ETHICAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE INNOVATIONS. WHAT IS THE ROLE OF PUBLIC POLICY?

Beatriz Vassiliadis

Dissertation

Master's in Innovation and Technological Entrepreneurship

Supervised by

Prof. Joana Costa

2024

I declare that the present research proposal for my master's dissertation is of my own authorship and has not been previously used in another course or curricular unit of this or any other institution. References to other authors (statements, ideas, thoughts) scrupulously respect the rules of attribution and are duly indicated in the text and the bibliographical references, in accordance with the referencing rules. I am aware that the practice of plagiarism and self-plagiarism constitute an academic offense.

Acknowledgments

First, I would like to thank my supervisor, Prof. Joana Costa, for her invaluable guidance, support, and expertise throughout the entire research process.

I also would like to thank Prof. João José Pinto Ferreira for his mentorship and constructive feedback that have been instrumental in shaping this thesis and in my academic growth.

I am grateful to my parents for their unwavering support throughout my Master's journey in Innovation and Technological Entrepreneurship. Their encouragement has been instrumental in my success.

I also thank my friends and all MIET colleagues for their constant support, wise counsel, and motivation. Their presence and encouragement have been invaluable throughout this journey.

Abstract

This thesis delves into the ethical considerations surrounding the integration of artificial intelligence (AI) technologies in business operations. It explores the complex moral dilemmas arising from AI usage, including concerns about bias, privacy, transparency, accountability, safety, and societal impact.

This dissertation aims to propose guidelines (i.e., a framework) offering a comprehensive approach to navigating the ethical complexities of AI implementation in the workplace. Through an iterative development process involving literature reviews and real employees for feedback and validation, the framework emerges as a responsive and adaptable tool, grounded in practical experiences and real-world concerns.

Distinguishing itself from standardized approaches prevalent in existing literature, the proposed framework provides tailored, actionable measures for each identified ethical concern. By offering clear guidance and practical solutions, it empowers organizations to translate ethical principles into tangible practices, fostering a culture of responsible AI usage.

Despite the nascent nature of AI and its associated ethical discourse, this study navigates the rapid technological transformation with diligence and foresight. It fills critical gaps in the discourse surrounding AI ethics, offering a forward-thinking perspective on the ethical implications of AI adoption in the workplace.

Keywords: Artificial, intelligence, business, ethics, society.

Resumo

Esta tese investiga as considerações éticas que cercam a integração de tecnologias de inteligência artificial (IA) nas operações comerciais. Explora os complexos dilemas morais decorrentes do uso da IA, incluindo preocupações sobre preconceito, privacidade, transparência, responsabilidade, segurança e impacto social.

Esta dissertação tem como objetivo propor diretrizes (ou seja, uma estrutura) que ofereçam uma abordagem abrangente para navegar pelas complexidades éticas da implementação da IA no local de trabalho. Através de um processo de desenvolvimento iterativo que envolve revisões de literatura e funcionários reais para feedback e validação, a estrutura surge como uma ferramenta ágil e adaptável, baseada em experiências práticas e preocupações do mundo real.

Distinguindo-se das abordagens padronizadas predominantes na literatura existente, a estrutura proposta fornece medidas personalizadas e viáveis para cada preocupação ética identificada. Ao oferecer orientações claras e soluções práticas, capacita as organizações a traduzir princípios éticos em práticas tangíveis, promovendo uma cultura de utilização responsável da IA.

Apesar da natureza nascente da IA e do discurso ético associado, este estudo navega pela rápida transformação tecnológica com diligência e visão. Preenche lacunas críticas no discurso em torno da ética da IA, oferecendo uma perspectiva inovadora sobre as implicações éticas da adopção da IA no local de trabalho.

Palavras-chave: Inteligência, artificial, negócios, ética, sociedade.

Index

1. Introduction

In the rapidly evolving landscape of artificial intelligence (AI), advancements and innovations are reshaping the way businesses operate, providing them with unprecedented opportunities to enhance efficiency, optimize processes, and offer new, data-driven products and services. A recent survey has revealed that the adoption of workplace AI has not only changed 82% of job roles but has also necessitated a shift in the required skills (Zirar, A. et al, 2023).

However, the integration of AI into corporate operations brings with it a host of ethical challenges that demand immediate attention. For instance, despite 89% of workers utilizing AI, only 56% of companies educate their employees on its risks (Atwell, E., 2023).

On top of that, the potential for AI job displacement looms large, with 40%–50% of the workforce at risk, particularly in developing countries, as AI has the capacity to outperform workers (Fisher, E. et al, 2023). This emphasizes the pressing need for proactive measures in the face of these technological innovations.

As AI becomes increasingly integrated into business strategies, it has the potential to affect a multitude of factors, including individual privacy, fairness, transparency, and accountability. This transformation underscores the necessity of addressing the ethical implications of AI integration in corporate operations.

1.1 Dissertation Proposal

The title of this project is as follows:

Ethical implications of artificial intelligence innovations. What is the role of public policy?

As AI continues to advance in multiple dimensions of human lives, questions and concerns arise regarding its impact on privacy, security, and accountability.

The ongoing debate on the boundaries of the transition must balance innovation and the protection of human values, ultimately shaping the trajectory of technological change in a responsible and inclusive manner.

The purpose of the work is to find evidence in the extant literature on which is the most effective path for policy actions taking place, accounting for their costs and benefits.

1.2 Research Motivation

In this context, the motivation for this research is to understand and analyze the central ethical challenges that corporations face as they explore the innovative opportunities that AI presents. This exploration extends to identifying potential areas of concern, assessing the need for ethical considerations, and defining the critical role of public policy in fostering responsible AI integration.

The ethical landscape of AI in corporate settings is multifaceted, and the dilemmas it poses are complex. Understanding these challenges and determining how innovation can effectively mitigate them is essential. This research aims to focus on this relationship and provide insights into how companies can navigate these challenges responsibly.

With this in mind, the central initial research question that drives this investigation was elaborated as follows: "What are the main ethical challenges that companies face as a result of the innovation opportunities involving the integration of AI into their operations? ".

Addressing this question is of utmost importance in the context of this dissertation, as it sets the stage for a comprehensive examination of the complex relationship between AI innovation and ethical considerations in the corporate environment.

2. Literature review

This literature review seeks to delve into the ethical challenges that surface when companies adopt the power of AI in their day-to-day activities. The primary focus here is to gain a comprehensive understanding of the ethical dilemmas encountered by companies as they embrace AI technology.

Navigating the AI terrain while upholding ethical standards is no straightforward task. It requires companies to balance between reaping the benefits of AI and shouldering the responsibility of using it in a manner that respects the rights and values of individuals.

This review serves as a window into the real-world concerns faced by businesses, policymakers, and society at large. Through examination of existing research, this study intends to highlight the challenges associated with AI adoption and explore potential solutions and recommendations aimed at mitigating these challenges within the corporate world.

2.1 Methodology

To gather relevant research on the topic, a systematic search was conducted using specific keywords. In this order, the keywords employed in the databases library encompassed 'artificial intelligence', 'corporate', and 'ethics'. The two main platforms utilized for this search were Web of Science (WOS) and SCOPUS, apart from the seminal articles found in Research Gate and ArXiv platforms.

The search retrieved a comprehensive response, retrieving 212 articles across both platforms, WOS and SCOPUS. Specifically, 108 articles were identified in SCOPUS, while 104 were sourced from WOS (Figures 1 and 2, respectively). The first date that this search took place on both platforms was on the 10th of October.

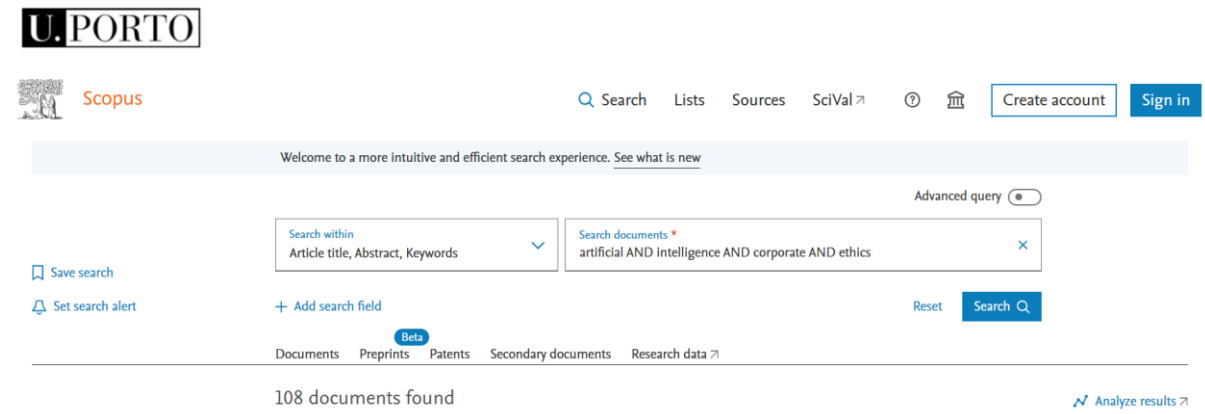

Figure 1: A screenshot of the retrieved bibliography research results via the SCOPUS platform utilizing the chosen keywords (10th October 2023). Keywords: 'artificial', 'intelligence', 'corporate', and 'ethics'.

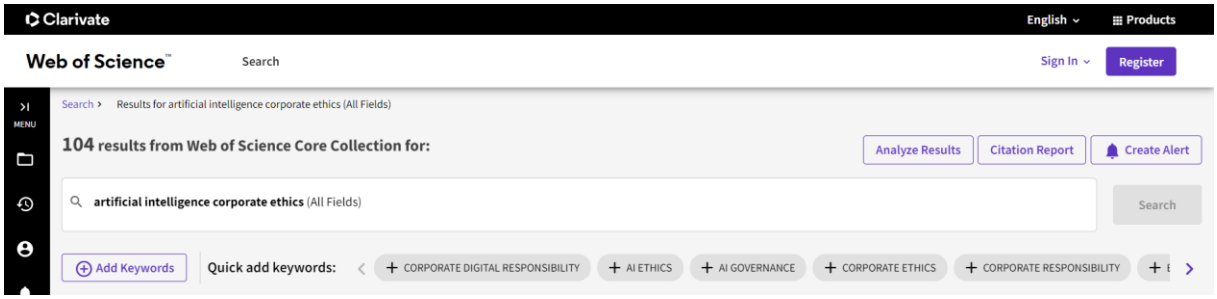

Figure 2: A screenshot of the retrieved bibliography research results via the WOS platform utilizing the chosen keywords (10th October 2023). Keywords: 'artificial', 'intelligence', 'corporate', and 'ethics'.

This method ensured the compilation of a wide-ranging collection of literature to inform the subsequent review. SCOPUS and WOS were considered as the main tools for search papers as both platforms are widely recognized, reputable databases known for their extensive coverage of academic literature. By focusing on these platforms, the aim is to ensure the inclusion of high-quality, peer-reviewed articles relevant to this research topic. This decision is justified by the desire to prioritize accuracy and scholarly rigor over quantity, thus contributing to the robustness of this literature review (Machi, L. A., & McEvoy, B. T., 2016).

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology flowchart was used to guide the study selection process. Figure 3 provides an overview of the steps taken to identify, screen, and include relevant studies.

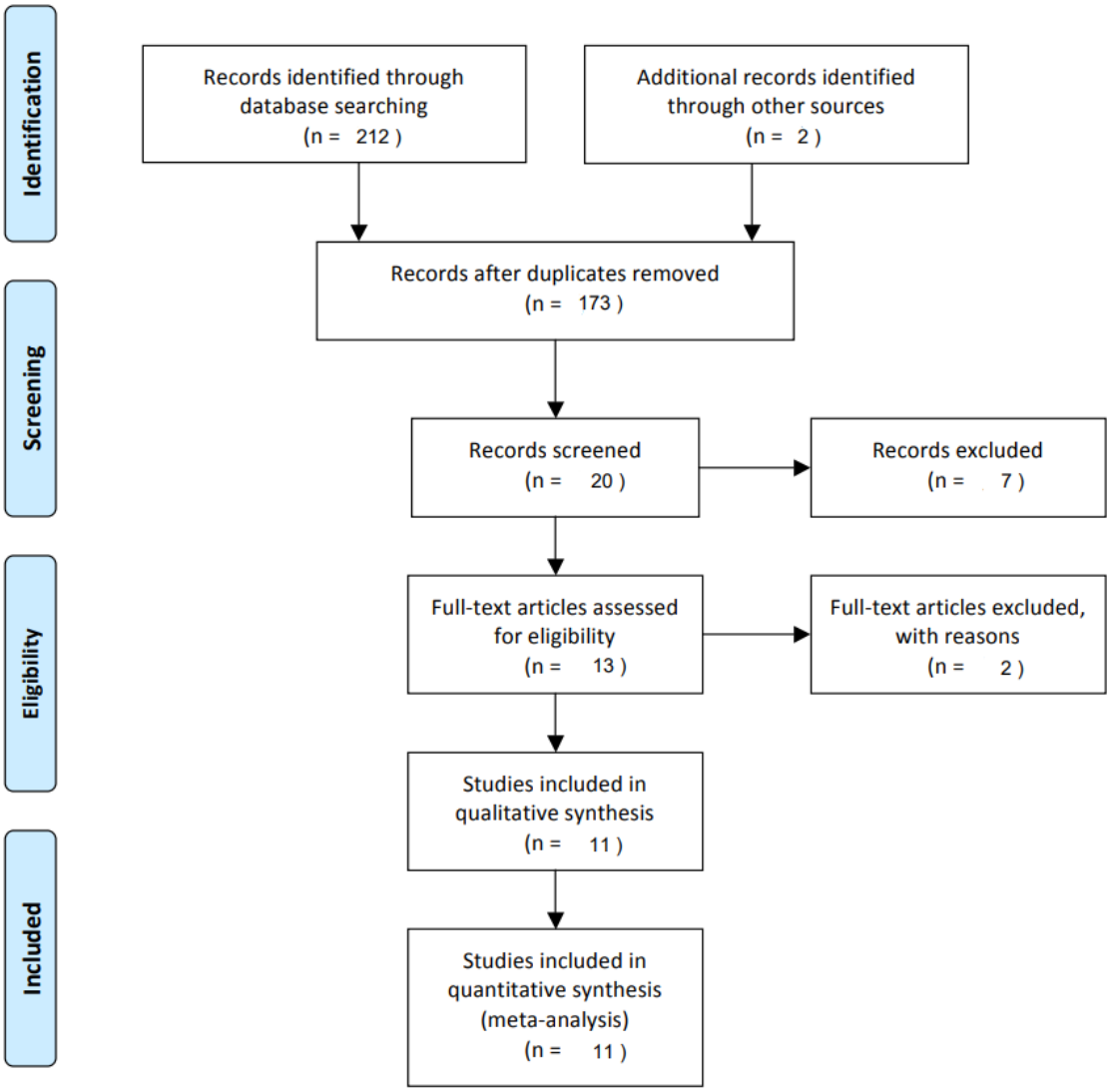

Figure 3: PRISMA flowchart elaborated to select the papers to be studied.

From the initial 214 papers identified across the mentioned platforms, 41 were duplicates found in more than one database, resulting in 173 unique articles.

In the screening phase, 20 articles were selected for further assessment. Each article's abstract was analyzed to determine its relevance to the research focus. During this stage, seven articles were excluded. Moving to the eligibility assessment, thirteen of the remaining fifteen articles were fully studied. A few articles were excluded from the analysis as they did not align closely with the study's primary objectives, that is, the predefined inclusion criteria (i.e. not addressing the specific ethical challenges in AI integration within companies). For the in-depth analysis phase, eleven articles were fully reviewed.

2.2 Literature Review Table

In this chapter, all the key literature in the course of this work will be analysed. First, two seminal articles, Vincent C. Müller (2020) and Brundage, M., et al (2018), have profoundly influenced the discourse on ethical challenges posed by AI in corporate settings. The articles were found in Research Gate and ArXiv platforms, respectively.

Vincent C. Müller (2020) provides a comprehensive framework for understanding AI ethics within corporations. It explores ethical dilemmas spanning machine ethics, legal considerations, and societal impacts. The paper advocates for transparent and responsible AI development practices, emphasizing the ethical responsibilities of companies.

Brundage, M., et al (2018), explore the malicious applications of AI, forecasting potential impacts and proposing preventive measures. It underscores the responsibility of companies to ensure ethical AI development and deployment.

These seminal papers form the cornerstone of this study, providing rich insights into the ethical complexities of AI integration in corporate contexts. Their thorough analysis and detailed exploration influenced the selection of this dissertation's theme, laying the groundwork for further investigation.

Furthermore, the following table (Table 1) serves as a structured overview of the key research articles analyzed in this literature review (i.e. the nine articles selected in the in-depth analysis phase explained in the previous section). Each article is accompanied by essential

information and insights relevant to the central research question concerning the ethical challenges, innovative AI strategies, and the role of public policy in AI integration within corporate operations.

These elements, presented in each column of table 1, are relevant because they provide a structured and comprehensive overview of the key aspects related to the research question and the broader context of AI integration within corporate operations.

The 'Key Findings and Relevance to the Research Question' column highlights the main findings of each article and how those findings relate to the central research question. It allows for a quick assessment of the article's alignment with the research objectives. The 'Ethical Challenges Addressed' column identifies the specific ethical challenges discussed in each article, which is crucial for categorizing and understanding the ethical dimensions relevant to AI integration.

The 'Innovative AI Strategies' column reveals the innovative approaches, strategies, or solutions proposed or discussed in the articles to address the identified ethical challenges. It provides insights into potential best practices. The 'Role of Public Policy' section is essential for grasping how governments and regulatory bodies impact the ethical considerations of AI in corporate operations.

Lastly, the 'Limitations and Future Research' section highlight a final personal conclusion on any limitations or areas for future research based on the specific article studied. As a matter of connecting with the gaps and opportunities for further exploration in the field.

In summary, these columns were chosen/constructed to collectively offer a comprehensive view of the literature, allowing for the systematic organization of information and insights relevant to the ethical challenges, innovative strategies, and the role of public policy in AI integration within corporate operations. This structured approach aids in synthesizing and comparing findings across the different sources, contributing to a more in-depth and well-rounded literature review.

Table 1: Elaborated literature review table of the nine chosen articles.

Articles	Key findings and relevance to the research question	Ethical challenges addressed	Innovative AI strategies	Role of public policy	Limitations and Future Research
Kriebitz, A., Lütge, C. (2020).	The article highlights potential conflicts between AI and human rights, stressing the importance of ethical AI development. It emphasizes the obligation of companies to prevent AI from violating human rights. Keywords: artificial intelligence, corporate ethics, data privacy, digitization, human rights.	- Bias and discrimination; - Privacy and Data Protection; - Accountability and transparency; - Safety and security; - Democratic values; - Employment and economic disruption.	Autonomous driving, facial recognition technology, and predictive policing.	- Ethical AI promotion through government oversight; - Accountability for AI's society impacts; - Aligning AI with public policy for the greater good.	- Incomplete practical guidance on implementing a human rights due diligence approach in AI; - How can companies address the challenge of full transparency?; - Calls for further research to fully understand AI-human rights relationships and protective strategies - solutions presented are too diverse.
Krkac, K. (2019).	This paper discusses reframing human notions of AI responsibility in corporate social irresponsibility. It investigates the ethical responsibility of robots and AI, highlighting the importance of ethical guidelines. Keywords: Business ethics, Artificial intelligence, Corporate social irresponsibility, Human irresponsibility, Robot irresponsibility.	- Socially irresponsible use (autonomous weapons or perpetuating bias); - Defining AI responsibility and CSR challenges, requiring ethical guidelines; - Ensuring AI transparency, accountability, and fair decision-making; - AI's impact on jobs, with ethical concerns about displacement and inequality; - Balancing AI benefits with risks and aligning	Human-centered AI.	- Shaping ethical AI practices, emphasizing collaborative efforts among stakeholders; - Facilitate discussions and collaboration among diverse stakeholders to address AI's complex ethical challenges.	- It does not provide a detailed discussion of the role of public policy in addressing the ethical challenges of AI; - It acknowledges that the pace of technological change is rapid, which can make it difficult to keep up with the ethical implications of new technologies;

Articles	Key findings and relevance to the research question	Ethical challenges addressed	Innovative AI strategies	Role of public policy	Limitations and Future Research
		development with human values.			- There is no one-size-fits-all solution to these challenges.
Hickman, E., Petrin, M. (2021).	The paper outlines the EU's Ethics Guidelines for Trustworthy AI, based on four ethical principles. It discusses how AI systems have to meet these guidelines to be deemed trustworthy. Keywords: Corporate governance, Artificial intelligence, Company law, Ethics, European Union.	Human agency and oversight, safety, and security, privacy and data protection, lack of transparency, bias, and discrimination.	It presents the European Commission's AI strategy (capacity, socio-economic change, and ethical and legal framework).	- EU's Ethics Guidelines for trustworthy AI; - Policymakers should develop regulations for AI's transparency and accountability.	- Policymakers and regulators may have limited resources to devote to addressing the ethical challenges of AI; - AI systems can be complex and difficult to understand; -There is often a lack of consensus on what ethical principles should guide the development and use of AI.

Articles	Key findings and relevance to the research question	Ethical challenges addressed	Innovative AI strategies	Role of public policy	Limitations and Future Research
Roemmich, K., Rosenberg, T. I., Fan, S., & Andalibi, N. (2023).	Ethical challenges in AI hiring, include demographic bias and the effectiveness of local legislation in preventing AI hiring discrimination. The authors call for ongoing deliberation on AI's ethical role in hiring and the values we want for our future. Keywords: Artificial intelligence, future of work, affective computing, emotion recognition.	Demographic bias, legislative effectiveness, and the potential for local services claims to obscure harms and reinforce exclusionary practices.	AI strategies used in hiring (Elevatus, Bob, and iMocha).	- The UK's data protection laws and regulatory methods; - Policymakers to address AI hiring vendors' transparency and ethical implications.	- Implicit racial biases are more difficult to detect and mitigate; - The effectiveness of existing legislation to prevent discriminatory effects of AI in hiring is limited.
Nyenno, I. (2023).	Harmonizing ethics, legality, and management performance amid AI impact, addressing the skills gap, and integrating ethics into AI R&D. It stresses a human-first approach in AI, focusing on people, solidarity, safety, and freedom of choice. Keywords: AI, AI personality, artificial intelligence, ethics; management, managerial work	Adhering to ethical standards and EU law, prioritizing a human-centric approach to AI development, and promoting responsibility in the technological age for human survival and the planet's future.	- AI for augmented innovation of enterprises; - Workforce automation.	- Harmonize ethical, legal, and management performance; - Ethical guidelines and codes of conduct for AI developers and users; - Human-first approach to AI development.	- Assessment of the negative impacts of AI on employment and the labor market (i.e. replacement of human-kind work); - Environmental impacts of AI.

Articles	Key findings and relevance to the research question	Ethical challenges addressed	Innovative AI strategies	Role of public policy	Limitations and Future Research
Zhao, J., Fariñas, B. (2022).	Potential for bias and discrimination, lack of transparency and accountability, and the need for responsible and ethical decision-making. However, it also highlights the potential benefits of using AI to promote more socially responsible companies and achieve sustainable decisions. Keywords: artificial intelligence, sustainable decisions, regulation, risk-based approach, company law.	- Automated bias and conflicts with human ethics; - Infringement on privacy and additional algorithmic bias; - Potential conflicts with human ethics and exacerbation of inequalities in society.	- AI training; - Cybersecurity.	The authors propose a proactive regulatory framework and rigorous corporate policies to ensure accountable and sustainable AI.	- Lack of in-depth analysis of specific case studies or real-world examples where AI has been successfully used to address sustainability challenges in companies. - Examination of the potential long-term societal and environmental impacts of AI implementation in corporate decision-making processes.
Wirtz, J., et al (2023).	- The era of self-regulation in AI ethics is over; legal frameworks are now imperative. - AI's predictive power raises ethical concerns, enabling bias-driven manipulations without awareness or oversight. Keywords: corporate digital responsibility, artificial intelligence, data, ethics, privacy, fairness.	Coercion of data disclosure, threat to human dignity, social engineering, violations of consumer privacy, and fairness concerns related to AI in service contexts.	- Predictive analytics; - Automated decision-making; - Machine learning.	- Corporate Digital Responsibility (CDR) as a framework for guiding service firms in their ethical, fair, and protective use of data and technology.	- Limitations of self-regulation in addressing ethical challenges related to AI, suggesting the need for legal intervention; - Lack of research on the practical implementation of CDR principles in service firms and the potential impact of regulatory interventions on managing ethical challenges

Articles	Key findings and relevance to the research question	Ethical challenges addressed	Innovative AI strategies	Role of public policy	Limitations and Future Research
					related to AI in service contexts.
Dr. Varsha P. S. (2023).	Different types of biases, including cognitive bias and incomplete data, it emphasizes the importance of responsible AI in firms to reduce the risk of biases and discrimination. Keywords: Artificial intelligence, bias, vulnerabilities, responsible AI, AI ethics, AI systems.	AI bias: gender, racial discrimination, recruitment.	General AI.	- Establishment of policies and techniques to detect and mitigate bias in AI systems; - Fact-based discussions with respect to human decision-making biases.	Lack of analysis of the potential impact of AI biases on different stakeholders, such as customers, employees, and society as a whole.
Lysova, E.I., et al (2023).	It discusses concerns about technological advancements potentially leading to a world without work and the impact of technology on employee autonomy and meaningful work. Keywords: meaningful work, business ethics, future of work, artificial intelligence, corporate social responsibility.	- Economic inequality; -Unequal treatment of workers in different jurisdictions; - Structural injustices exacerbated by the COVID-19 pandemic.	General AI.	- Development of regulatory frameworks; - Development of policies that protect workers' rights and promote fair treatment in the workplace.	It could potentially explore more in-depth discussions on the practical implementation of ethical frameworks in organizations, the role of leadership in addressing ethical challenges, and the long-term societal implications of the changing landscape of work and technology.

The subsequent subtopics of this chapter have the intention to cross-analyze the insights obtained in the above table and provide conclusions.

2.2.1 Foundational Ethical Considerations

Kriebitz and Lütge (2020) lay the groundwork by examining potential conflicts between AI and human rights, establishing an essential ethical foundation for subsequent discussions. This exploration is further reinforced by Krkac (2019), who emphasizes the necessity of ethical guidelines and responsible AI practices, setting a critical tone for further examination.

2.2.2 Regulatory Frameworks and Trustworthy AI

Hickman and Petrin (2021) introduce the concept of trustworthy AI through the EU's Ethics Guidelines, aligning with Roemmich et al.'s (2023) emphasis on transparency and accountability in AI hiring. These works underscore the importance of regulatory frameworks in ensuring ethical AI practices within corporate environments.

2.2.3 Interrelation of Ethics, Legality, and Management

Nyenno (2023) expands the discussion by highlighting the interconnectedness of ethics, legality, and management, echoing Zhao and Fariñas' (2022) call for a proactive regulatory framework to address bias and discrimination. These contributions provide a comprehensive perspective on responsible AI adoption, emphasizing the multifaceted nature of ethical considerations.

2.2.4 Shift Towards Legal Intervention

Wirtz et al. (2023) signal a shift towards legal intervention, suggesting a departure from self-regulation. This aligns with Kriebitz and Lütge's (2020) observation that the era of self-regulation is over, indicating a growing recognition of the need for legal frameworks to govern AI practices.

2.2.5 Societal Implications and Human-Centric Approaches

Dr. Varsha P. S. (2023) and Lysova et al. (2023) explore the societal implications of AI integration, with a focus on biases and the impact on employment. While Dr. Varsha P. S. advocates for responsible AI practices, Lysova et al. bridge ethical considerations with

broader societal implications, highlighting the link between AI adoption with socioeconomic factors.

2.2.6 Limitations and Areas for Further Research

Despite the valuable insights provided by these works, it is essential to acknowledge their limitations. Kriebitz and Lütge (2020) call for further research to fully understand AI-human rights relationships, while Krkac (2019) highlights the challenges of keeping up with the ethical implications of rapid technological advancements. Moreover, the lack of consensus on ethical principles, as noted by Hickman and Petrin (2021), poses a significant challenge in the development and use of AI. These limitations collectively point to areas where additional research is needed to enhance the understanding of the ethical complexities surrounding AI integration in corporate settings.

2.2.7 Conclusion

The extensive literature review conducted on the ethical challenges of integrating AI into corporate settings has provided valuable insights into the major issues and considerations surrounding this complex topic. Across the various articles examined, a consistent pattern emerges, highlighting common ethical concerns such as bias, privacy, transparency, and accountability. The authors emphasize the imperative of addressing these challenges in AI adoption within corporate environments.

While each paper contributes unique perspectives and insights, the overarching ethical dilemmas remain strikingly similar, indicating a shared understanding of the fundamental issues at stake. Despite the abundance of ethical concerns raised, a common thread throughout the literature is the absence of a universal guide or set of instructions for navigating the ethical complexities of AI integration in the workplace. While some papers propose precautionary measures there is a notable lack of consensus on standardized protocols or frameworks applicable across diverse organizational contexts.

This gap in the literature underscores the evolving nature of the AI landscape and the early phase of research in this domain. As AI continues to advance and permeate various aspects

of corporate operations, the need for comprehensive and universally applicable ethical guidelines becomes increasingly urgent. However, the relative novelty of the topic may explain the absence of a definitive framework at this stage. The ongoing discourse and research efforts in AI ethics are essential for addressing these gaps and establishing robust ethical standards to guide responsible AI adoption in corporate settings.

2.3 Research Design

In light of the insights obtained from the comprehensive literature review, a refinement of the central research question was also obtained. The refined research question now encapsulates an understanding of the ethical intricacies surrounding the integration of AI into corporate operations, as well as the imperative for innovative strategies and the pivotal role of public policy.

The revision was prompted by the identification of a significant gap in the existing literature, specifically in addressing the complex challenges of bias, privacy, transparency, accountability, safety, and societal impact within the context of AI incorporation into corporate settings. The revised research question, therefore, seeks to fill this critical gap by exploring how ethical considerations shape the incorporation of AI in corporate settings and what innovation strategies, alongside requisite public policy measures, are essential for addressing these multifaceted challenges. The revised research question reads:

"How do ethics and innovation influence AI integration in business, addressing bias, privacy, transparency, accountability, safety, and societal impact?".

Hevner's Design Science is a research design model commonly used to guide the creation and evaluation of innovative artifacts. It focuses on designing practical solutions to real-world problems and emphasizes the importance of rigor and relevance in the research process (Hevner, A. et al, 2004).

Hevner's research design model was used in this study to better assist the research processes and to establish and envision a practical solution to real-world corporate problems in the context of this study. Figure 4 below illustrates Hevner's model by the research design of this project.

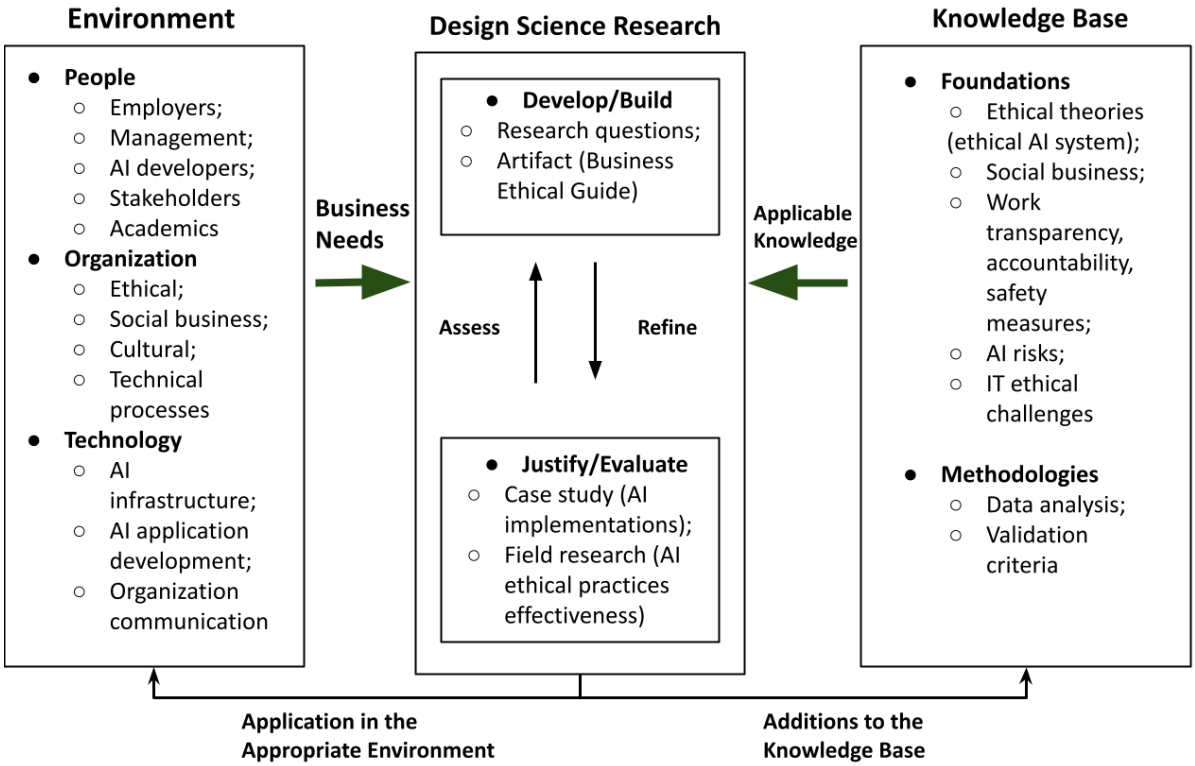

Figure 4: Hevner's framework to assist research and establish an artifact for this project.

The proposed artifact for addressing ethical challenges in AI integration within companies takes the form of a comprehensive framework or toolkit. This practical resource offers ethical principles, strategies, and best practices to guide organizations in responsibly adopting AI in their operations. It has a significant impact on individuals, organizational culture, and technology to guarantee the ethical integration of AI. It is essential for formulating plans, cultivating an ethical culture, and optimizing procedures to give ethical issues priority while using AI.

Furthermore, the artifact is a consequence of ethical theories as well as an essential tool for developing ethically grounded AI systems. Beyond theory, its usage allows for thorough

evaluations through field research and justification of moral actions through real-world case studies. It can be applied across multiple dimensions in the real world. Table 2 below illustrates how this artifact would serve as a solution for different sectors in the corporate environment.

Table 2: The proposed artifact's purpose for different sectors.

Sector	**Artifact's purpose**
Ethical framework implementation	The artifact serves as a practical guide for companies to implement ethical AI practices, ensuring that AI aligns with human rights, transparency, and fairness.
Policy development	By using the artifact, organizations may create ethical internal AI rules that promote responsible AI integration.
Employee training	The artifact serves as a valuable training resource, educating staff about ethical AI considerations and equipping them to make ethical decisions in their AI interactions.
Vendor and partner assessment	Companies can employ the artifact to evaluate the ethical practices of AI vendors and partners, enabling the selection of ethical collaborators.
Continuous improvement	The framework supports ongoing improvement efforts, allowing companies to assess their AI practices against ethical benchmarks and refine their strategies.

Through the integration of theoretical knowledge from case studies and practical examples from literature, this proposed artifact (i.e., a comprehensive framework or toolkit for ethical AI integration in companies) can be developed and made into an effective tool for organizations. It is a useful tool for responsible AI integration in business settings since it is based on academic ethical concepts and tested in real-world contexts.

3. Framing The Problem

In this section, it is possible to delve into a comprehensive understanding of the primary ethical challenges identified in the literature review (section 2 of this report). The consensus among authors highlights several key issues:

3.1 Prejudice and Unequal Treatment

Authors underscore the pervasive presence of prejudice and unequal treatment in the integration of AI systems. These biases manifest in various forms, including gender, racial discrimination, and unequal opportunities for different demographic groups.

3.2 Lack of Privacy and Transparency

A recurring concern highlighted in the literature is the lack of privacy and transparency surrounding AI technologies. This encompasses issues related to data privacy, algorithmic transparency, and the potential misuse of personal information.

3.3 Irresponsibility and Injustice

Ethical considerations often revolve around the concept of irresponsibility and injustice in AI implementation. This involves discussions on accountability for AI-driven decisions, fairness in algorithmic outcomes, and the ethical implications of AI systems' actions.

3.4 Security

Security emerges as a critical aspect of ethical AI adoption, with authors emphasizing the potential vulnerabilities and risks associated with AI systems. Concerns include data breaches, cybersecurity threats, and AI manipulation for malicious purposes.

3.5 Job Loss, Pay Inequality, and/or Financial Instability

The literature also reflects on the socioeconomic impacts of AI, particularly in terms of job displacement, pay disparities, and economic instability. The authors explore the ethical dimensions of these consequences and advocate for measures to address them.

3.6 Legal and Regulatory Concerns

Finally, legal and regulatory considerations play a pivotal role in the ethical discourse surrounding AI. The authors discuss the need for robust legal frameworks, regulatory oversight, and compliance mechanisms to ensure ethical AI practices.

4. Analysis of Ethical Frameworks and Guidelines

In this section, the analysis delves into existing ethical frameworks and guidelines pertinent to AI implementation in the workplace, focusing particularly on the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the European Commission's Ethics Guidelines for Trustworthy AI. By critically examining these frameworks and comparing them with the ethical challenges identified in the literature review, the aim is to assess their adequacy in addressing the complex ethical landscape of AI challenges outlined earlier.

4.1 Scope and Focus

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems is a comprehensive effort aimed at establishing ethical guidelines and standards for autonomous and intelligent systems. These guidelines, first launched in 2016, are designed to be applicable across a wide range of domains, including but not limited to artificial intelligence (AI).

The initiative brings together experts from diverse backgrounds to develop consensus-based ethical frameworks that address the complex ethical challenges posed by emerging technologies. By focusing on autonomy and intelligence, the IEEE Global Initiative aims to promote responsible and ethical design, development, and deployment of AI systems worldwide (Chatila & Havens, 2019).

In contrast, the European Union's Ethics Guidelines for Trustworthy Artificial Intelligence (AI), first launched in 2019, are tailored to the environment of corporate governance and company law. Unlike a comprehensive review of the European Commission's guidelines, this study focuses on their alignment with corporate governance frameworks and legal mandates affecting companies. Highlighting the implications for corporate entities, the analysis delves into how ethical AI principles can be harmonized with governance standards and integrated into corporate practices. Its goal is to furnish practical recommendations for companies struggling with AI adoption amidst the regulatory confines of company law (Hickman & Petrin, 2021).

4.2 Development Process

The development process of the IEEE Global Initiative involves collaboration among experts spanning various disciplines such as technology, ethics, policy, and academia. This initiative emphasizes a global outlook and seeks input from stakeholders worldwide to ensure inclusivity and relevance across diverse contexts (Chatila & Havens, 2019).

On the other hand, the European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence are crafted by the European Union's High-Level Expert Group on Artificial Intelligence (AI HLEG). This group comprises experts from academia, industry, and civil society within the EU. The guidelines are customized to align with the fundamental values, legal frameworks, and societal norms prevalent in the EU region (Hickman & Petrin, 2021).

4.3 Content and Structure

The IEEE Global Initiative offers a thorough framework that encompasses a wide array of ethical principles, standards, and best practices pertinent to the entire lifecycle of AI and autonomous systems. It addresses various ethical considerations such as transparency, accountability, fairness, and societal impact, providing a comprehensive guide for stakeholders involved in the design, development, and deployment of AI technologies (Economou N., 2019).

On the other hand, the European Commission's guidelines for trustworthy AI are structured around seven core requirements, each addressing specific aspects crucial for ensuring the trustworthiness of AI systems. These requirements include human agency and oversight, technical robustness, and safety, privacy and data governance, transparency, diversity, non-discrimination, and societal and environmental well-being. This structured approach provides clear guidance on key areas that AI developers and implementers need to prioritize to foster trust and ethical deployment of AI technologies (Smuha, N., 2019).

4.4 Geographical Coverage

The IEEE Global Initiative aims to establish universal ethical practices and standards for AI development worldwide by collaborating with experts from diverse backgrounds (Chatila & Havens, 2019). In contrast, the European Commission's guidelines are focused on addressing the unique needs of the European Union. While EU-centric, they may influence global AI ethics discussions due to the EU's significant role in setting ethical standards and regulations (Hickman & Petrin, 2021).

4.5 Enforceability and Legal Status

The IEEE Global Initiative offers voluntary guidelines and recommendations for ethical AI practices, lacking legal enforceability. Its influence relies on adoption by stakeholders and industry players (Adamson, G. et al, 2019).

On the other hand, the European Commission's guidelines, although initially non-binding, can inform future regulatory efforts within the EU. Proposed regulations like the AI Act aim to enforce specific aspects of AI ethics and governance, indicating a potential shift towards legal enforceability in the EU (Ebers, M., 2021).

4.6 Stakeholder Engagement

The IEEE Global Initiative prioritizes inclusivity and collaboration by involving stakeholders from academia, industry, government, and civil society. This diverse participation ensures a comprehensive range of perspectives and expertise, contributing to the development of well-rounded guidelines (Adamson, G. et al, 2019).

Alternatively, the European Commission engages with stakeholders within the EU, such as policymakers, industry representatives, researchers, and advocacy groups. This engagement facilitates the gathering of input and ensures alignment with EU values and priorities, making the guidelines reflective of the region's specific needs and objectives (Ebers, M., 2021).

4.7 Disadvantages of the Ethical Frameworks and Guidelines

The disadvantages of the IEEE Global Initiative include the lack of enforcement, as voluntary guidelines, there is no mechanism for ensuring compliance, potentially leading to inconsistent adoption. Also, the limited global impact, the guidelines may not align with all regional regulations and cultural norms, limiting their influence in certain areas. The complexity of the comprehensive framework may be challenging to interpret and implement, especially for smaller organizations. It is resource-intensive, its implementation may require significant time, expertise, and financial investment. Lastly, the potential bias, despite efforts for inclusivity, biases in expert groups and stakeholders may still exist (Shahriari, K. and Shahriari, M., 2017).

The disadvantages of the European Commission's Guidelines include the lack of flexibility, the prescriptive nature that may hinder adaptability to evolving technology, and ethical challenges. The EU-centric focus is tailored to EU values and legal frameworks, potentially limiting global relevance. Also, the regulatory burden, compliance may impose administrative and financial burdens, especially on smaller businesses. The uncertain legal status and ambiguity surrounding legal status and enforceability raises questions about practical impact. Lastly, the potential overregulation and stricter adherence could oppressive innovation and competitiveness in the European AI ecosystem (Larsson, S., 2020).

5. Proposed Artifact

The proposed artifact for this thesis is a new framework (with guidelines), based on the challenges identified in the literature review, that draws upon, and improves upon, the existing guidelines/frameworks analyzed, namely the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the European Commission's Ethics Guidelines for Trustworthy AI. This new framework was developed to address the specific ethical challenges identified in the literature review, with a focus on AI implementation in the workplace. Additionally, the proposed framework/guidelines were validated through feedback obtained from real employees, ensuring that it reflects the practical experiences and concerns of those directly affected by AI technologies in the workplace.

The analysis of both guidelines is crucial for developing a comprehensive framework that effectively addresses ethical considerations in AI implementation within the workplace. This comparative analysis allows for the identification of strengths and weaknesses in each approach, enabling the synthesis of the most effective elements while mitigating potential limitations. Moreover, understanding the specific features and focus areas of each existing guideline facilitates the tailored adaptation of the new framework to address the unique challenges and priorities relevant to the workplace environment. Additionally, involving real employees in the validation process ensures that the developed framework resonates with their practical experiences and concerns regarding AI technologies in the workplace, ultimately enhancing its relevance and effectiveness in real-world scenarios.

In developing the framework for this thesis, a comprehensive approach was taken to address the main ethical challenges outlined in Section 3. These challenges were particularly analyzed, and for each, a set of measurements was developed based on the analysis of both the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the European Commission's Ethics Guidelines for Trustworthy AI, and related study reviews of them. By synthesizing insights from these frameworks and tailoring them to the specific context of AI implementation in the workplace, a robust framework was formulated.

This process involved a systematic examination of ethical principles and practices relevant to AI ethics, ensuring that the framework adequately addressed the complex ethical

landscape of AI challenges identified earlier. Furthermore, the development process included iterative refinement and validation stages, wherein real employees provided feedback on the proposed measurements to ensure alignment with practical experiences and concerns in real-world workplace settings.

Through this rigorous approach, the framework was iteratively refined and optimized to enhance its relevance and effectiveness in addressing ethical challenges associated with AI implementation in the workplace.

5.1 Initial Development of the Artifact

In the development of the proposed framework for this thesis, a comprehensive approach was undertaken, drawing upon insights from existing guidelines and frameworks, namely the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the European Commission's Ethics Guidelines for Trustworthy AI. The objective was to address the specific ethical challenges identified in Section 3, focusing on AI implementation in the workplace.

To achieve this, a systematic analysis of each ethical challenge was conducted, and corresponding measurements were formulated based on the analysis of both frameworks (and study reviews of these materials). These measurements serve as actionable guidelines for addressing the identified ethical concerns in AI implementation.

The proposed artifact consists of eight key ethical considerations, each accompanied by a set of recommended actions:

1. Bias and Discrimination:

- Implement algorithms capable of detecting and mitigating bias and discrimination in decision-making processes by utilizing techniques such as fairness-aware machine learning algorithms and bias detection frameworks.

- Conduct regular audits on AI systems to identify and correct discriminatory trends, utilizing methodologies such as algorithmic impact assessments and bias audits to ensure fairness and equity in AI-driven decision-making.

2. Privacy and Transparency:

- Ensure transparency in the use of employee data by providing clear and accessible information about data collection practices, purposes, and processing methods.

- Obtain explicit and informed consent from employees for data collection and processing activities, adhering to principles of privacy by design and default.

- Adopt anonymization and pseudonymization practices to protect employee privacy during data processing, ensuring that personally identifiable information is appropriately masked or encrypted to prevent unauthorized access or disclosure.

3. Responsibility and Justice:

- Establish clear accountability mechanisms to ensure ethical and fair use of AI systems, defining roles and responsibilities for stakeholders involved in the design, development, and deployment of AI technologies.

- Create multidisciplinary ethics committees composed of representatives from various departments and disciplines to assess and monitor the impact of AI systems on work practices and employees, providing oversight and guidance on ethical decision-making processes.

4. Security:

- Implement robust security measures to protect AI systems against cyberattacks and malicious manipulations, including encryption, access controls, and intrusion detection systems.

- Conduct regular risk and vulnerability assessments on AI systems to identify and mitigate potential security threats, employing techniques such as penetration testing and threat modeling to proactively identify and address security vulnerabilities.

5. Economic Disruption:

- Develop 'change management' strategies to minimize the impact of AI-driven automation on job roles and skills, including workforce planning, reskilling, and redeployment initiatives.

- Invest in retraining and upskilling programs to empower employees to adapt to new market demands and acquire the skills necessary to work alongside AI technologies effectively.

6. Legal and Regulatory Concerns:

- Stay updated on laws and regulations related to AI ethics and usage in the workplace, including data protection regulations, labor laws, and industry-specific standards.

- Collaborate with regulatory bodies and government authorities to develop policies and guidelines promoting ethical AI usage, engaging in industry associations, and public-private partnerships to shape regulatory frameworks that balance innovation and protection.

7. Workers' Rights and Equity:

- Protect workers' rights by ensuring that AI systems are not used for discrimination or unfair treatment, implementing safeguards such as algorithmic fairness and anti-discrimination policies.

- Promote diversity and inclusion in the development and implementation of AI systems, ensuring equitable representation of different demographic groups in dataset collection, model training, and decision-making processes.

8. Technological Impact on Employment:

- Carefully assess the potential impact of AI adoption on job opportunities and working conditions for employees, conducting impact assessments and scenario analyses to anticipate and address workforce displacement.

- Implement mitigation measures such as job support programs, income redistribution policies, and social safety nets to address potential adverse effects on the workforce and ensure a just transition to the AI-enabled economy.

This framework/set of guidelines serves as a comprehensive roadmap for guiding the ethical and responsible implementation of AI systems in the workplace, prioritizing ethical values at every stage of the process.

6. Artifact Validation

For the validation of the proposed artifact, a quantitative approach is ideal for this thesis due to its ability to gather data from a large sample size. This allows for comprehensive insights into AI's ethical challenges. The numerical data is easily analyzed using statistical methods, identifying patterns and correlations. Quantitative findings can be generalized to broader populations and enable comparative analysis between different respondent groups. The efficiency of quantitative methods ensures timely data collection, while structured responses minimize researcher bias, enhancing the reliability and validity of the findings.

During the validation phase of this research, a questionnaire was circulated among groups of artificial intelligence workers on LinkedIn, a group of employees of a Brazilian company in the stock exchange sector, and a group of collaborators of a French insurance company. The purpose was to gain valuable insights into their perceptions and experiences regarding the ethical challenges discussed in Chapter 3, particularly their interactions with AI technologies in the workplace. Respondents were also invited to share feedback on the proposed artifact, expressing their agreement levels and providing suggestions for potential enhancements or adjustments.

In respect to security policies and confidentiality agreements within the participating organizations, the specific names of the companies cannot be divulged for this study, due to both companies' preferences of not disclosure. Some respondents chose to disclose their names and occupations, while others preferred to remain anonymous. This approach ensures participant privacy, allowing them to offer candid feedback without any concerns or repercussions. All collected data will be anonymized and aggregated to preserve confidentiality and uphold ethical research standards.

The decision to involve employees from this Brazilian stock exchange company was deliberate due to the organization's involvement with AI, offering valuable insights into the nuances of AI integration within the financial sector and highlighting industry-specific challenges and opportunities. Additionally, the collaborators of the French insurance company added a more entry-level experience with AI, as AI is particularly newly introduced

into the company. Furthermore, engaging professionals from diverse industries through LinkedIn ensures a diverse range of perspectives and experiences. Despite proactive outreach efforts to other companies, some did not respond, while others cited privacy concerns, time constraints, or personal reasons for declining participation.

6.1 Validation Methodology

For the validation of the research findings, a Google Forms questionnaire consisting of 13 questions was distributed. Among these questions were two non-mandatory inquiries, namely the respondent's name and occupation, and an additional query inviting participants to provide further details regarding the theme of the research. The questionnaire was disseminated to employees of the two companies and LinkedIn users, encompassing a diverse pool of respondents from different professional backgrounds.

The primary objective of this questionnaire was to validate whether the ethical challenges identified in the literature review corresponded with the real-life experiences of workers. If respondents did not agree with the ethical challenges outlined in the literature review (Section 2), it would have been necessary to revisit the process of identifying ethical AI workplace issues in the literature to justify such discrepancies. However, as anticipated, the responses from the participants aligned with the findings documented in the literature, affirming the relevance and applicability of the identified ethical challenges in practical work environments.

6.2 Questionnaire Overview

The following table explains the details regarding the 13 questions contained in the questionnaire that were used as an instrument of validation for this research.

Table 4: Explanation of the questionnaire questions of this research.

Order	Type of question	Question	Type of answer	Possible answers (any or more than one)
1	Optional	What is your name and occupation? (this step is optional, it is not necessary to provide such information)	Text	Name and occupation of the candidate or a blank response.
2	Mandatory	Do you agree that the following ethical challenges are relevant in the context of using artificial intelligence in the workplace? Please select the ones you agree with.	Checkbox (selection of multiple options)	• Prejudice and unequal treatment; • Lack of privacy and transparency; • Irresponsibility and injustice; • Insecurity; • Job loss, pay inequality, and/or financial instability; • Legal and regulatory concerns.
3	Mandatory	Of the options presented in the previous question, which one is the most important in your opinion?	Dropdown (selection of only one)	
4	Mandatory	Which one is the second most important in your opinion?		
5	Mandatory	Which one is the third most important in your opinion?		
6	Mandatory	Are there any other ethical challenges related to the use of artificial intelligence in the workplace that you would like to mention? (Please enter N/A if not applicable).	Text	Any other ethical challenge(s) according to the correspondents' opinion.
Below are proposals to mitigate the ethical challenges of using AI. Please rate them based on your opinion.				
7	Mandatory	(1) Implement AI algorithms that are capable of detecting and mitigating bias and discrimination in decision-making processes. (2) Conduct regular audits of AI systems	Rating scale between 1 to 5; 1 being considered of 'non-relevant' and 5 'very relevant'	Option to choose between 1 and 5.

Order	Type of question	Question	Type of answer	Possible answers (any or more than one)
		to identify and correct potential discriminatory biases.		
8	Mandatory	(1) Ensure full transparency in the use of employee data by obtaining explicit and informed consent for the collection and processing of personal data. (2) Adopt anonymization and pseudonymization practices to protect employee privacy during data processing.		
9	Mandatory	(1) Establish clear accountability mechanisms to ensure that AI systems are used ethically and fairly. (2) Create multidisciplinary ethics committees to evaluate and monitor the impact of AI systems on work practices and employees.		
10	Mandatory	(1) Implement robust security measures to protect AI systems against cyberattacks and malicious manipulations. (2) Conduct regular risk and vulnerability assessments on AI systems to identify and mitigate potential security threats.		
11	Mandatory	(1) Develop change management strategies to minimize the impact of AI-driven automation on worker roles and skills. (2) Invest in recycling and		

Order	Type of question	Question	Type of answer	Possible answers (any or more than one)
		retraining programs to enable employees to adapt to the new demands of the job market.		
12	Mandatory	(1) Ensure that AI is not used for discrimination or unfair treatment. (2) Promote diversity and inclusion in AI systems.		
13	Optional	Please share any additional comments, suggestions, or thoughts you have about the topic discussed in this form.	Text	Any additional information from the respondent regarding this research's theme.

To understand better the relevance of each question contained in the questionnaire, please consider for each, the explanation below:

Question 1: This question allows respondents to provide their name and occupation, although it is optional. While not required, this information can help in understanding the demographics and professional backgrounds of the respondents. However, respondents may choose to skip this question if they prefer to remain anonymous.

Question 2: In this question, respondents are asked to indicate their agreement with a list of ethical challenges relevant to AI usage in the workplace (i.e. the ethical challenges found in the literature review). Respondents can select multiple options from the list of challenges provided. This question aims to gauge the perceived relevance of these challenges among workers.

Questions 3 to 5: Respondents are required to select the 3 ethical challenges they consider the most important from the options provided in the previous question. These questions

help prioritize the 3 most important identified challenges based on their perceived significance in the workplace.

Question 6: Respondents are asked if they have any additional ethical challenges related to AI usage in the workplace that were not covered in the predefined list. This open-ended question allows respondents to express any unique or overlooked concerns they may have.

Question 7 to 12: Respondents are presented with the proposed mitigation measure (i.e. the first draft of the artifact contained in Section 5.1) for each ethical challenge identified earlier. They are then asked to rate the relevance of each measure on a scale of 1 to 5, with 1 indicating "non-relevant" and 5 indicating "very relevant." This question assesses the perceived effectiveness of proposed solutions in addressing the identified ethical challenges.

Question 13: Finally, respondents are allowed to share any additional comments, suggestions, or thoughts they may have about the research topic. This open-ended question allows respondents to provide qualitative feedback and insights beyond the structured questions in the questionnaire.

7. Results Obtained

The questionnaire was made available to participants for 3 weeks. Throughout this period, proactive reminders were dispatched via LinkedIn and within the collaborative groups affiliated with the participating companies, ensuring sustained engagement and participation.

In total, the questionnaire garnered responses from 74 participants, comprising 27 collaborators representing the stock exchange company, 15 collaborators from the French insurance company, and 32 individuals from LinkedIn. This diverse pool of respondents

offers valuable insights from both external industry professionals and internal stakeholders, enriching the depth and breadth of the research findings.

7.1 Profiling of the Respondents

The profiling of the respondents provides insight into the demographics and perspectives of individuals participating in the questionnaire. This section aims to categorize respondents based on their occupation, concerns regarding AI ethics, and prioritization of ethical challenges.

7.1.1 Brazilian Stock Exchange Company Collaborators

The company's collaborators represent a diverse array of professionals spanning various organizational functions. Among the 27 respondents, 10 disclosed their roles within the company, encompassing positions such as IT consultants, analysts, programmers, web designers, and managers.

Their firsthand experiences and perspectives offer invaluable insights into the ethical considerations surrounding AI deployment within their specific domains. Furthermore, it is worth noting that the company has been actively utilizing AI tools since last year, with widespread adoption and continuous promotion through avenues such as lectures and training sessions.

Notably, the company has forged a partnership with Microsoft, establishing itself as one of the pioneering organizations in Brazil to leverage Copilot and corporate ChatGPT.

7.1.2 French Insurance Company Collaborators

The collaborators from the French insurance company encompass a diverse mix of professionals primarily concentrated in business analysis, IT consultancy, and programming roles offering key insights into the ethical dimensions of AI integration within their respective domains. Specifically, their feedback sheds light on the particular challenges and priorities pertinent to the insurance industry's operations.

The company's venture into AI is relatively recent, having commenced at the beginning of this year, 2024. Despite being in the early stages of adoption, they have already begun incorporating AI tools into their daily operations, amidst a training period to further refine their usage. Additionally, the company has solidified a partnership with Microsoft as well.

7.1.3 Other Respondents

While profiling LinkedIn users is more generalized and detailed profiling is not feasible, the questionnaire was disseminated in groups where members have varying levels of interaction with artificial intelligence in their work. The questionnaire was shared multiple times into these groups on LinkedIn to maximize visibility, ultimately reaching various types of workers, predominantly those in the technology field. Their diverse perspectives and experiences contributed valuable insights to this research.

8. Analyzing the Results

This analysis aims to provide an overarching view of the ethical challenges in AI usage across all groups involved in this study based on the answers from the distributed questionnaire. Notably, there was no disagreement with the ethical challenges identified in the literature review, suggesting that the challenges discussed in the literature reflect the day-to-day experiences of AI workers.

1. Primary Ethical Concerns: Among the three groups surveyed, the most significant concern, indicated by 43% of respondents, is 'job loss, pay inequality, and/or financial instability.' This is followed by 'insecurity' at 19%, and 'lack of privacy and transparency' at 17%.

2. Additional Ethical Challenges: While the majority of respondents did not cite additional challenges beyond those in the literature review, some mentioned issues such as respect and

dignity in human-AI interactions, biased modeling, loss of control, immorality, and concerns about machines correcting human work.

3. Solution Perspectives: The proposed artifact (in Section 5.1) was separated into groups and made it shorter in the questionnaire so that is easier for the readers to assess the proposed artifact.

- First Solution Group: '(1) Implement AI algorithms that are capable of detecting and mitigating bias and discrimination in decision-making processes. (2) Conduct regular audits of AI systems to identify and correct potential discriminatory biases'. The results were that 60% of respondents rated this solution related to implementing AI algorithms to detect and mitigate bias favorably (rating 4 out of 5) and with 21% strongly agreeing (rating 5 out 5).

- Second Solution Group: '(1) Ensure full transparency in the use of employee data by obtaining explicit and informed consent for the collection and processing of personal data. (2) Adopt anonymization and pseudonymization practices to protect employee privacy during data processing'. The results were that 40% of respondents rated the solution regarding ensuring transparency in data usage favorably (rating 4 out of 5) and 32% strongly agreed (rating 5 out 5).

- Third Solution Group: '(1) Establish clear accountability mechanisms to ensure that AI systems are used ethically and fairly. (2) Create multidisciplinary ethics committees to evaluate and monitor the impact of AI systems on work practices and employees'. The results were that 38.3% of respondents strongly agreed (rating 5 out of 5) with establishing clear accountability mechanisms, while 38.3% also agreed (rating 4 out of 5).

- Fourth Solution Group: '(1) Implement robust security measures to protect AI systems against cyberattacks and malicious manipulations. (2) Conduct regular risk and vulnerability assessments on AI systems to identify and mitigate potential security threats'. The results were that 36% of respondents strongly agreed (rating 5 out of 5) with implementing robust security measures and 40% agreed (rating 4 out of 5).

- Fifth Solution Group: '(1) Develop change management strategies to minimize the impact of AI-driven automation on worker roles and skills. (2) Invest in recycling and retraining programs to enable employees to adapt to the new demands of the job market'. The results were that 47% of respondents agreed (rating 4 out of 5) with the solution involving change management strategies and 23% strongly agreed (rating 5 out of 5).

- Sixth Solution Group: '(1) Ensure that AI is not used for discrimination or unfair treatment. (2) Promote diversity and inclusion in AI systems'. The results were that 47% of respondents agreed (rating 4 out of 5) with ensuring AI is not used for discrimination and 26% strongly agreed (rating 5 out of 5).

4. Additional Insights Provided:

In the section where respondents could share additional details, various perspectives emerged:

1. Concerns about AI providing incorrect information without human supervision;

2. Importance of protecting personal biodata in the era of AI;

3. Philosophical reflections on the nature of intelligence and consciousness;

4. Recognition of bias stemming from fear-based parameters;

5. Ethical concerns about relying solely on algorithms for candidate selection in recruitment;

6. Challenges in designing interpretable AI systems, particularly in critical domains like healthcare and justice;

7. Questioning the limitations of AI in replacing human work;

8. Noting the opacity of many AI models, especially deep learning models, and the importance of explainability.

The quantitative analysis reveals significant consensus on the primary ethical concerns related to AI usage in the workplace (discovered in the literature review) and favorable attitudes toward the proposed solutions. However, the additional insights provided by respondents highlight the complexity and multifaceted nature of ethical challenges, such as respect and

dignity in AI interactions, fear-based modeling, bias, loss of control, immorality, AI correcting human work, task automation, and AI explainability.

This underscores the importance of ongoing dialogue and further research in this area. The engagement and depth of feedback were both surprising and enlightening, indicating that these ethical challenges are indeed relevant in the day-to-day experiences of AI workers and necessitating continuous examination and refinement of ethical frameworks to keep pace with technological advancements and real-world applications.

9. Artifact Refinement

As the artifact was refined in response to insights derived from the research findings, it became apparent that the ethical challenges associated with AI implementation are multifaceted and necessitate nuanced strategies for effective resolution.

The initial draft presented a comprehensive framework aimed at tackling key issues such as bias, privacy, accountability, and economic disruption. Nevertheless, the extensive insights obtained from the research questionnaire have served to further inform and enrich the approach to addressing these challenges. In this refined artifact, the perspectives and concerns articulated by AI professionals across various industries are integrated, ensuring that the solutions proposed are not only theoretically sound but also pragmatic and applicable in real-world settings.

By aligning the artifact with the experiences and priorities expressed by AI practitioners, the objective is to cultivate a culture of responsible AI development and deployment, emphasizing principles of fairness, transparency, and equity. The following guidelines is the artifact refined based on the insights collected from the questionnaire:

1. Bias and Discrimination:

- Implement algorithms capable of detecting and mitigating bias and discrimination, as highlighted by respondents' concerns about the potential for AI systems to perpetuate or exacerbate biases in decision-making processes.

- Conduct regular audits on AI systems to identify and correct discriminatory trends, leveraging insights from respondents who emphasized the importance of ongoing monitoring and evaluation to ensure fairness and equity.

2. Privacy and Transparency:

- Enhance transparency in data usage by providing clear and accessible information about data collection practices, responding to respondents' calls for greater transparency and accountability in AI-driven decision-making.

- Strengthen practices for obtaining explicit consent from employees for data processing activities, addressing concerns raised by respondents about the need for informed consent and privacy protections.

3. Responsibility and Justice:

- Strengthen accountability mechanisms to ensure ethical and fair use of AI systems, drawing from respondents' suggestions for clear roles and responsibilities in the governance of AI technologies.

- Empower multidisciplinary ethics committees to assess and monitor the impact of AI systems on work practices and employees, incorporating feedback from respondents who emphasized the importance of diverse perspectives in ethical decision-making.

4. Security:

- Enhance security measures to protect AI systems against cyberattacks and malicious manipulations, addressing concerns raised by respondents about the vulnerability of AI systems to security breaches.

- Conduct regular risk and vulnerability assessments on AI systems, incorporating insights from respondents who highlighted the need for proactive measures to identify and mitigate security threats.

5. Economic Disruption:

- Develop comprehensive ‘change management’ strategies to minimize the impact of AI-driven automation on job roles and skills, integrating feedback from respondents who emphasized the importance of reskilling and upskilling initiatives.

- Invest in targeted retraining programs to empower employees to adapt to new market demands, aligning with respondents' suggestions for supporting workforce transition in the AI-enabled economy.

6. Legal and Regulatory Concerns:

- Stay in accordance of evolving laws and regulations related to AI ethics and usage in the workplace, reflecting respondents' concerns about gaps in existing regulatory frameworks.

- Engage with regulatory bodies and government authorities to shape policies and guidelines promoting ethical AI usage, addressing respondents' calls for proactive engagement with policymakers.

7. Workers' Rights and Equity:

- Strengthen protections against discrimination and unfair treatment in AI systems, incorporating feedback from respondents who highlighted the importance of algorithmic fairness and diversity in AI development.

- Foster a culture of inclusion and equity in AI systems, reflecting respondents' suggestions for equitable representation and decision-making processes.

8. Technological Impact on Employment:

- Conduct comprehensive assessments of the potential impact of AI adoption on job opportunities and working conditions, integrating insights from respondents who emphasized the need for proactive measures to address workforce displacement.

- Implement targeted mitigation measures such as job support programs and income redistribution policies, aligning with respondents' suggestions for ensuring a just transition to the AI-enabled economy.

The refinement of their artifact means a pivotal advancement in the pursuit of ethical AI practices that are firmly rooted in theory and bolstered by empirical evidence. By assimilating the insights and recommendations furnished by AI practitioners via the research questionnaire, has fortified the strategy for tackling pivotal ethical dilemmas inherent in AI deployment.

The refined artifact exemplifies a dedication to upholding ethical standards in AI development and utilization, spanning endeavors such as increasing transparency, fostering accountability, ameliorating bias, and fortifying workforce adaptability. Moving forward, it must continue to engage with stakeholders across industry sectors, regulatory bodies, and academic institutions to ensure that ethical frameworks evolve in conjunction with technological advancements. By collaboratively shaping the future of AI ethics, it is possible to create a more inclusive, equitable, and sustainable digital landscape for all.

9.1 Significance of the Proposed Artifact in Real-World Applications

The refined artifact represents a significant leap forward in the practical application of ethical AI practices across various real-world scenarios. By integrating insights from AI workers and industry stakeholders, the artifact offers tangible solutions to address pressing ethical challenges in AI implementation. The following subsection will discuss how the refined artifact can be effectively put into practice.

9.1.1 Implementation Process

The artifact can be implemented in companies by following the below suggestions:

1. Comprehensive Training Programs:

- Implement comprehensive training programs for AI developers, data scientists, and other stakeholders involved in AI development and deployment. These programs should emphasize the ethical considerations outlined in the refined artifact and provide practical guidelines for integrating ethical principles into AI systems.

2. Adoption of Ethical Frameworks:

- Encourage organizations to adopt the ethical frameworks proposed in the refined artifact as part of their AI development and deployment processes. This involves integrating ethical considerations into project management workflows, decision-making frameworks, and risk assessment processes.

3. Stakeholder Engagement:

- Foster collaboration and dialogue among stakeholders, including AI practitioners, policymakers, regulatory bodies, and advocacy groups. Engage in ongoing discussions to refine and adapt ethical frameworks in response to emerging challenges and technological advancements.

The next subsection will discuss the necessary resources to put the refined artifact into practice.

9.1.2 Necessary Resources

The necessary resources to insert the refined artifact into practice in corporations include:

1. Educational Resources:

- Develop educational resources, such as online courses, workshops, and seminars, to educate AI practitioners and decision-makers about the ethical implications of AI technologies. These resources should provide practical guidance on implementing the recommendations outlined in the refined artifact.

2. Ethics Review Boards:

- Establish ethics review boards within organizations to oversee the ethical development and deployment of AI systems. These boards should consist of multidisciplinary experts who can evaluate the ethical implications of AI projects and provide guidance on mitigating risks.

3. Technical Tools:

- Invest in the development of technical tools and algorithms designed to detect and mitigate bias, enhance transparency, and ensure accountability in AI systems. These tools can help operationalize the recommendations outlined in the refined artifact and facilitate ethical decision-making in AI development.

The next subsection will discuss the potential impact, related to AI, that the refined artifact can have in organizations.

9.1.3 Potential Impact

1. Enhanced Trust and Accountability:

- By adhering to the ethical principles outlined in the refined artifact, organizations can enhance trust and accountability in their AI systems. This, in turn, can foster greater acceptance and adoption of AI technologies among users and stakeholders.

2. Reduced Bias and Discrimination:

- Implementing measures to detect and mitigate bias can help reduce the risk of discriminatory outcomes in AI-driven decision-making processes. This can lead to fairer and more equitable outcomes for individuals affected by AI systems.

3. Improved Social and Economic Outcomes:

- Ethical AI practices can contribute to improved social and economic outcomes by minimizing the negative impact of AI-driven automation on job displacement and inequality. By prioritizing workforce resilience and inclusivity, organizations can create a more sustainable and equitable future for workers and communities.

The refined artifact offers a roadmap for the ethical development and deployment of AI technologies in real-world applications. By embracing these principles and integrating them into their AI initiatives, organizations can uphold ethical standards, foster trust and accountability, and promote positive social and economic outcomes in the AI-driven era.

10. Conclusion

In conclusion, this study has thoroughly examined the ethical dilemmas posed by the integration of AI technologies in business operations, addressing issues such as bias, privacy, transparency, accountability, safety, and societal impact. Supported by significant literature, this thesis has provided a foundational framework for understanding and tackling these multifaceted challenges.

The proposed artifact offers practical solutions within a comprehensive ethical framework. Designed for application across diverse corporate contexts, including continuous improvement initiatives, employee training programs, and policy formulation, this artifact aims to enhance existing literature while demonstrating its pragmatic utility in real-world scenarios.

This framework, distinguished from existing guidelines such as the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and the European Commission's Ethics Guidelines for Trustworthy AI, synthesizes insights from multiple sources. By integrating diverse perspectives, it presents a holistic approach that navigates the complexities of ethical AI implementation in the workplace, effectively addressing gaps and limitations inherent in singular frameworks.

The iterative development process of this framework, which engages real employees for feedback and validation, ensures its relevance and effectiveness. Grounded in practical experiences and real-world concerns, the framework is a responsive and adaptable tool, capable of evolving alongside the dynamic landscape of workplace environments.

The urgency of addressing these ethical challenges cannot be overstated. While the proposed solutions are central and urgent, they are not exhaustive; many other concerns remain. Abandoning the technological revolution is not an option. Instead, embracing AI with a focus on ethical practices is crucial. Failing to act will only exacerbate the ethical issues.

Hiding the potential of AI technology is not an option; it must be embraced responsibly. Public policy plays a critical role in this endeavor. A broader vision encompassing public policy is necessary to ensure that AI technologies are developed and deployed ethically, benefiting society as a whole.

Despite the nascent nature of AI and the rapid pace of technological transformation, this study contributes to the broader discourse by offering a framework that is both adaptable and applicable across various sectors and scenarios. It fills critical gaps in the discourse surrounding AI ethics, providing tailored, actionable measures for each identified concern. By translating ethical principles into tangible practices, the framework empowers organizations to foster a culture of responsible AI usage.

In essence, this investigation contributes to society by offering a comprehensive, stakeholder-informed, and practical approach to addressing ethical challenges in AI implementation within the workplace. As organizations navigate the ethical complexities of AI adoption, this framework guides them toward responsible and ethical practices, ensuring a more inclusive, equitable, and sustainable future for all.

References

*References in bold are from SCOPUS and WOS (main platforms used in the literature review)

Adamson, G., Havens, J. C., and Chatila, R. (2019). Designing a Value-Driven Future for Ethical Autonomous and Intelligent Systems, in Proceedings of the IEEE, vol. 107, no. 3, pp. 518-525, March 2019, doi: 10.1109/JPROC.2018.2884923.

Atwell, E. (2023). Kolide. https://www.kolide.com/blog/89-of-workers-use-ai-far-fewer-understand-the-risks#download-kolides-shadow-it-report.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H. S., Roff, H. M., Allen, G. C., Steinhardt, J., Flynn, C., Héigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., . . . Amodei, D. (2018). The Malicious Use of Artificial intelligence: *Forecasting, Prevention, and Mitigation*. arXiv (Cornell University). https://doi.org/10.17863/cam.22520.

Chatila, R., & Havens, J. C. (2019). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. In International series on intelligent systems, control and automation: science and engineering (pp. 11–16). https://doi.org/10.1007/978-3-030-12524-0_2.

Ebers, M. (2021). Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act. The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics.https://ssrn.com/abstract=3900378.

Economou, N. (2019). Principles for the Trustworthy Adoption of AI in Legal Systems: The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. https://ceur-ws.org/Vol-2484/keynote1.pdf.

Fisher E., Flynn M.A., Pratap P., Vietas J.A. (2023). Occupational Safety and Health Equity Impacts of Artificial Intelligence: A Scoping Review. Int J Environ Res Public Health. 24;20(13):6221. doi: 10.3390/ijerph20136221.

Hevner, Alan & R, Alan & March, Salvatore & T, Salvatore & Park, & Park, Jinsoo & Ram, & Sudha,. (2004). Design Science in Information Systems Research. Management Information Systems Quarterly. 28. 75-.

Hickman, E., Petrin, M. (2021). Trustworthy AI and Corporate Governance: *The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective*. Eur Bus Org Law Rev 22, 593–625. doi: 10.1007/s40804-021-00224-0

Shahriari, K. and Shahriari, M. (2017). IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems," 2017 IEEE Canada International Humanitarian Technology Conference (IHTC), Toronto, ON, Canada, 2017, pp. 197-201. doi: 10.1109/IHTC.2017.8058187.

Kriebitz, A., Lutge, C. (2020). Artificial Intelligence and Human Rights: A Business Ethical Assessment. *Business and Human Rights Journal*, 5(1), 84-104. doi:10.1017/bhj.2019.28.

Krkac, K. (2019). Corporate social irresponsibility: humans vs artificial intelligence. *Social Responsibility Journal*, Vol. 15 No. 6, pp. doi: 10.1108/SRJ-09-2018-0219.

Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. Asian Journal of Law and Society, 7(3), 437–451. doi:10.1017/als.2020.19

Lysova, E. I., Tosti-Kharas, J., Michaelson, C., Fletcher, L., Bailey, C., & McGhee, P. (2023). Ethics and the Future of Meaningful Work: Introduction to the special

issue. Journal of Business Ethics, 185(4), 713–723. https://doi.org/10.1007/s10551-023-05345-9

Machi, L. A., & McEvoy, B. T. (2016). The Literature Review: Six Steps to Success. Corwin Press. *Research Gate.* https://www.researchgate.net/publication/339032640_Book_Review_The_Literature_Review_Six_Steps_to_Success_3rd_edition_by_Lawrence_A_Machi_and_Brenda_T_McEvoy_2016

Müller, V. C. (2020). Ethics of Artificial Intelligence and Robotics. *ResearchGate*. https://www.researchgate.net/publication/340435326_Ethics_of_Artificial_Intelligence_and_Robotics

Nyenno, I. (2023). Managerial future of artificial intelligence. *Virtual Economics*. doi: 10.34021/ve.2023.06.02(5).

Roemmich, K., Rosenberg, T. I., Fan, S., & Andalibi, N. (2023). Values in Emotion Artificial Intelligence Hiring Services: Technosolutions to Organizational problems. *Proceedings of the ACM on Human-computer Interaction*, 7(CSCW1), 1–28. doi: 10.1145/3579543.

Smuha, N. (2019). The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. Computer Law Review International, 20(4), 97-106. https://doi.org/10.9785/cri-2019-200402.

Varsha, P. (2023). How can we manage biases in artificial intelligence systems – A systematic literature review. International Journal of Information Management Data Insights, 3(1), 100165. https://doi.org/10.1016/j.jjimei.2023.100165.

Wirtz, J., Kunz, W. H., Hartley, N., & Tarbit, J. (2022). Corporate digital responsibility in service firms and their ecosystems. Journal of Service Research, 26(2), 173–190. https://doi.org/10.1177/10946705221130467.

Zhao, J., & Fariñas, B. G. (2022). Artificial intelligence and sustainable decisions. European Business Organization Law Review, 24(1), 1–39. https://doi.org/10.1007/s40804-022-00262-2.

Zirar, A., Ali, S. I., & Islam, N. (2023). Worker and workplace Artificial Intelligence (AI) coexistence: Emerging themes and research agenda. Technovation, 124, 102747. https://doi.org/10.1016/j.technovation.2023.102747