman intelligence, is crucial (Kilian et al., 2023; Bostrom and Yudkowsky, 2014).

Reducing AI bias in data gathering and development is crucial for trust and minimizing speculation, protecting both AI technology and society. A comprehensive approach to addressing ethical dimensions involves advocating for independent agencies overseeing AI ethics, as proposed by Anderson and Anderson (2010). These agencies can scrutinize AI behaviour, manage biases, and foster universal ethical standards. Unregulated AI mechanisms may worsen societal disparities and resource allocation inequities, highlighting the importance of cultivating moral frameworks for AI, like human ethics (Rossi, 2015). As AI deeply integrates into society, the interaction of human and machine intelligence, from autonomous vehicles to voice automation, raises complex ethical questions (Johnson et al., 2022; Wien & Peluso, 2021). This prompts exploration of transparency, accountability particularly in healthcare, recruitment, and criminal justice (Tubadji et al., 2021). AI impact on individuals, businesses, and institutions necessitates vigilant evaluation despite human decision-making imperfections (Burström et al., 2021). Considering data-driven innovation, central to the AI revolution, introduces algorithmic biases from training data and analytics models (Akter et al., 2021), which may result jeopardizing individuals based on ethnicity, income, sexual orientation, religion, or gender (Akter et al., 2021, 2022), or lead to discriminatory pricing, limiting resources for vulnerable customers or minorities (Akter et al., 2021). Figure 2 illustrates a propositional framework, capturing AI's transformative capabilities and ethical obligations for fairness, accountability, and transparency in the evolving smart society.

<<Figure 2 goes here>>

Therefore, the need for proactive approach is inevitable to ensure that technological advancements are paralleled by ethical considerations, fostering a society that reaps the benefits of AI while upholding core principles of fairness and accountability. Table 1 shows a narrative review of existing literature about the ethics of AI.

<<Table 1 goes here>>

The effortlessness of Information Accessibility: In the context of AI advancement, addressing data accessibility is a critical concern, linked to the fair generation and availability of information. The rise of digital technologies and diverse platforms have dramatically amplified data accessibility, reshaping business operations and consumer information access (Morris, 2009; Langan et al., 2019). This data accessibility evolves into a capability that

enhances stakeholder communication and plays a pivotal role in organisational strategies amid digital transformation, with a focus on digital ethics to avoid disruptions (Stonehouse and Konina, 2020). This accessibility supports overall business value creation within the digital ecosystem, facilitating informed decision-making and effective stakeholder engagement.

After the COVID-19 pandemic outbreak, almost all of the organisations globally embedded the humans-machine collaboration so promptly. With the exponentially growing amount of data that is attempting to humanise technology, AI is expected to learn at a faster pace day by day and therefore is expected to solve complex problems that undoubtedly improve our lives. However, as AI integration deepens, new ethical concerns are emerging, particularly around *moral choice* and *moral hazard* in situations where technological decisions may conflict with human values. Moral choice becomes particularly critical in scenarios where AI-driven automation may disproportionately affect certain job sectors, raise questions around workforce well-being, or prioritize efficiency over ethical considerations. Moreover, moral hazard- a situation where organizations may excessively rely on technology, neglecting broader societal impacts- poses significant risk if AI is deployed without adequate oversight or ethical standards to safeguard workers' rights and inclusivity. (Tabaghdehi, 2022). Addressing these ethical tensions is essential to ensure that technological advancement aligns with fair and responsible practices in a workplace.

It remains uncertain whether the rise of technological advancements will boost productivity and create long-term value, or if it will ultimately reduce labour productivity by shifting more tasks from humans to machines. Preceding studies highlighted that technological advancement could amplify workforce skills or urge them to compete with technology (Acemoglu and Autor, 2011; Brynjolfsson and McAfee, 2012; 2014). Yet, many countries are heavily investing in AI-based production and digital transformation, such as China which is making digital technology and AI the key effective factor in their economic development (He ABY, 2017). Although, the technological progression improves the workforces' skills discrepancy and professional mobility, which can help economies to develop more rapidly but workforces need to be equipped well with digital skills and soft skills to remain resilient with swift technological changes (Jaiswal et al., 2022).

Engaging humans into production with collaborative robots has developed the fears and debates about the substitution of humans for robots. This concern of being replaced by robots pushes employees to develop soft skills and abilities to be the part of the new value-creation

process (Cartwright, 2018). Nevertheless, moral concerns arise when workers' fears of replacement are not ethically addressed, raising questions about how organizations can morally justify automation choices that may adversely impact job security. By incorporating moral hazard considerations, companies can mitigate excessive reliance on automation at the expense of workforce stability, ensuring that AI-enabled transitions are ethically grounded and socially responsible. To address these concerns, optimizing digital work transformation requires an ethical framework that not only enhances job security but also prioritizes workforce well-being and fairness. In this light, ethical AI offers an approach that balances the benefits of automation with the need for inclusive job preservation and worker empowerment, ensuring that technological advancements contribute positively to both organizational objectives and societal welfare.

Principles, Standards and Regulations of AI ethics: The inclusion of AI is no longer about productivity enhancement, but it is more about the reordering of value creation by human effort regarding the nature of work itself (Phan et al., 2017). Consequently, ethical concerns such as diversity, equality, fairness and the boundaries of responsibility have become central questions in relation to technological transformation and digital adoption for policy makers and businesses. Thus, digital ethical practice that could include the voices of all affected communities are essential for avoiding discrimination and marginalisation. The algorithms, or technological system behind AI, need to be designed by a diverse group of people to examine multiple perspectives from different users' viewpoints to reduce discrimination, bias, and power-imbalances in the way the algorithms work (Zou and Schiebinger, 2018).

The negligent AI and algorithm can target or eliminate certain categories of people based on their ethnicity, gender, sexual orientation and religious beliefs from products and services, or price (Wachter, 2019). For example, resulting from the biases in training the data, recruitment apps favour one gender over another, which indicates the breaches in anti-discrimination laws (Eitel-Porter, 2021; Dastin, 2018). From a Kantian perspective, the ethical issues surrounding AI underscore a critical failure to treat individuals as ends in themselves. In AI applications, this principle implies that systems must be designed and implemented with a respect for the autonomy, rights, and well-being of each individual affected by them. When AI is developed or deployed in ways that disregard this respect- such as through biased algorithms or decisions that disadvantage certain groups- the technology violates Kantian principles by failing to afford all individuals equal consideration and respect.

In the context of agency theory, the issue of fairness gains relevance as it relates to the interactions between agents and principals and algorithmic systems within organizations. Unfair algorithms can create misalignments between the interests of all parties, leading to conflicts and mistrust. Such algorithmic biases can increase the agency problem by introducing potentially harmful consequences to both individuals and businesses (Osonde et al., 2017). Kantian ethics complements agency theory in this analysis by proposing that fairness is not merely a contractual obligation but a moral duty that extends to all stakeholders. According to Kantian principles, organizations should ensure that the interests of all individuals affected by AI- whether they are employees, customers, or other community members- are equally weighted in decisions related to AI deployment. This moral duty to fairness, as posited by Kantian ethics, strengthens the case for an ethical framework that mandates transparency, inclusivity, and impartiality in AI-driven decision-making.

Furthermore, the ethical issue of accountability emphasises that individuals or organisations who are designing the tools and applications and engineering the algorithm are responsible for its further consequences and implications. Kim et al. (2020) indicated that the issue of transparency and accountability in algorithms can be tackled efficiently by designing more human-interpretable algorithms with sufficient explanations. Furthermore, Gunning et al, (2019) emphasise that the use of explainable AI in simulating decision-making scenarios can improve the transparency and accountability aspects in decision making processes and output. The AI accountability is more about the control system among the humans and technology (Wirtz et al, 2019) and in the current society, when relying on machines and technology so vastly, it is essential to justify the requirements and assessments in relation to accountability of the AI based decision-making process. Within Kantian ethics, accountability is not merely a regulatory measure but a moral obligation that respects stakeholders' autonomy. Kantian principles dictate that actions (including technological ones) should be guided by a moral intention to benefit stakeholders, thus obliging organizations to maintain transparency as a form of respect for the right of individuals to informed consent.

According to Kantianism theory, what makes an action ‘the right thing to do’ must be found in the agent’s intention (De Colle and Werhane, 2008). Agency theory supports the development of AI ethical standards by focusing on conflicts of interest that might arise between organizational principals and agents, especially when deploying AI solutions. This study proposes that agency theory aids in creating accountability structures that ensure ethical

alignment across organizational levels, promoting trust and transparency in AI governance (Clarke, 2019; Dwivedi et al., 2019).

Therefore, the current digital era introduces a key challenge as agents and principals with diverse interests and priorities engage in ethical decision-making within organizations (Floridi et al., 2021). In such a dynamic landscape, addressing potential conflicts and ensuring harmonious ethical outcomes is challenging, necessitating a strategic integration of Kantian ethics and agency theory for ethical decision-making. Kantian ethics emphasizes considering the interests of all stakeholders (Bowie, 2002), while agency theory focuses on managing differing interests between principals, agents, and algorithmic systems when dealing with AI. Through Kantian ethics, organizations are reminded to regard each stakeholder's dignity and autonomy as paramount, fostering an ethical climate that transcends mere profit motives and prioritizes fair and humane treatment. By practicing an ethical framework grounded in these dual theoretical perspectives, any organization can cultivate a culture that empowers agents and principals to collaboratively shape responsible decision-making, ensuring a balance between individual priorities and collective ethical imperatives.

AI Governance Across Countries and Cultures: AI governance is rapidly becoming a critical issue on the global stage, with various nations and regions developing their own principles for responsible AI innovation. While AI has the potential to transform industries and societies, its impact is heavily influenced by the regulatory frameworks that govern its development, deployment, and use. As AI technology continues to evolve, governments, international bodies, and organizations have been working to establish frameworks to ensure AI operates in ways that are transparent, accountable, ethical, and beneficial to society.

The United States, the United Kingdom, Australia, Canada, the European Union, and China are at the forefront of creating AI governance principles that align with their societal values, economic priorities, and cultural contexts. For example, the United States emphasizes AI innovation and technological advancement, with a focus on maintaining leadership in AI and fostering competitiveness (US Executive Office of the President, 2016). The EU, on the other hand, is focusing on ensuring that AI systems are human-centric, transparent, and respectful of fundamental rights, as reflected in its White Paper on Artificial Intelligence (European Commission, 2020). Meanwhile, China has adopted a more centralized approach to AI governance, with the government actively promoting AI as a key component of national

strategy while emphasizing the need for regulation that aligns with societal stability and the broader public good (Beijing Academy of Artificial Intelligence, 2019).

In addition to these regional efforts, international cooperation is becoming increasingly vital in addressing the global challenges AI presents. The uneven development of AI governance frameworks between countries highlights the need for international standards and collaboration, especially given the cross-border nature of AI technologies. This includes not only the regulation of data privacy and ethical AI design but also the alignment of AI systems with broader global challenges, such as climate change, inequality, and human rights (Binns, 2018). One of the major challenges in AI governance is balancing national interests with global norms. As countries like China and the United States prioritize different aspects of AI development- such as economic growth versus individual privacy- there is growing recognition of the need for international dialogues and standards that can foster fairness, inclusivity, and accountability in AI systems worldwide. These international frameworks will need to address differing regulatory environments, varying cultural norms, and the risks of AI-driven inequalities across regions (Dwivedi et al., 2019).

Furthermore, the cultural context plays a significant role in shaping how AI governance frameworks are developed and implemented. While Western approaches to AI governance often emphasize individual rights and democratic values, other regions may place greater importance on collective well-being, societal harmony, and state oversight. This divergence necessitates a deeper understanding of how AI governance can be shaped by cultural contexts and how different cultural perspectives can influence the ethical and practical deployment of AI technologies. As AI continues to shape the future of work, health, education, and governance, understanding and reconciling these cultural differences in AI regulation will be crucial for ensuring that AI development benefits society at large while mitigating risks such as biases, discrimination, and privacy violations (Gunkel, 2018).

The development of AI governance principles must be a collaborative and inclusive process that engages multiple stakeholders- from governments and international organizations to corporations and civil society. This engagement will help ensure that AI is not only technologically advanced but also ethically sound and aligned with the broader goals of social justice, equity, and environmental sustainability. As various nations move forward with their AI strategies, it is critical to establish common ethical principles and governance frameworks that can guide the responsible development and deployment of AI technologies globally.

3. Discussion and Implications

By addressing principal-agent dynamics, agency theory provides a framework for understanding the ethical challenges in AI adoption. This framework enables organizations to implement Responsible AI practices that align agents' ethical concerns with principals' business objectives, fostering sustainable and ethically sound digital transformation (Faraj & Leonardi, 2022; Makridakis, 2017).

Beyond identifying principal-agent conflicts in organizational contexts, agency theory provides valuable insights into how responsibility and accountability can become diluted when decision-making authority is transferred to AI systems (Dawson et al., 2016). For example, when organizations deploy algorithmic decision-making tools in recruitment or lending, it is not always clear whether the organization (the principal) or the developers and users of the AI system (the agents) should be held accountable for biased or unfair outcomes. This diffusion of responsibility exacerbates the ethical risks associated with AI and makes governance more challenging (Dwivedi et al., 2019: Ayaz et al., 2025). By applying agency theory, this study highlights how the lack of transparency in algorithmic decision-making can worsen principal-agent problems and ethical hazards within digital ecosystems.

Ethics in Digital Governance: Ethics and governance are essential for the sustainable and efficient AI development in current workplaces and society. As AI systems gain autonomy and decision-making capabilities, ensuring ethical conduct becomes primary requirement to prevent potential failures and negative outcomes (Murray et al., 2021; Balasubramanian et al., 2020). AI bias could lead the users to make an unfair decision which could be harmful for individuals and organisations. To minimise the risk of biased algorithms we require the governance of AI applications to manage and mitigate the related risk (Wu et al., 2020). This will provide a chance for regulators to review, evaluate and audit the input and generated output through the AI and algorithm systems consistently. Many countries including UK, US, Australia, Canada, Europe and China including others are developing AI governance principles to innovate responsible AI for greater society, people and environments (Beijing Academy of Artificial Intelligence, 2019). The EU High-Level Expert Group on AI and he National Institute of Standard and Technology in the US produced the ethics guidelines for trustworthy AI (European Commission, 2019). Moreover, the United Nation (UN) developed the AI ethics by emphasising on the use of AI with human participation for global sustainable development and growth. The uprising promotion of AI governance technology globally indicates the global

concerned around the issue of ethics in AI and algorithm design at individual, business, national and international level. However, due to cross-border interactions and global business engagement in digital ecosystem, ethical principles for AI necessitate a diverse and comprehensive compliance strategy aligned with global ethical initiatives and human values. Kantian ethics reinforces the necessity of embedding ethical conduct in AI governance by emphasizing that organizations have a moral duty to respect stakeholders' autonomy, beyond mere compliance with regulatory frameworks (Gal et al., 2022). This aligns with the European Parliament's position that ethical AI governance must go beyond legal compliance to include respect for fundamental rights, human dignity, and the promotion of the common good, requiring organizations to embed ethical principles proactively within their AI systems and decision-making processes (European Parliament, 2020).

Regulated AI: While there is no established method for building trust with technology, digital trust can be cultivated when technology provides tangible benefits and safety, supported by effective regulation (Winfield and Jirotka, 2018). Current advanced technologies contribute to rule formalization by enhancing service delivery, reducing corruption, and mitigating discrimination (Barker and Jewitt, 2022), helping the crisis management, real-time data analysis, big data insights, strengthening public sector service provision and reinforcing the democratic role of bureaucracy (Newman et al., 2022). However, building digital trust necessitates robust regulation and governance that regulatory bodies could align the guidelines with public activities to enhance the impact of governing procedures in society and promote transparency (Wilsdon and Willis, 2004; Wilkinson and Weitkamp, 2016). Despite the importance of the global need for the regulated AI, yet there is insufficient evidence of good practices. Effective AI regulation is essential for fostering digital trust and mitigating risk of ethical harm to privacy, trust, and future employment (Sugianto et al., 2021). Ethical initiatives such as transparency, accountability, and inclusivity are fundamental properties of ethical regulations, with transparency being a particularly crucial principle (Winfield and Jirotka, 2018), and accountability is yet the puzzle in the AI ecosystem. This requires that particular attention should be directed toward the quality of decisions made by responsible AI, especially when ethical hazards arise. In the current AI ecosystem, there is a serious need for a universal code of ethics using regulated AI to optimize the AI-based decision-making process and production while enhancing workforce and citizens' trust, confidence, and well-being. In the context of regulated AI, establishing digital trust is the most important aspect to ensure the responsible and ethical use of technology by citizens voluntarily. Utilizing Kantian ethics,

which emphasize the moral worth of actions based on the motivations of agents, indicates the importance of ethical governance in the AI ecosystem (De Colle and Werhane, 2008). Similar to Kantian ethics, promoting actions driven by principles and duties that respect the rights of individuals, ethical governance in AI requires the principles that safeguard the rights, privacy, and well-being of both users and society at large. From a Kantian perspective, the effort to establish digital trust through regulation is not only a strategic necessity but also a manifestation of an organization's duty to treat individuals as ends in themselves, ensuring their rights and dignity are preserved in AI-enabled environments (Sama et al., 2022; Ayaz et al., 2025).

Risks to Responsible AI: Machine learning algorithms increasingly aid businesses in various setting including document and data handling, sales forecasts, cross-selling, drug discovery, customer targeting, executive recruitment, and human resources (Agrawal et al., 2018). Concerns arise as algorithmic evaluations might replace human judgment, altering morality irreversibly (Moser et al., 2021). Unethical use of AI poses risks to trust, reputation, and privacy, yet public confidence in AI decisions is lacking (Accenture, 2019; Dignum, 2019; Schmidt et al., 2022). The risks that arise due to insufficient compliance and weak AI governance can be categorised as: (i) breaching regulations, (ii) brand damage, as exemplified by Microsoft's Tay chatbot incident, and (iii) third-party transparency issues arising from opaque AI tools (Eitel-Porter, 2021). Aligning AI decisions with universal ethical principles can mitigate risks, ensuring compliance, promoting transparency, and follow ethical norms. Practitioners and policymakers seek AI solutions that consider societal values, ethics, stakeholder priorities, and transparency (Dignum, 2017), promoting more responsible use of AI across various applications. To achieve this, incorporating Kantian ethics for transparent AI explanations and reinforcing accountability and ethical responsibility could enhance the commitment to broader ethical principles for positive societal contributions. Moreover, agency theory helps illuminate how insufficient ethical oversight can amplify principal-agent problems (Westphal et al., 2013), particularly when algorithmic decision-making reduces human intervention and ethical accountability, increasing the risk of opportunistic behavior or ethical negligence (Young et al., 2019).

Algorithm Fairness: In machine learning, the problem of algorithmic bias is widely studied in a wide range of disciplines (Cowgill and Tucker, 2020; Corbett-Davies et al., 2017). Eliminating human interaction from the processes, automation is promising in terms of fairness and equality (Raisch and Krakowski, 2021). The AI algorithms could be perceived as a fair

system when they provide better inclusion of civil society or other relevant stakeholders in an interactive manner (Jobin et al., 2019). The issue of fairness can be argued in both ways. In banking and finance, AI based systems are used for credit approval to augment their processes and reduce human bias that might exclude some customers from reaching enough credit score due to their ethnicity, gender, or postal code (Daugherty and Wilson, 2018). On the other hand, the outcome of decisions made by a machine might be considered unfair due to unbiased positioning of the range of factors in their system design. Therefore, human intelligence augmented by AI is expected to find new solutions by sustaining justice and fairness which rely on different objectives including eliminating unfair discrimination, ensuring that the use of AI will generate benefits for the relevant audience, prevent the harms to the existing social structures (Floridi et al., 2021). The Kantian ethics aligns well with the notions of algorithm fairness that introduces a foundation rooted in ethical principles that emphasize on how AI capabilities can be harnessed to provide unbiased outcomes for all (See Figure 3). In this regard, Kantian ethics offers a normative foundation for algorithmic fairness by asserting that fairness is not contingent on outcomes alone but on the inherent respect for every individual's right to unbiased treatment, irrespective of efficiency considerations.

<<Figure 3 goes here>>

Although, the technological revolution and rise of AI enhance productivity but also it compromises job opportunities for the workforce. In some sectors such as manufacturing the AI advancement weakens wages and job opportunities which benefits organisations and businesses as a cost-saving strategy (Autor, 2010; Acemoglu and Restrepo, 2017). It was predicted in 1952 by Wassily Leontief that the value of labour diminishes, and machine will progressively replace the workforce. In the recent AI ecosystem, rapid technological advancements have not only brought about structural changes but also shifted work values and psychological contracts within the labour market. This entails that employers and human resource systems will be designed to mitigate uncertainties between principals and agents. (Cohen and Baruch, 2010) (See Figure 4).

<< Figure 4 goes here>>

Hence, we develop a circular framework (see Figure 5) that focuses on ethical use of AI and algorithm fairness, to extend the workforce's proficiencies and digital- and soft- skills for the benefit of citizens, society and economy. The proposed framework delineates a comprehensive journey, spanning from the repercussions of technological disruption to the

emergence of social and economic value through ethical AI actions. The initial segment of the framework scrutinizes the ramifications of technological disruption, categorized as AI bias, accessibility, and automation and job replacement. To understand how organizations enhance their value creation strategy within the fast-moving digital era, current organisations require to prompt the re-evaluation of value- capturing and creation mechanisms within their digital ecosystem. Transitioning to the principles, standards, and regulations of AI ethics, the framework underscores inclusivity, equality, fairness, and accountability as core ethical principles. These principles intersect directly with our inquiry into the strategic incorporation of value creation, as businesses focus on ethical digital adoption, they ensure sustainable and responsible value generation. Finally, the ethical digital governance, regulated AI practices, responsible AI approaches, and algorithm fairness are the pivotal requirements for ethical AI actions. The latter portion of Figure 5 highlights the social and economic value creation resulting from the responsible implementation of AI. In conclusion, our circular framework and analysis emphasize on how organizations strategically enhance value creation within the dynamic digital era, while navigating the ethical implications of technological advancements.

<<Figure 5 goes here>>

This study holds significant practical implications for policymakers, practitioners, and organizations aiming to manage the complexities of the AI ecosystem. It is crucial to understand that regulated AI underscores the need for organizations to implement governance frameworks that align with global ethical principles, such as transparency, accountability, and inclusivity. For practitioners, these principles serve as a blueprint for embedding fairness and inclusivity into AI-driven decision-making processes. This is particularly critical in fields such as finance, healthcare, and human resources, where the risks of algorithmic bias and unfair outcomes can have profound effects on stakeholders. Ensuring fairness in AI systems not only improves decision quality but also fosters trust among employees, customers, and the broader public (Tabaghdehi, 2024).

For policymakers, this research highlights the urgent need to develop adaptive governance policies that strike a balance between innovation and ethical considerations. The variation in AI governance principles across counties highlights the need for harmonized standards that goes beyond the national boundaries, developing a unified approach to responsible AI development. Regulatory frameworks should aim not only to mitigate risks while incentivising the creation of technologies that align with societal values and address global challenges such

as sustainability and equitable growth. By emphasizing the principles of inclusivity, equality, and fairness, the proposed circular framework (Fig 2 and Fig 3) offers organizations practical strategies for addressing the psychological and social impacts of technological disruption. For example, it enables companies to align AI-driven innovations with workforce development, ensuring that technological advancements contribute positively to employee well-being and organizational sustainability. Ultimately, adopting responsible AI is not merely a compliance issue but a strategic imperative that will shape the future of work and organizational success. The integrated application of Kantian ethics and agency theory allows us to conceptualize Responsible AI not merely as a technical or managerial concern but as a moral imperative, bridging the gap between individual ethical intentions and structural governance mechanisms within and across organizational boundaries.

4. Conclusion

This study highlights the critical tension between automation and augmentation in AI's business applications, emphasizing the complexities of machine learning algorithms that, while fast and efficient, often lack interpretability (Baum and Haveman, 2020; Alpaydin, 2016; Burrell, 2016). This lack of transparency presents a significant challenge, particularly as AI transforms ecosystems of production, management, and governance. As AI reshapes organizational landscapes, it is essential to establish frameworks for managing and regulating automated systems to prevent ethical pitfalls (Shestakovsky, 2017). While AI offers substantial opportunities, its potential for bias and unpredictability requires ongoing human oversight. To mitigate risks such as data privacy violations and unintended consequences from opaque algorithms, businesses must prioritize transparency, explainability, and accountability in their data governance strategies (Gregory et al., 2021). The responsible use of AI fosters organizational trust and aligns technology with a "human-first" approach (Russell et al., 2015). Ethical decision-making must be embedded within digital governance frameworks, with upper management ensuring that AI systems are designed, implemented, and audited responsibly (Arrieta et al., 2020; Eitel-Porter, 2021).

As AI systems aim to replicate or surpass human capabilities, they introduce greater unpredictability, demanding new safety assurances (Bostrom and Yudkowsky, 2014). Without proper regulation, AI may perpetuate inequalities, from biased recruitment to the unfair allocation of social resources. Inclusivity, fairness, and transparency are critical to building trust and mitigating societal concerns about AI’s potential harms. Informed by Kantian ethics,

this study emphasizes fairness, inclusivity, and privacy as foundational ethical principles (Bowie, 2002). Organizations must treat individuals as ends in themselves, not merely as means to an economic or technological end. By combining Kantian ethics with agency theory, we present a framework for aligning AI-driven initiatives with diverse stakeholder interests, fostering accountability within AI governance (Bosse and Phillips, 2016).

The ethical dimensions of AI are vital for driving sustainable digital transformation. By prioritizing responsible AI that encompasses transparency, privacy, inclusivity, and fairness, organizations can build trust, empower employees, and positively impact society (Dwivedi et al., 2019). This study stresses the importance of algorithmic fairness and ethical practices in AI development (Faraj and Leonardi, 2022; Makridakis, 2017), offering a comprehensive framework for responsible AI deployment that shapes a future-oriented, equitable digital ecosystem. Finally, this research calls for further exploration of ethical AI frameworks to align technology with societal values, ensuring responsible decision-making and promoting a more inclusive and transparent digital future. Addressing global challenges- such as differing regulatory environments, cultural values, and AI-induced inequalities- requires the development of international ethical standards. By fostering a unified approach to AI governance, we can bridge the gap between technological progress and societal well-being, contributing to a more ethical and equitable global AI landscape.

Managerial Implications: The findings of this study provide essential guidance for managers overseeing AI adoption in their organizations. As AI continues to influence organizational structures and decision-making processes, managers must take proactive steps to ensure that AI technologies are developed and applied ethically. It is crucial for managers to advocate for the development of explainable AI systems, as the complexity of machine learning models (Burrell, 2016) can often obscure the decision-making process. Transparency in these systems ensures that stakeholders can understand how decisions are made, fostering trust and enabling more effective accountability mechanisms (Arrieta et al., 2020). By promoting algorithmic transparency, organizations can also reduce the risk of legal challenges related to discriminatory practices or violations of data privacy.

As AI systems become more integrated into business processes, organizations must consider the potential biases that may be embedded in these systems (Russell et al., 2015). Involving diverse teams in the design and deployment of AI can help mitigate biases by ensuring that different perspectives are considered, which promotes more equitable outcomes

(Rossi, 2015). Managers need to emphasize fairness in AI applications, particularly in areas such as recruitment and resource allocation, where biased algorithms can cause significant social harm (Dwivedi et al., 2019). Upper management must also play an active role in fostering a culture of ethical AI use within the organization. This includes creating robust AI governance frameworks that establish clear guidelines for the ethical design, deployment, and auditing of AI systems (Gregory et al., 2021). Managers should ensure that AI systems adhere to both internal and external ethical standards, aligning technological outcomes with the organization's broader commitment to social responsibility (Bowie, 2002; Makridakis, 2017).

Practical Implications: From a practical standpoint, this study offers actionable recommendations for businesses, policymakers, and AI practitioners. As AI continues to permeate various sectors, there is an urgent need for regulatory frameworks to guide the ethical use of AI. Businesses must adopt data governance strategies that ensure AI systems respect privacy and comply with ethical standards (Eitel-Porter, 2021). A "human-first" approach to AI governance, which emphasizes privacy, data security, and accountability, is necessary to build and maintain public trust (Shestakovsky, 2017).

In a globalized economy, AI systems often operate across borders, making it essential for organizations to address varying regulatory environments and cultural differences in AI governance. The development of international standards for ethical AI would help mitigate disparities in AI regulations, ensuring that these systems are designed and deployed responsibly, respecting local values while promoting fairness and inclusivity (Dwivedi et al., 2019). Furthermore, practitioners and developers need to be educated on the ethical implications of AI. Organizations should invest in training programs that foster an understanding of both the technical and ethical aspects of AI. This would help create a workforce that is not only skilled in deploying AI but also aware of its societal impact and ethical responsibilities (Russell et al., 2015).

Future Research Avenues: While this study provides a comprehensive framework for ethical AI governance, further research is needed to explore the evolving ethical challenges posed by AI. More work is required to design and evaluate AI governance models that effectively integrate ethical considerations with technological advancements (Bostrom and Yudkowsky, 2014). Future studies could focus on the scalability of these governance frameworks across different industries and sectors.

Further research is also necessary to assess the long-term societal impact of AI, particularly in relation to its potential to exacerbate inequalities. Investigating how AI interacts with social structures, such as education, healthcare, and employment, will provide deeper insights into the ethical implications of widespread AI adoption (Makridakis, 2017). As AI technologies are increasingly deployed across borders, research into the development of international AI ethics standards will become crucial. Future studies could explore how diverse cultural perspectives influence AI design and implementation, and how these differences can be reconciled to promote global fairness and inclusivity (Dwivedi et al., 2019).

Ultimately, AI's transformative potential extends beyond its ability to enhance operational efficiency. It underscores the importance of ethical considerations in shaping a responsible, fair, and transparent digital ecosystem. By focusing on transparency, fairness, inclusivity, and accountability, organizations can not only ensure the ethical deployment of AI but also foster trust, respect individual autonomy, and contribute to broader societal well-being. As AI technology continues to evolve, the need for ethical governance frameworks will grow more pressing, and organizations must remain vigilant in aligning their AI strategies with both organizational and societal values. This study calls for further exploration into cross-border AI ethics, international regulatory frameworks, and the long-term societal implications of AI to ensure that AI's future is not only technologically advanced but also socially just and responsible.

Figure 1: Propositional framework of emerging technology as technological facilitator and driver of business growth (Source: Created by Authors)

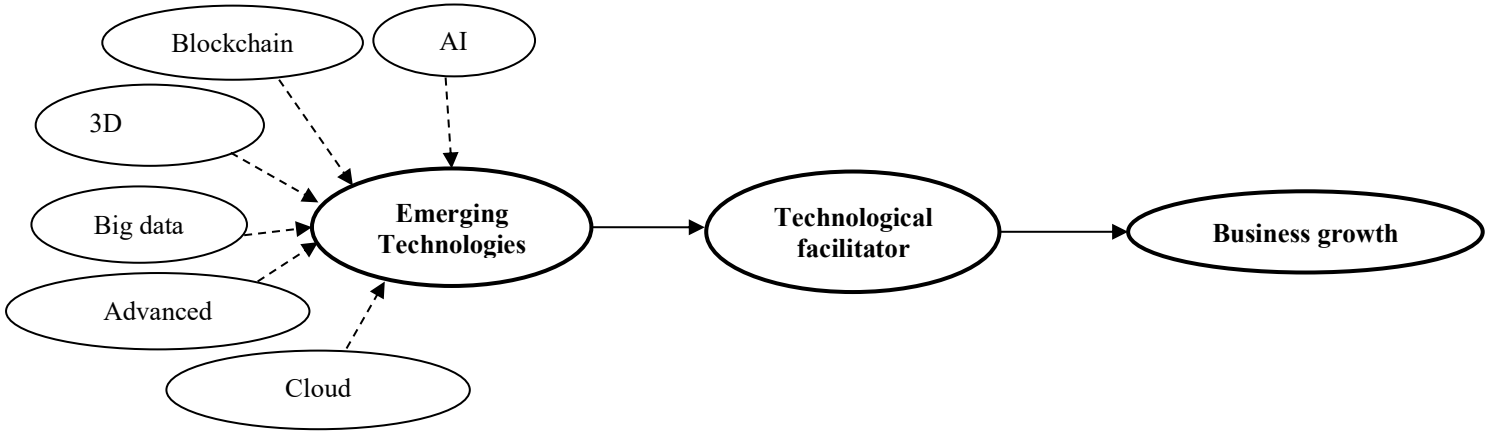

Figure 2: Propositional framework of a technological disruption and raise of ethical concerns (Source: Created by Authors)

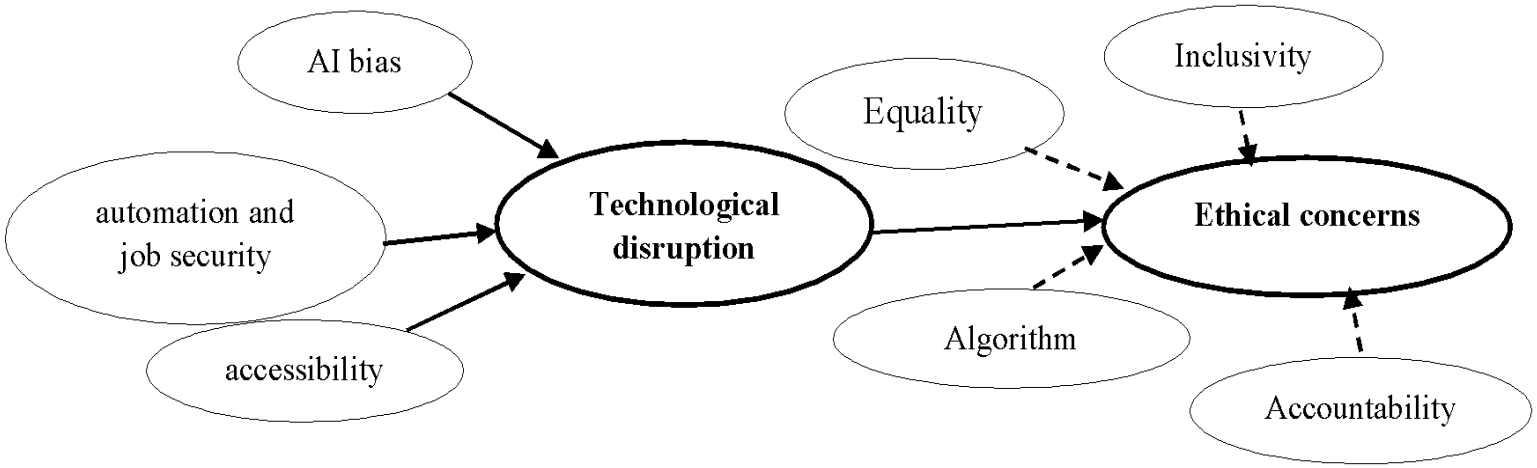

Figure 3: Propositional framework of deriving factors on AI ethical principle and standards (Source: Created by Authors)

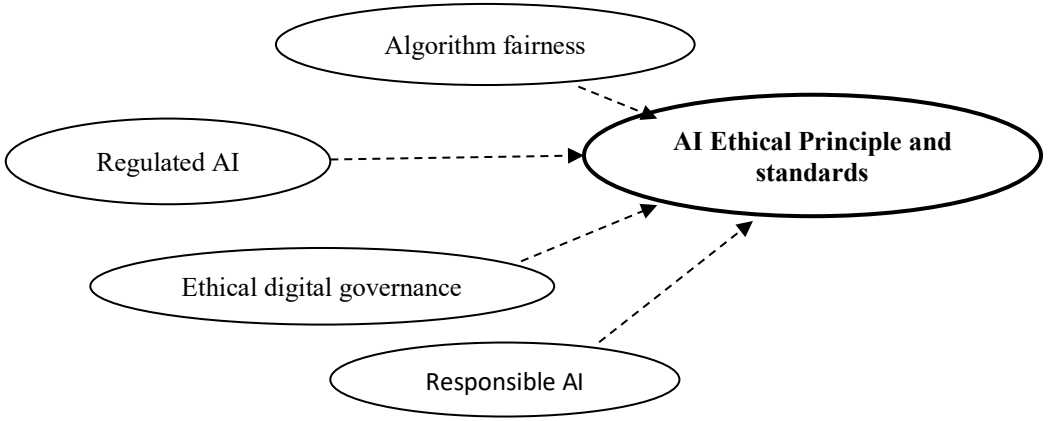

Figure 4: Propositional framework of the effect of ethical AI advancement on economic growth and social development (Source: Created by Authors)

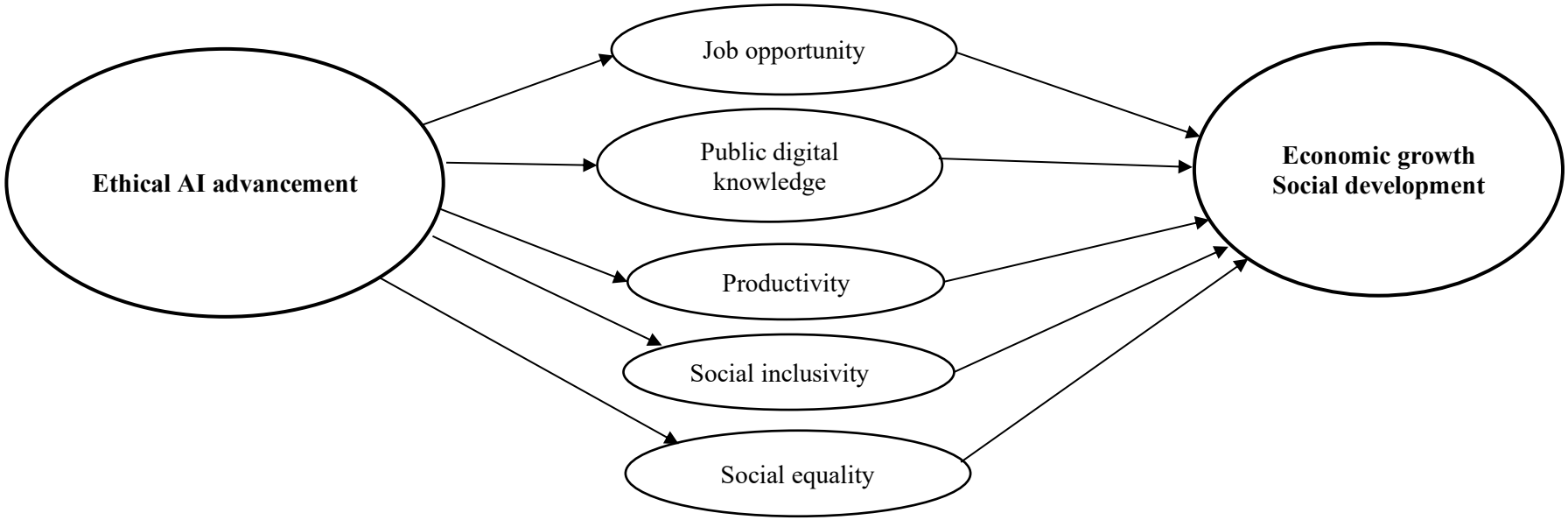

Figure 5- A propositional model of transition from disruptive AI to a sustainable digital economy and society (Source: Created by Authors)

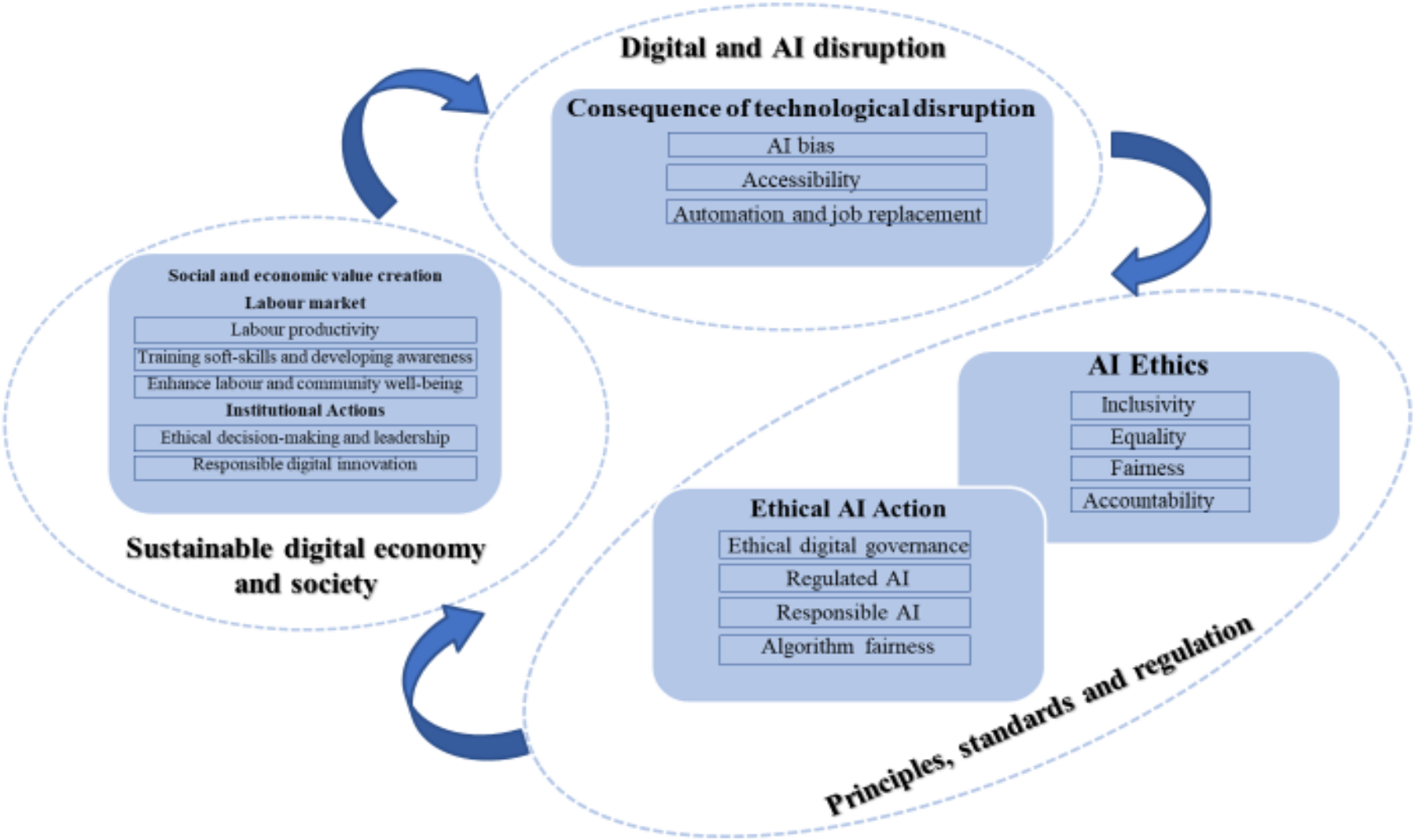

Table 1: Summary of reviewed highly cited articles in Ethics for AI from 2018 to 2022

Theme	Citation	Reference	Relevant Discipline	Future Research Questions	Highlights of the study
Adoption of ethical AI and best practise	185	(Kellogg et al., 2020)	Economic value of algorithms	What factors shape proprietors' use of AI-based systems negatively?	First, labour process theory helps to highlight potential problems with the largely positive view of algorithms at work. Second, the technical capabilities of algorithmic systems facilitate a form of rational control that is distinct from the technical and bureaucratic control used by employers for the past century. Third, employers' use of algorithms is sparking the development of new algorithmic occupations. Finally, workers are individually and collectively resisting algorithmic control through a set of emerging tactics we call algoactivism. These insights sketch the contested terrain of algorithmic control and map critical areas for future research.
				How can businesses design AI systems with an understanding of how different types of workers may have different needs?	
				How can businesses include employees and consumers' opinions in the design and implementation of AI technologies to improve user experiences and livelihoods?	
				How can businesses give their workforce the training they need to be able to use AI technology more efficiently?	
				In what ways can businesses use sectoral training to attract and keep workers who are skilled in AI while also helping them improve their long-term employment and income prospects?	
			Algorithms and control	How can employees who do "ghost work," like data curation, change or adapt algorithmic production technologies as they do their work?	

				How could articulation work be used to address algorithmic failure in a proactive way?	
				What are the full-scale consequences of implementing AI-based technologies without a "man in the middle" (human supervision in general)?	
				How does algorithmic opacity affect workforce identities and performance?	
				Is it inevitable that AI technologies foster a climate of anxiety, timidity, and frustration among the workforces?	
				How does existing case law about privacy rights and third-party tracking affect how algorithms are used in workplaces?	
Design and development of AI	79	(Kuo & Kusiak, 2019)	Sustainability of production systems	What are the ramifications of big data analytics concepts on sustainable business processing?	The review of the literature suggests that production research enabled by data has shifted from that based on analytical models to data-driven. Manufacturing and data envelopment analysis have been the most popular application areas of data-driven methodologies.
				How can AI-enabled technologies address environmental issues resulting from production processes?	

			Personalised production	What influence would AI-enabled manufacturing technologies have on customisation and manufacturing process implementation?	The research published to date indicates that data mining is becoming a dominant methodology in production research
Responsible and Accountable AI	74	(McClure, 2018)	Algorithm fairness	What is the best metrics to measure disproportionate fears of technological advancement in the area of workforce robotics and trusting AI to do the work of humans?	The study finds that there exists a sizable population of “technophobes" or those who fear robots, AI, and technology they do not understand. Technophobes are also more likely than non-technophobes to report having anxiety-related mental health issues and to fear unemployment and financial insecurity. With advances in robotics and AI, the threat of technological unemployment is discussed as a real concern among a substantial portion of the American population.
Design and development of AI	63	(Eling & Lehmann, 2018)	Strategic management, innovation	What exactly is digitalization? How will it impact the insurance industry?	The results illustrate four major tasks the industry is facing: enhancing the customer experience, improving its business processes, offering new products, and preparing for competition with other industries. Moreover, we identify three key areas of change with respect to insurability: the effect of new and more information on information asymmetry and risk pooling, the implications of new technologies on loss frequency and severity, and the increasing dependencies of systems through connectivity.
Adoption of ethical AI and best practise	54	(Boyd & Holton, 2018)	Strategic management, innovation	How can businesses adopt new technology from a less deterministic point of view that is aware of power and uncertainty?	This article provides a critical evaluation of how far recent innovations in robotics and artificial intelligence herald an unprecedented economic and social transformation, challenging the relentless technological determinism of much debate and reframing the issues involved within a political-economic and sociological approach. This focuses on the economic, political, and historical dynamics of technological innovation and its consequences for employment and economic restructuring, mediated through sovereign and discursive power. A range of epistemological and empirical problems with the transformationist position are identified, and an alternative perspective is proposed, emphasising complexity and uncertainty around contemporary and future trends.

Digital innovation and the 4.0 and 5.0 industrial revolution	54	(Caruso, 2018)	Strategic management, innovation	What sorts of effects would the initiative known as "Industry 5.0" have on the wider society?	Currently, firms are succeeding in making the second pole of these dichotomies (digital Taylorism, verticalisation, marketization, individualization) dominant on the first one (autonomy, participation, peer cooperation and socialisation of production). As it has always occurred in the history of the relationship between capital and labour, the possibility that the production process will shift in a direction favourable to labour mainly depends on the capacity for coalition and conflict and on the bargaining power of the latter. These elements develop within the labour relationship also thanks to the support of dynamics (political, cultural, organisational) and actors which are external to the production process, as the history of the workers' movements demonstrates (Bartolini 2000). Therefore, positive outcomes of 'Industry 4.0.' for workers will depend on social conflict and politics.
	51	(Fakhar Manesh et al., 2021)	Strategic management, innovation	How can artificial intelligence technologies transform new business innovation paradigms (e.g., inbound open innovation)?	The present article investigates the intellectual structure and trends of KM in Industry 4.0. Bibliometric analysis and a systematic literature review are conducted on a total of 90 relevant articles. The results reveal six clusters of keywords, subsequently explored via a systematic literature review to identify potential streams of this emergent field and future research avenues capable of producing meaningful advances in managerial knowledge of Industry 4.0 and its consequences.
Digital Transformative Capability				How can digital Transformation lead to a successful business strategy?	
				How can big data capabilities transform new knowledge into competitive advantage?	

Adoption of ethical AI and best practise	47	(Menard et al., 2018)	Privacy/Security	How can information security-related behaviours be used to look at possible differences between two individual traits, collectiveness, and psychological ownership of information?	In this study the survey design tests effects of collectivism and ownership on security decisions. Furthermore, collectivism affects intention and psychological ownership. Psychological ownership affects native PMT variables and intention, and there is a significant difference shown between U.S. and China sample populations.
Digital innovation and the 4.0 and 5.0 industrial revolution	46	(Garay-Rondero et al., 2020)	Strategic management, innovation	What constructs can be used to better understand the evolution of new organisational cultures and norms as a result of Industry 4.0 digitization?	Having achieved an exploration of the different conceptual frameworks, there is no compelling evidence of the existence of a conceptual SCM that incorporates the basic theoretical constructs and the new roles and elements of Industry 4.0. Therefore, the main components of Industry 4.0 and their impact on DSC Management are described, driving the proposal for a new conceptual model which addresses and accelerates a vision of the future of the interconnectivity between different DSCs, grouped in clusters in order to add value, through new forms of cooperation and digital integration.
Digital innovation and the 4.0 and 5.0 industrial revolution	41	(osonde, 2018)	Strategic management, innovation	How can digitisation help businesses and their workforce understand distributive justice concerns, such as providing them with resources to help them understand and evaluate other specialties that underpin their work?	This review concludes by arguing that technology and artificial intelligence are entwined with social relations, being sites of class struggle. How this is played out is an outcome of the balance of power, not only within the social formation but also globally. How far the development of the forces of production is compatible with capitalist relations is a moot point, as this is also a site of struggle. The paper draws out the implications for VET and considers progressive educational responses. However, such a practice needs to be set within a broader politics that is committed to the development of a society.

References:

Accenture. (2019). The Sphere of Control: Accenture 2019 Global Risk Management Study, Capital Markets Report. Available at: https://images.info.accenture.com/Web/ACCENTURE/%7B7c939083-a6b3-41c0-8709-7cd442c4cbfd%7D_Accenture-Global-Risk-Study-Capital-Markets-2019.pdf?elqcst=272&elqcsid=164

Acemoglu D., Restrepo P. (2017). Robots and jobs: Evidence from US labor markets. NBER Working Paper 23285 (National Bureau of Economic Research, Cambridge, MA).

Acemoglu D., Autor D. (2011) Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics* (Elsevier, Amsterdam), Vol 4, pp 1043–1171.

Agrawal, A., Gans, J. S., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. Cambridge, MA: *Harvard Business School Press*.

Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, *144*, 201-216.

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60, 102387.

Alpaydin, E. (2016). Machine learning: the new AI. *MIT press.*

Anderson, L. A., & Anderson, D. (2010). The change leader's roadmap: How to navigate your organization's transformation (Vol. 384). *John Wiley & Sons.*

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Chatila, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.

Askell, A., Brundage, M., & Hadfield, G. (2019). The role of cooperation in responsible AI development. *arXiv preprint arXiv:1907.04534.*

Ayaz, O., Tabaghdehi, S. A. H., Rosli, A., & Tambay, P. (2025). Ethical implications of employee and customer digital footprint: SMEs perspective. *Journal of Business Research, 188*, 115088.

Autor D. (2010). The polarization of job opportunities in the US labor market: Implications for employment and earnings (Center for American Progress and The Hamilton Project).

Balasubramanian, N., Ye, Y., & Xu, M. (2020). Substituting human decision-making with machine learning: Implications for organizational learning. *Academy of Management Review*, in press.

Barker, N., & Jewitt, C. (2022). Future touch in industry: Exploring sociotechnical imaginaries of tactile (tele) robots. *Futures, 136,* 102885.

Baum, J. A., & Haveman, H. A. (2020). Editors' comments: the future of organizational theory. *Academy of Management Review,* 45(2), 268-272.

Beijing AI principles [Internet]. Beijing: Beijing Academy of Artificial Intelligence; c2019. Available from: https://www.baai.ac.cn/blog/beijing-ai-principles.

Bosse, D. A., & Phillips, R. A. (2016). Agency theory and bounded self-interest. *Academy of Management Review*, 41(2), 276-297.

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

Bonyuet, D. (2020). Overview and impact of blockchain on auditing. *International Journal of Digital Accounting Research*, 20, 31-43.

Bowie, N. E., & Freeman, R. E. (1992). Ethics and agency theory: An introduction. *Oxford University Press.*

Bowie, N. E. (2002). A Kantian approach to business ethics. Ethical issues in business: A philosophical approach, 7, 61-71.

Brynjolfsson E., McAfee A. (2012). Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy (Digital Frontier, Lexington, MA).

Brynjolfsson E, McAfee A (2014) The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies (Norton, New York).

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society,* 3(1), 2053951715622512.

Burström, T., Parida, V., Lahti, T., & Wincent, J. (2021). AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research. *Journal of Business Research,* 127, 85–95.

Calo, R. (2021). Modelling through. *Duke LJ*, *71*, 1391.

Calp, M. H., & Bütüner, R. (2022). Society 5.0: Effective technology for a smart society. In *Artificial Intelligence and Industry 4.0* (pp. 175-194). Academic Press.

Cartwright, E. (2018). Behavioral economics. Routledge.

Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., & Abadie, A. (2022). AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research,* 144, 31–49.

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review,* 35(4), 410-422.

Cohen, A., & Baruch, Y. (2010). An agency theory perspective of the Israeli labor market segmentation: Past, present, and future. *Human resource management review, 20*(3), 186-193.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, August). Algorithmic decision making and the cost of fairness. In Proceedings of the 23rd international conference on knowledge discovery and data mining (pp. 797-806).

Cowgill, B., & Tucker, C. E. (2020). Algorithmic fairness and economics. Columbia Business School Research Paper. Available at SSRN: https://ssrn.com/abstract=3361280 or http://dx.doi.org/10.2139/ssrn.3361280.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Available online at: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Dawson, G. S., Denford, J. S., Williams, C. K., Preston, D., & Desouza, K. C. (2016). An examination of effective IT governance in the public sector using the legal view of agency theory. *Journal of Management Information Systems, 33*(4), 1180-1208.

Daugherty, P. R., & Wilson, H. J. (2018). Human+ machine: Reimagining work in the age of AI. Boston, MA: *Harvard Business Press*.

De Colle, S., & Werhane, P. H. (2008). Moral motivation across ethical theories: what can we learn for designing corporate ethics programs?. *Journal of Business Ethics*, 81(4), 751-764.

Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way. *Springer Nature.*

Dignum, V. (2017). Responsible artificial intelligence: designing AI for human values. *Digital Institutional Repository*, 1-8.

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, *20*(1), 1-3.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management,* 101994.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of management review*, *14*(1), 57-74.

Eitel-Porter, R. (2021). Beyond the promise: implementing ethical AI. *AI and Ethics,* 1(1), 73-80.

Eling, M., & Lehmann, M. (2018). The Impact of Digitalization on the Insurance Value Chain and the Insurability of Risks. Geneva Papers on Risk and Insurance: Issues and Practice, 43(3), 359–396.

European Parliament. (2020). Artificial Intelligence: Opportunities, risks and policy implications. European Parliamentary Research Service. Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf

European Commission (2019). Economic impacts of artificial intelligence (AI). Available at:https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf.

Faraj, S., & Leonardi, P. M. (2022). Strategic organization in the digital age: Rethinking the concept of technology. *Strategic Organization,* 20(4), 771-785.

Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180081.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2021). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Ethics, governance, and policies in artificial intelligence, 19-39.

Foresti, R., Rossi, S., Magnani, M., Bianco, C. G. L., & Delmonte, N. (2020). Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. *Engineering,* 6(7), 835-846.

Gal, U., Hansen, S., & Lee, A. S. (2022). Research perspectives: toward theoretical rigor in ethical analysis: the case of algorithmic decision-making systems. *Journal of the Association for Information Systems, 23*(6), 1634-1661.

Garay-Rondero, C. L., Martinez-Flores, J. L., Smith, N. R., Caballero Morales, S. O., & Aldrette-Malacara, A. (2020). Digital supply chain model in Industry 4.0. *Journal of Manufacturing Technology Management*, 31(5), 887–933.

Gautier, A., & Lamesch, J. (2021). Mergers in the digital economy. *Information Economics and Policy,* 54, 100890.

Ghosh, S., Hughes, M., Hodgkinson, I., & Hughes, P. (2021). Digital transformation of industrial businesses: A dynamic capability approach. *Technovation.*

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review,* 46(3), 534-551.

Gruetzemacher, R., & Whittlestone, J. (2022). The transformative potential of artificial intelligence. *Futures, 135*, 102884.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics*, *4*(37).

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines,* 30(1), 99–120.

Henry-Nickie, M., Sun, H. (2019). Skills and opportunity pathways: building an inclusive workforce for the future. Brookings Institution: Washington.

He ABY. (2017). How China is preparing for an AI-powered future. Wilson Center, Washington, DC.

Jaiswal, A., Arun, C. J., & Varma, A. (2022). Rebooting employees: Upskilling for artificial intelligence in multinational corporations. *The International Journal of Human Resource Management,* 33(6), 1179-1208.

Jensen, M., & Meckling, W. (1976). Theory of the firm: Managerial behaviour, agency costs and ownership structure. Journal of Financial Economics, 3(4), 305–360

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

Johnson, P. C., Laurell, C., Ots, M., & Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development – Automation or augmentation, exploration or exploitation? *Technological Forecasting and Social Change*, 179, 121636.

Kellogg, K. C., Valentine, M. A., Ang`, A., & Christin, A. (2020). ALGORITHMS AT WORK: THE NEW CONTESTED TERRAIN OF CONTROL Work and Organization Studies MIT Sloan School of Management. *Academy of Management Annals*, 2020(1), 366–410.

Kilian, K. A., Ventura, C. J., & Bailey, M. M. (2023). Examining the differential risk from high-level artificial intelligence and the question of control. *Futures, 151*, 103182.

Kim, B., Park, J., & Suh, J. (2020). Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information. *Decision Support Systems*, *134*, 113302.

Kohtamäki, M., Parida, V., Oghazi, P., Gebauer, H., & Baines, T. (2019). Digital servitization business models in ecosystems: A theory of the firm. *Journal of Business Research*, 104, 380–392.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America, 111(24), 8788–8790.

Kumar, V., Ramachandran, D., & Kumar, B. (2021). Influence of new-age technologies on marketing: A research agenda. *Journal of Business Research*, 125, 864–877.

Kuo, Y. H., & Kusiak, A. (2019). From data to big data in production research: the past and future trends. *International Journal of Production Research*, *57*(15-16), 4828-4853.

L'etang, J. (1992). A Kantian approach to codes of ethics. *Journal of Business Ethics*, 11(10), 737-744.

Langan, R., Cowley, S., & Nguyen, C. (2019). The state of digital marketing in academia: An examination of marketing curriculum's response to digital disruption. *Journal of Marketing Education,* 41(1), 32-46.

Leontief, W. (1952). Some Basic Problems of Structural Analysis. *The Review of Economics and Statistics,* 34(1), 1-9.

Li, L. (2022). Blockchain technology in industry 4.0. *Enterprise Information Systems,* 16(12), 2095535.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures, 90,* 46-60.

Martin, K. (2019). Designing ethical algorithms. *MIS Quarterly Executive*.

McClure, P. K. (2018). "You're Fired", Says the Robot: The Rise of Automation in the Workplace, Technophobes, and Fears of Unemployment. *Social Science Computer Review,* 36(2), 139–156.

Menard, P., Warkentin, M., & Lowry, P. B. (2018). The impact of collectivism and psychological ownership on protection motivation: A cross-cultural examination. *Computers and Security,* 75, 147–166.

Meijerink, J., & Keegan, A. (2019). Conceptualizing human resource management in the gig economy: Toward a platform ecosystem perspective. *Journal of Managerial Psychology, 34*(4), 214-232.

Mihalache, O. R., & Volberda, H. W. (2021). Business Model Innovation in Transforming Economies: A Co-evolutionary Perspective for a Global and Digital World. *Management and Organization Review*, *17*(2), 202-225.

Mittelstadt, B., Russell, C., & Wachter, S. (2019, January). Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 279-288).

Morris, N. (2009). Understanding digital marketing: marketing strategies for engaging the digital generation, *Journal of Direct, Data and Digital Marketing Practice*, 10, 384–387.

Moser, C., den Hond, F., & Lindebaum, D. (2021). Morality in the age of artificially intelligent algorithms. *Academy of management learning & education*, (ja).

Mujtaba, D. F., & Mahapatra, N. R. (2019). Ethical considerations in AI-Based recruitment. In 2019 IEEE International Symposium on Technology and Society (ISTAS) (pp. 1-7). IEEE.

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review*, 46(3), 552-571.

Naicker, V., & Van Der Merwe, D. B. (2018). Managers' perception of mobile technology adoption in the Life Insurance industry. *Information Technology & People, 31*(2), 507-526.

Newman, J., Mintrom, M., & O'Neill, D. (2022). Digital technologies, artificial intelligence, and bureaucratic transformation. *Futures, 136,* 102886.

A. Osoba and William Welser IV. (2017). An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. *Rand Corporation* Osonde.

Phan, P., Wright, M., & Lee, S. H. (2017). Of robots, artificial intelligence, and work. *Academy of Management Perspectives*, 31(4), 253-255.

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

Roselli, D., Matthews, J., & Talagala, N. (2019, May). Managing bias in ai. In Companion Proceedings of the 2019 World Wide Web Conference (pp. 539-544).

Rossi, F. (2015, September). Safety constraints and ethical principles in collective decision making systems. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)* (pp. 3-15). Cham: Springer International Publishing.

Russell, S., Hauert, S., Altman, R., & Veloso, M. (2015). Ethics of artificial intelligence. *Nature*, 521(7553), 415-416.

Sama, L. M., Stefanidis, A., & Casselman, R. M. (2022). Rethinking corporate governance in the digital economy: The role of stewardship. *Business Horizons, 65*(5), 535-546.

Schneier, B. (2018). Click here to kill everybody. New York: W. W. *Norton & Company.*

Schmidt, G. B., Philip, J., Van Dellen, S. A., & Islam, S. (2022). Gig worker organizing: toward an adapted Attraction-Selection-Attrition framework. *Journal of Managerial Psychology, 38*(1), 47-59.

Shestakofsky, B. (2017). Working algorithms: Software automation and the future of work. *Work and Occupations*, 44(4), 376-423.

Sugianto, N., Tjondronegoro, D., Stockdale, R., & Yuwono, E. I. (2021). Privacy-preserving AI-enabled video surveillance for social distancbartoliniing: Responsible design and deployment for public spaces. Information Technology & People.

Stonehouse, G. H., & Konina, N. Y. (2020, February). Management Challenges in the Age of Digital Disruption. In Proceedings of the 1st International Conference on Emerging Trends and Challenges in the Management Theory and Practice (ETCMTP 2019), Advances in Economics, *Business and Management Research* (Vol. 119, pp. 1-6).

Tabaghdehi, S. A. H., (2022). Connected tech: smart or sinister, https://committees.parliament.uk/work/6686/connected-tech-smart-or-sinister/publications/oral-evidence/

Tabaghdehi, S. A. H., & Foroudi, P. (eds). (2024). Business Strategies and Ethical Challenges in the Digital Ecosystem. Emerald Publishing Limited.

Tabaghdehi, S. A. H. (2024). Ethical Governance of Digital Footprint Data: A Journey Towards a Responsible Society. In *Business Strategies and Ethical Challenges in the Digital Ecosystem* (pp. 369-379). Emerald Publishing Limited.

Theres, C., & Strohmeier, S. (2023). Met the expectations? A meta-analysis of the performance consequences of digital HRM. *The International Journal of Human Resource Management*, 1-36.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. *Ethics, Governance, and Policies in Artificial Intelligence*, 97-123.

Tubadji, A., Huang, H., & Webber, D. J. (2021). Cultural proximity bias in AI-acceptability: The importance of being human. *Technological Forecasting and Social Change,* 173, 121100.

Varian, H. R. (2021). Seven deadly sins of tech? *Information Economics and Policy*, 54, 100893.

Van Noorden, R. (2020). The ethical questions that haunt facial-recognition research. *Nature*, 587(7834), 354-359.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97-112.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *The International Journal of Human Resource Management, 33*(6), 1237-1266.

Wachter, S. (2019). Data protection in the age of big data. *Nature Electronics*, 2(1), 6-7.

Westphal, J. D., & Zajac, E. J. (2013). A behavioral theory of corporate governance: Explicating the mechanisms of socially situated and socially constituted agency. *Academy of Management Annals, 7*(1), 607-661.

Wien, A. H., & Peluso, A. M. (2021). Influence of human versus AI recommenders: The roles of product type and cognitive processes. *Journal of Business Research,* 137, 13–27.

Wilkinson C, Weitkamp E. (2016). Creative research communication: theory and practice. Manchester, UK: *Manchester University Press.*

Wilsdon J, Willis R. (2004). See-through science: why public engagement needs to move upstream. Project Report. London, UK: Demos.

Winfield, A. 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), 20180085.

Wirtz, B. J. Weyerer, and Geyer C. (2019). Artificial Intelligence and the Public Sector - Applications and Challenges", *International Journal of Public Administration*, 42(7), 596-615.

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

Young, M. M., Himmelreich, J., Bullock, J. B., & Kim, K. C. (2019). Artificial intelligence and administrative evil. *Perspectives on Public Management and Governance, 4*(3), 244-258.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—it's time to make it fair. Nature, available on line: https://www.nature.com/articles/d41586-018-05707-8?source=post_page-----817fa60d75e