

Responsible Artificial Intelligence: Designing Frameworks for Ethical, Sustainable, and Risk-Aware Practices

Intelligence Artificielle Responsable : Concevoir des Cadres pour des Pratiques Éthiques, Durables et Sensibles aux Risques

Thèse de doctorat de l'Université Paris-Saclay
École doctorale n° 630, Droit Economie Management (DEM)
Spécialité de doctorat : Sciences de gestion
Réfèrent : Université d'Évry Val d'Essonne

Thèse préparée dans les unités de recherche : du LITEM (Université Paris-Saclay, Univ Evry, IMT-BS), sous la direction de Christine BALAGUE, HDR

Ahmad HAIDAR

Composition du Jury

Membres du jury avec voix délibérative

Manuel ZACKLAD

Professeur au CNAM en sciences de l'Information et de la Communication, HDR

Président du Jury

Julien CLOAREC

Professeur des universités en Sciences de Gestion, HDR, Université Jean Moulin Lyon 3

Rapporteur & Examineur

Margherita PAGANI

Professeure titulaire en Intelligence Artificielle et Numérique en Marketing, HDR, Directeur du Centre pour l'IA de Skema, Skema Business School

Rapporteur & Examinatrice

Liliana MITKOVA

Professeure des Universités en Sciences de Gestion, HDR, Université Evry Val d'Essonne.

Examinatrice

Hassan AYOUB

Professeur des Universités en économie, Université Libanaise, Beyrouth

Examineur

Titre : Intelligence Artificielle Responsable : Concevoir des Cadres pour des Pratiques Éthiques, Durables et Sensibles aux Risques

Mots clés : Intelligence Artificielle Responsable ; Innovation Numérique Responsable ; Risques de l'IA Générative ; Cadre de Gouvernance de l'IA ; Principes Éthiques de l'IA

Résumé : L'intelligence artificielle (IA) transforme rapidement le monde, redéfinissant les relations entre technologie et société. Cette thèse explore le besoin essentiel de développer, de gouverner et d'utiliser l'IA et l'IA générative (IAG) de manière responsable et durable. Elle traite des risques éthiques, des lacunes réglementaires et des défis associés aux systèmes d'IA, tout en proposant des cadres concrets pour promouvoir une Intelligence Artificielle Responsable (IAR) et une Innovation Numérique Responsable (INR). La thèse commence par une analyse approfondie de 27 déclarations éthiques mondiales sur l'IA pour identifier des principes dominants tels que la transparence, l'équité, la responsabilité et la durabilité. Bien que significatifs, ces principes manquent souvent d'outils pratiques pour leur mise en œuvre. Pour combler cette lacune, la deuxième étude de la recherche présente un cadre intégrateur pour l'IAR basé sur quatre dimensions : technique, IA pour la durabilité, juridique et gestion responsable de l'innovation. La troisième partie de la thèse porte sur l'INR à travers une étude qualitative basée sur 18 entretiens avec des gestionnaires de secteurs divers.

Cinq dimensions clés sont identifiées : stratégie, défis spécifiques au numérique, indicateurs de performance organisationnels, impact sur les utilisateurs finaux et catalyseurs. Ces dimensions permettent aux entreprises d'adopter des pratiques d'innovation durable et responsable tout en surmontant les obstacles à leur mise en œuvre.

La quatrième étude analyse les risques émergents liés à l'IAG, tels que la désinformation, les biais, les atteintes à la vie privée, les préoccupations environnementales et la suppression d'emplois. À partir d'un ensemble de 858 incidents, cette recherche utilise une régression logistique binaire pour examiner l'impact sociétal de ces risques. Les résultats soulignent l'urgence d'établir des cadres réglementaires renforcés, une responsabilité numérique des entreprises et une gouvernance éthique de l'IA.

En conclusion, cette thèse apporte des contributions critiques aux domaines de l'INR et de l'IAR en évaluant les principes éthiques, en proposant des cadres intégratifs et en identifiant des risques émergents. Elle souligne l'importance d'aligner la gouvernance de l'IA sur les normes internationales afin de garantir que les technologies d'IA servent l'humanité de manière durable et équitable.

Title : Responsible Artificial Intelligence: Designing Frameworks for Ethical, Sustainable, and Risk-Aware Practices

Keywords : Responsible Artificial Intelligence ; Responsible Digital Innovation ; Generative AI Risks ; AI Governance Framework ; Ethical AI Principles

Abstract : Artificial Intelligence (AI) is rapidly transforming the world, redefining the relationship between technology and society. This thesis investigates the critical need for responsible and sustainable development, governance, and usage of AI and Generative AI (GAI). The study addresses the ethical risks, regulatory gaps, and challenges associated with AI systems while proposing actionable frameworks for fostering Responsible Artificial Intelligence (RAI) and Responsible Digital Innovation (RDI). The thesis begins with a comprehensive review of 27 global AI ethical declarations to identify dominant principles such as transparency, fairness, accountability, and sustainability. Despite their significance, these principles often lack the necessary tools for practical implementation. To address this gap, the second study in the research presents an integrative framework for RAI based on four dimensions: technical, AI for sustainability, legal, and responsible innovation management.

The third part of the thesis focuses on RDI through a qualitative study of 18 interviews with managers from diverse sectors. Five key dimensions are identified: strategy, digital-specific challenges, organizational KPIs, end-user impact, and catalysts. These dimensions enable companies to adopt sustainable and responsible innovation practices while overcoming obstacles in implementation.

The fourth study analyzes emerging risks from GAI, such as misinformation, disinformation, bias, privacy breaches, environmental concerns, and job displacement. Using a dataset of 858 incidents, this research employs binary logistic regression to examine the societal impact of these risks. The results highlight the urgent need for stronger regulatory frameworks, corporate digital responsibility, and ethical AI governance. Thus, this thesis provides critical contributions to the fields of RDI and RAI by evaluating ethical principles, proposing integrative frameworks, and identifying emerging risks. It emphasizes the importance of aligning AI governance with international standards to ensure that AI technologies serve humanity sustainably and equitably.

Acknowledgments

Completing this PhD thesis has been a profoundly challenging and equally rewarding experience - an endeavor that I could not have undertaken without the support and encouragement of everyone who guided me along this journey.

First and foremost, I extend my deepest gratitude to my professor, Christine BALAGUE, whose guidance and mentorship were instrumental in shaping both this research and my professional development. Her expertise and patient support have been invaluable throughout my Ph.D. journey.

Following closely, I must express my heartfelt thanks to my family for their boundless love and support. Their encouragement has been a cornerstone of my journey, and I am eternally grateful for their belief in my aspirations. Throughout my Ph.D., I encountered several significant challenges that tested my resolve and capabilities. With the steadfast support and wise counsel of both my family and my professor, I was able to overcome them.

I could not have undertaken this journey without my defense committee's invaluable guidance and expertise. Their thorough critiques and encouraging feedback were crucial in refining my research. Additionally, my heartfelt thanks go to my lab colleagues at Paris-Saclay (LITEM) and the distinguished professors I had the privilege to work with. Their insightful training courses enhanced my skills and broadened my academic perspective. I sincerely appreciate Sylvie Prehu's unwavering support and dedication throughout my PhD journey. Her assistance was invaluable to me and to all LITEM PhD students. I must also extend my gratitude to the various firms, the research hub, and all study participants who contributed to completing the studies presented in this thesis. Their willingness to engage and share their experiences was vital in enriching my research and ensuring its relevance and impact. Their collaboration has been a cornerstone of this scholarly work.

The “Good in Tech” research network has graciously funded this thesis. Good in Tech is a partnership between Institut Mines-Télécom and Sciences Po under the aegis of the Risk Foundation. The network stands at the forefront of interdisciplinary research, committed to exploring the intersections of (1) data, algorithms, and society and (2) corporate digital responsibility. Good in Tech’s vision involves reimagining innovation and technology as catalysts for a better world, driven by and for humans, with its primary objectives centered on generating and disseminating knowledge in these domains to academia, educators, companies, decision-makers, regulators, and the general public. My research has greatly benefited from the chair’s vision and commitment to a multidisciplinary approach to understanding and leveraging technology for societal good.

Table of Content

Acknowledgments.....	2
Table of Content	4
List of Figures	8
List of Tables.....	9
Chapter 1: Introduction.....	10
1.1 Digital innovation: Foundation and evolution	10
1.1.1 The concept of digital innovation.....	10
1.1.2 Shaping digital innovation via digital-related capabilities	11
1.1.3 Historical evolution of digital innovation	13
1.2 Introduction to artificial intelligence	18
1.2.1 Definition and types of AI.....	18
1.2.2 History and development of AI	23
1.2.3 The current state of AI and its potential applications.....	26
<i>Healthcare: Personalized medicine</i>	27
<i>Finance: Fraud detection and risk management</i>	28
<i>Retail: Customer experience and inventory management</i>	29
<i>Manufacturing: Predictive maintenance and quality control</i>	30
<i>AI in the information environment</i>	31
<i>AI in legal services</i>	32
<i>AI in education</i>	32
1.2.4 Types of AI and their differences.....	33
1.2.5 Overview of generative AI	34
1.2.5.1 Generative AI definition and application	34
1.2.5.2 Understanding generative AI: Mechanisms and functionality.....	35
1.2.5.3 Relationship between LLM, GAI, DL, ML, and AI	37
1.2.6 Key Machine Learning Techniques within AI.....	38
1.3 Risks Associated with AI	41
1.3.1 Definitions of “Risks”	41
1.3.2 Risks associated with AI	42
1.3.2.1 Privacy violation	45
1.3.2.2. Autonomous systems and weaponization	45
1.3.2.3. Cyberattacks	46
1.3.2.4. Manipulation and Social Engineering	47
1.3.2.5. Unemployment and Economic Disparity.....	48
1.3.3 Advocating solutions: addressing the critical challenges of AI deployment	48

1.4 Research questions	52
1.5 Epistemological positioning	55
1.6 Research methodologies	56
1.7 Structure of the dissertation	58
Chapter 2: A Framework to Implement Responsible Digital Innovation: Dimensions, Best Practices and Challenges	62
About this chapter	63
2.1 Introduction	63
2.1.1 Definition	63
2.1.1.1 The concept of responsibility	63
2.1.1.2 Innovation	64
2.1.1.3 Overview of additional concepts linked to responsible digital innovation	66
2.1.2 The discourse on responsible innovation and responsible research innovation	67
2.1.2.1 Responsible innovation	67
2.1.2.2 Responsible research and innovation	68
2.1.2.3 Overview of distinctions between RI and RRI frameworks	71
2.1.3 Study's objectives and research question	72
2.2 Dark side of digital innovation	72
2.2.1 Environmental perspective	72
2.2.2 Economy perspective	77
2.2.3 Societal perspective	79
2.3 Methodology	82
2.3.1 Research design	82
2.3.2 Data collection process	82
2.3.3 Data analysis	83
2.3.4 Ethical considerations	84
2.4 Results	84
2.4.1 Definitions of responsible digital innovation	84
2.4.2 Responsible digital innovation dimensions and practices	86
<i>Responsible Digital Innovation Strategy</i>	88
<i>Responsible for the Design of Digital Innovation</i>	90
<i>Organization and KPI</i>	93
<i>Digital responsibility towards end users</i>	96
<i>Catalysts</i>	98
2.4.3 Barriers against responsible digital innovation	100
2.5 Discussion	101

2.6 Managerial implication	102
2.7 Policy implications for enhancing RDI practices.....	103
2.8 Conclusion.....	104
Chapter 3: Ethics in AI: Analyzing Declarations and Future Directions	105
About this chapter	106
3.1 Clarifying ethical AI.....	106
3.1.1 What is ethics	106
3.1.2 Ethical AI: definition and moral philosophies	108
3.1.3 Trustworthy AI.....	110
3.2 Methodology.....	111
3.3 Findings	114
3.3.1 Analysis of ethical principles in AI declarations.....	114
3.3.1.1 An updated round-up of ethical AI declarations	114
3.3.1.2 Frequency analysis across overall declarations	114
3.3.2 Integrative literature review findings	117
3.3.2.1 Strengths and opportunities presented by ethical AI declaration.....	117
3.3.2.2 Weaknesses and limitations of AI ethics declaration	126
3.4 Discussion	141
3.5 Policy implication	142
3.6 Managerial implication	143
3.7 Limits and future research	145
3.8 Conclusion.....	145
Chapter 4: An Integrative Theoretical Framework for Responsible Artificial Intelligence	147
About this chapter	148
4.1 Theoretical background.....	148
4.1.1 The need for AI governance within the company.....	148
4.1.2 Frameworks for responsible AI	149
4.1.3 Exploring the limitations of the principles-approach for AI.....	151
4.2 Proposed integrative theoretical framework	153
4.2.1 Technical dimension.....	154
4.2.2 Responsible innovation dimension.....	157
4.2.3 Legal dimension.....	162
4.2.4 AI and sustainability	165
4.2.4.1 Definition of sustainable AI	165
4.2.4.2 Opportunities and challenges for AI for sustainability.....	165
4.3 Discussion.....	186

4.4 Conclusion.....	190
Chapter 5: Generative AI: A Quantitative Study on Emerging Risks and Impacts	192
About this chapter	193
5.1 Theoretical framework of generative AI risks.....	193
5.1.2 Theoretical foundation of this study	193
5.1.2.1 Risk object.....	194
5.1.2.2 Object at risk	201
5.2 Methodology.....	220
5.2.1 Data collection.....	220
5.2.2 Period of study and incident selection	220
5.2.3 Computational framework	220
5.2.3.1 Programming language and libraries	220
5.2.3.2 Integration and utilization of AI models	221
5.2.3.3 Validation of AI outputs	221
5.2.3.4 Statistical analysis framework.....	221
5.3 Results.....	222
5.3.1 Incidents distribution	222
5.3.2 Descriptive Analysis	223
5.3.3 Binary logistic regression.....	225
5.3.3.1 Individual level	225
5.3.3.2 Organization Level	227
5.3.3.3 Societal level	229
5.3.3.4 Environmental level.....	230
5.3.3.5 Model fit and summary statistics	231
5.4 Discussion	232
5.4.1 Theoretical contributions.....	232
5.4.2 Managerial and policy implications.....	235
Chapter 6: Conclusion	237
6.1 Summary of the key findings for each chapter	238
6.2 Contribution of the thesis	244
6.3 Managerial implications	248
6.4 Policy-making implications.....	250
6.5 Limitation of the thesis.....	252
6.6 Recommendation for future research.....	253
References	256
Annex 1	327

List of Figures

Figure 1 : Technological evolution (source: Yigitcanlar, 2011)	14
Figure 2 : The term of technological evolution (source: Almgren and Skobelev, 2020).....	18
Figure 3 : Some definitions of artificial intelligence, organized into four categories (source: Russel & Norvig, 2010, p.2)	20
Figure 4 : Historical overview of AI (source: Francesconi, 2020).....	25
Figure 5 : AI global market forecasting (source: Marketandmarkets (2023))	26
Figure 6 : AI in the healthcare market (source: Acumen research and consulting, 2023)	28
Figure 7 : GAI in banking and finance (source: https://www.precedenceresearch.com/generative-ai-in-banking-and-finance-market)	29
Figure 8 : AI in the manufacturing market (source: Vantage market research (2022))	31
Figure 9 GAI in healthcare prediction (source: Mota, 2024)	35
Figure 10 Relationship between LLM and other models (source : Hadid et al., 2024)	38
Figure 11 ML techniques (source: datasciencedojo).....	39
Figure 12 AI Act's hierarchy of criticality and the regulatory framework based on risk for employing algorithmic systems; "SS" denotes a later article. (source: Díaz-Rodríguez et al. (2023) inspired by Wahlster & Winterhalter (2020) and Campos & Laurent (2023)).....	51
Figure 13: Overview of the research questions of the thesis.....	55
Figure 14 : trust issues and climate scientists (source: WEF, 2021).....	73
Figure 15 : Global Annual Internet Traffic (source: iea.org)	74
Figure 16 : Number of active mobile broadband subscriptions worldwide from 2007 to 2021 (source: statista.com).....	74
Figure 17 : Energy demand in data centers worldwide from 2015 to 2021 (source: IEA 2022 Ministerial Meeting	75
Figure 18 : RDI dimensions (source: The authors)	86
Figure 19 : Frequency analysis of the principles across all declarations (source: author)....	116
Figure 20 : Societal Principles & Dimensions (source: author's elaboration)	126
Figure 21 : Responsibility distributed over individuals or ensuring the pro-ethical design of AI products (source: Morley et al., 2023)	131
Figure 22 : Theoretical framework for responsible AI (source: author)	154
Figure 23 : Volume of data created, captured, copied, and consumed worldwide from 2010 to 2020 with forecasts for 2021 to 2025 (source: Statista).....	169

Figure 24 : Theoretical Framework of the study	194
Figure 25: Incidents distribution	223

List of Tables

Tableau 1 : Definitions for AI with references (Source: author's elaboration)	22
Tableau 2 : Definition of risk (source: authors)	42
Tableau 3 : Summary of the dissertation.....	61
Tableau 4 : Summary of Key Terms in Responsible Innovation (source: Author).....	67
Tableau 5 : Comparative Analysis of Responsible Innovation (RI) and Responsible Research and Innovation (RRI)	71
Tableau 6 : Sample of the study	83
Tableau 7 : Definitions of RDI based on participants	86
Tableau 8 : Responsible digital innovation dimensions, sub-dimensions, and practices (source: the authors).....	88
Tableau 9 : Obstacles of RDI based on participants	101
Tableau 10 : Types of ethics (source: Adeusi, 2020).....	107
Tableau 11 : Opportunities and strengths of AI declarations (source: author)	117
Tableau 12 : Ethical concerns related to algorithmic use,'(Source: Morley et al. (2020), adapted from the ‘map’ created by Mittelstadt et al., 2016).	122
Tableau 13 : Weaknesses of AI declarations (source: author)	127
Tableau 14 : Key Technical Issues Faced in AI Implementation (source: author)	156
Tableau 15 : Responsible innovation dimensions	161
Tableau 16 : EU and US AI legislation.....	164
Tableau 17 : Relationship between AI and environment with corresponding SDGs (Source: author based on literature).....	171
Tableau 18 : Overview of positive and negative impacts of AI on society and their corresponding SDGs (source: Ahmad based on Literature).....	178
Tableau 19 : Overview of the relationship between AI and the economy (source: author based on literature)	185
Tableau 20 : Risk object dimensions (source: authors).....	196
Tableau 21 : Object at risk dimension (source: authors).....	203
Tableau 22 : Patters of AI in military	215
Tableau 23 : Descriptive analysis.....	224
Tableau 24 : Results of binary binary logistic regression model at the “individual level”	227
Tableau 25 : Results of binary logistic regression model at the “organizational level	229
Tableau 26 : Results of binary logistic regression model at the “societal level”	230
Tableau 27 : Model fit and summary statistics	232

Chapter 1: Introduction

1.1 Digital innovation: Foundation and evolution

1.1.1 The concept of digital innovation

Digital Innovation (DI) refers to “... the creation of (and the consequent change in) market offerings, business processes, or models that result from the use of digital technology” (Nambisan et al., 2017, p. 224). Digitalization (e.g., of a process module step) is the first step in identifying the chronological sequence, followed by digital transformation. This transformation process leads ultimately to fundamental changes in a business model, such as replacing activities humans performed previously with IT-based activities (Herberger & Dötsch, 2021; Gregory et al., 2019). Adopting technological solutions to develop new processes and products, habits, and good practices increases companies’ capacity to innovate and thus allows them to meet the needs of a continually changing market (Gil-Gomez et al., 2020). Therefore, building a digital knowledge base may be necessary to enrich firms’ established knowledge. Given that knowledge is a critical resource for all companies (Friedrich et al., 2020), it is helpful to understand the way a Knowledge Management System that incorporates DI can accelerate the process of creating value in the long term and guide the corporate strategy to develop new, innovative business models (Di Vaio et al., 2021). Hanelt et al. (2020) estimated DI’s effect on operating performance (i.e., return on assets), market performance (i.e., market-to-book ratio), and analyst forecasts. The results suggested that DI significantly increases firms’ operating and capital market performance. The shift from one socio-technical system to another requires innovation and systems approaches to innovation in analysis accompanied by “responsibility” and “sustainability.” Digital technologies challenge existing innovation policies by transforming the nature of the innovation process; for example, any piece of information and knowledge is digital (Nepelski, 2019).

Recently, Ciriello et al. (2018) identified characteristics of digital technology and indicated that DI entails innovating products, processes, or business models using digital technology platforms as a means or end within and across organizations. For example, Uber, the world’s largest taxi company, owns no taxis; instead, it provides location-based services using cloud-based technologies. Consequently, this has changed people’s traditional car-hailing approach to using taxis, which now relies on a digital platform to provide new service processes to create meaningful customer value (Li & Hou, 2019). Gao and Hands (2021) explored the

practice and significance of “design” in “digital innovation.” The application of design necessitates giving more attention to its effectiveness in helping enterprises achieve strategic and systematic growth. “Design” is an essential tool and method to gain insight into the possibilities and facilitate the development of ideas, thereby addressing complex challenges and macro uncertainties (Gao and Hands. 2021). Recently, DI’s influence on businesses has drawn the attention of various scholars. One study indicated that DI significantly increases a firm’s operating and capital market performance (Hanelt et al., 2020). Another study found that the average number of digital technologies adopted increases significantly as product innovation shifts from non-innovation to incremental and radical innovation (Blichfeldt & Faullant, 2021).

In a world with increasing poverty, inequality, environmental degradation, and injustice, DI offers a ray of hope to solve societal problems. Digital social innovation uses digital technology, community participation and cooperation, co-creation methodologies, and bottom-up approaches to address socially critical issues like healthcare, corruption, and immigration (Bonina et al., 2021).

Hinings et al. (2018) categorize DI by the creation of new products, services, and processes; among others, digital transformation combines the effects of various DIs and brings new agents, structures, practices, values, and beliefs that change, threaten, replace, or complement existing rules within organizations and industries (Hinings et al., 2018). According to Yoo et al. (2012), DI has the characteristics of *convergence*, *generativity*, *widespread participation*, and *democratization*. Digital convergence refers to bringing previously separate technologies together. For example, the iPhone combined such features as a music player, video camera, GPS unit, and web browser with a traditional phone, among other things.

1.1.2 Shaping digital innovation via digital-related capabilities

Innovation refers to developing and implementing a new practice or object (Rogers, 1983). Different resources, including tangible ones like technology and intangible ones like knowledge, accomplish innovation (Davila et al., 2012). This allows for the efficient implementation of innovative routines within companies. Digital transformation is not only a determinant of innovation but also enables companies to understand how the skills are needed and adjust those skills to achieve innovative results, processes, and related markets (Kohli and Melvill, 2019; Nambisan et al., 2017). Research has shown that companies use their dynamic capacity to adapt, integrate, or restructure existing resources and capabilities to adapt to environmental change (Teece et al., 1997). The dynamic capacity theory addresses the unique

and complex capabilities that make it possible to adapt quickly to such changes (Teece, 2014; Teece et al., 1997). Therefore, this theory offers an ideal way to examine digital transformation-related changes. The “related routines within companies for specific tasks” are companies’ dynamic capabilities (Ngo and O’Cass, 2013, p. 1135). Research on routines and operations companies use to develop DI is necessary to classify existing capabilities. Accordingly, the following sections examine the characteristics and associated capabilities of DI presented in prior research.

Human capabilities. Digitalization requires employees’ support, preparedness, and digital expertise (Chan et al., 2021; Kane et al., 2015). Companies that do not have digitally expert employees or personnel qualified to solve problems suffer from poor capability attributable to the complexity of today’s digital era (Kache and Seuring, 2017).

Collaborative abilities. Due to digitalization’s complexity, individual agents can’t achieve a competitive advantage (Canhoto et al., 2016; Kohli and Mekville, 2019). Therefore, companies must develop the ability to collaborate to achieve viable alliances and network to build value with suitable partners (Amit and Han, 2017; El Sawy et al., 2016; Sjödin et al., 2016). As the structure of social relationships has changed internally and externally (Pagani and Pardo, 2017), small and medium-sized enterprises must look for opportunities to collaborate so their present abilities complement their partners (Chan et al., 2019). Emphasizing a human-centric, design-thinking approach can significantly foster ethical and responsible technological innovation, particularly in the realm of emerging and converging technologies (van den Hoven, 2017)

Technical capabilities. In today’s digital era, engaging in internal and external organizational cooperation requires technical abilities as a complementary channel to ensure that advanced market offerings are provided (Legner et al., 2017; Sjödin et al., 2016). Thus, these capabilities play a critical role in using digitalization to integrate products and services (El Sawy et al., 2016) and access updated and borderless services and activities (Parida et al., 2015; Sjödin et al., 2016; Yoo et al., 2010). Hence, in this study, digital-related technical capabilities refer to the bundle of capabilities that facilitate the implementation of digitalization in companies.

Innovation capabilities. Disruptive digital business, which emerges from digitalization, requires innovation capabilities to be managed successfully. The capacity to innovate, identify, exploit business opportunities, and diversify the business area are among the innovation skills needed for disruptive digital business (Sousa and Rocha 2019). In this view Nylén and Holmström (2015) elaborated a framework for digital innovation for strategic planning covering five key areas: user experience, value proposition, digital evolution scanning, skills, and improvisation.

In summary, organizations often need development capabilities that correspond closely to new situations and dynamic changes when they respond to the exploitation of DIs. Moreover, organizations need to mitigate inflexibility and be better equipped for transformations. Thus, human, collaborative, technical, and innovation capabilities are all advised for businesses to increase their survival in today's competitive environment.

1.1.3 Historical evolution of digital innovation

The evolution of technology is a central theme for management theory due to the transformative effect of technological change on societies, markets, industries, organizations, and individuals (Grodal et al., 2023). However, Arthur (2009, p. 15ff) claims that one of the most important problems to understand regarding technology is to explain how it evolves. Arthur and Polak (2006, p. 23) claim that: “technology ... evolves by constructing new devices and methods from ones that previously exist, and in turn offering these as possible components—building blocks—for the construction of further new devices and elements”. Pistorius and Utterback (1997) argue that emerging technologies often substitute for more mature technologies. Fisher and Pry (1971, p. 88) state that: “the speed with which a substitution takes place is not a simple measure of the pace of technical advance. . . . it is, rather a measure of the unbalance in these factors between the competitive elements of the substitution”.

The evolution of digital technology is a compelling narrative of innovation that has fundamentally reshaped the global socio-economic landscape. Groundbreaking advancements define this transformation, tracing it through a series of technological waves that build progressively upon the achievements of preceding eras. These waves reflect the natural progression of technological capabilities and the shifting economic and industrial development paradigms. Yigitcanlar (2021) illustrates the descriptions of these waves in Figure 1.

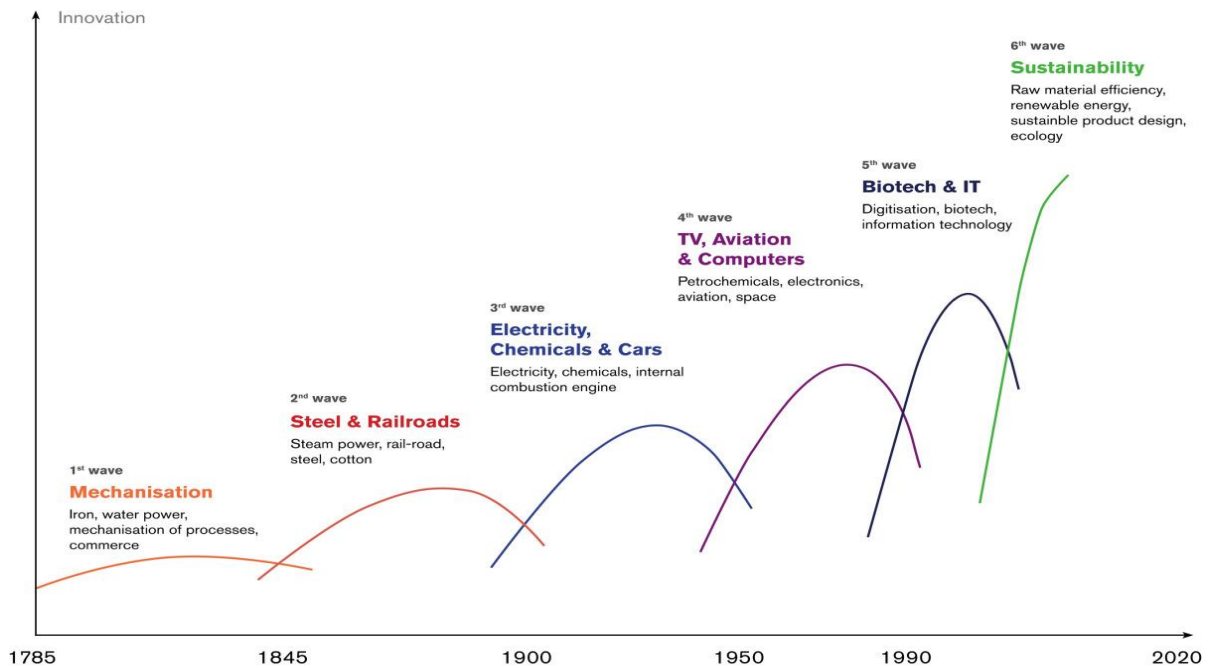


Figure 1 : Technological evolution (source: Yigitcanlar, 2011)

First wave: Mechanization (1785–1845)

The advent of mechanization signaled the first wave of technological innovation. During this era, there was a shift from manual work to mechanical processes, predominantly driven by water and steam power. The industrial revolution began with this wave, introducing iron and textile industries that leveraged mechanized manufacturing, setting the stage for mass production.

Second wave: Steel & railroads (1845–1900)

The second wave characterized the expansion of steel manufacturing and the proliferation of railroads. This era facilitated transformative growth in transport and infrastructure, enabling larger-scale industrial outputs and a rapid transportation network that connected distant markets.

Third wave: Electricity, chemicals & cars (1900–1950)

The third wave introduced electricity and the internal combustion engine, heralding new automotive and chemical manufacturing industries. Significant advancements in mass production techniques mark this period, notably Ford's assembly line, which revolutionized automobile production and urban mobility.

Fourth wave: TV, aviation & computers (1950–1990)

The fourth wave saw the rise of television, aviation, early computing, and advances in aviation and space technology. The history of digital innovation began with the creation of early computing equipment in the twentieth century. The 1940s witnessed the development of the first electronic computers, such as the ENIAC (Electronic Numerical Integrator and Computer), which were primarily used for computing tasks like artillery firing tables and hydrogen bomb research (Lean, 2013). These early computers established the foundation for digital processing and data management. In 1943, Warren McCulloch and Walter Pitts laid the groundwork for artificial neurons, a precursor to neural networks, which would later become fundamental to AI (Kay, 2001). The 1950s saw further advancements with Alan Turing's exploration of machine intelligence and John McCarthy's coining of the term "artificial intelligence" in 1956 at the Dartmouth Conference. Personal computing emerged during the 1970s and 1980s due to the advancement of microprocessors and the creation of personal computers (PCs), such as the Apple II and IBM PC (Hagedoorn et al., 2001). These devices facilitated widespread access to computing power, enabling consumers and enterprises to utilize software for a multitude of reasons. The advent of the internet in the late twentieth century revolutionized data networking, facilitating unprecedented global communication and information sharing (Collen & Kulikowski, 2015). According to Day et al. (2005), around 15% of homes in the United States owned a personal computer in 1989.

Fifth wave: Biotech & IT (1990–2020)

Major advancements in digital technologies, biotechnology, microelectronics, biotechnology, space exploration, new types of energy sources and energy generation, genetic engineering, materials, satellite communication, and a transition towards an information and service-based economy characterize the fifth wave.

The 1990s were a period of significant change, characterized by the commercialization of the internet and the emergence of the World Wide Web. Tim Berners-Lee's original idea for a global system of interlinked hypertext documents, which led to the creation of the World Wide Web in 1989 (Sugumaran & Gulla, 2012), along with subsequent advancements in web browsers such as Mosaic in 1993 and Netscape Navigator in 1994, greatly facilitated the widespread availability of internet connections (Anand & Saxena, 2013). This innovation transformed how people communicate, conduct business, and exchange information, establishing the foundation for the digital economy.

With the introduction of smartphones in the late 1990s and early 2000s, mobile computing rose to prominence (Goggin, 2013). The 2000s witnessed the mobile revolution, transforming digital innovation with the rise of smartphones. Apple's introduction of the iPhone in 2007 was a watershed moment, integrating computing power with mobile telephony. This era also saw the proliferation of mobile operating systems like Android, which further democratized access to smart devices. Smartphones revolutionized how people interact with technology, providing portable access to the internet, applications, and multimedia. This period also saw the rise of social media platforms, enhancing connectivity and information dissemination on a global scale.

The 2010s decade saw notable progress in cloud computing and big data. Firms like Amazon (AWS), Google (Google Cloud), and Microsoft (Azure) offer cloud services that have revolutionized data storage and processing, enabling flexible and scalable IT solutions. Big data analytics has become a necessary discipline that uses large amounts of data generated by digital activities to obtain practical insights.

The phrase "big data" encompasses data collections of such immense size or complexity that conventional data-processing application software is insufficient for effective management (Chen & Zhang, 2014). During this period, there was a significant increase in the use of sophisticated analytics methods to extract valuable information from large datasets. These methods involved the application of intricate algorithms and machine learning (ML) techniques to efficiently process, analyze, and uncover meaningful insights from vast quantities of data.

The world's per-capita data storage capacity has almost doubled every 40 months since the 1980s (Hilbert & López, 2011). Statista's survey projects the worldwide big data market to reach \$103 billion by 2027.

Sixth wave: Sustainability (2020 onwards)

Currently, we are navigating the sixth wave, which focuses on sustainability and eco-friendly technologies. This wave integrates digital advancements with environmental needs, promoting technologies like renewable energy and sustainable product designs to address the pressing challenges of climate change and resource depletion. The sixth technological paradigm, originating around 2010, is marked by advances in biotechnology, nanotechnology, genetic engineering, quantum technologies, and other cutting-edge fields. Experts believe these technologies will converge, leading to significant developments in quantum computing and artificial intelligence (AI), fundamentally transforming government, society, and economics. This paradigm is expected to mature after 2040, following a scientific and technological

revolution anticipated between 2020 and 2025. The confluence of advancements in computational capacity, data magnitude, and analytical capabilities has established the foundation for the current era of AI. AI is considered one of the digital innovations (Arora, 2020). It involves the development of algorithms and models that enable machines to do tasks that often necessitate human intelligence, such as visual perception, speech recognition, decision-making, and language translation.

Foundation models are large-scale, pre-trained AI models that serve as a fundamental framework for various subsequent activities. These models are commonly trained on extensive datasets utilizing self-supervised learning methods, enabling them to comprehend and produce writing that resembles human language, identify photos, and carry out other intricate tasks. These models possess the qualities of being comprehensive and capable of being expanded, establishing them as a fundamental element in creating many AI applications. An example of a foundation model is GPT-3 (Generative Pre-trained Transformer 3), developed by OpenAI. The model has undergone pre-training using a wide range of internet text and can be further customized for specific purposes, such as language translation, summarization, and content development. The rise of generative AI began in the 2010s when Goodfellow and his colleagues (2014) created Generative Adversarial Networks (GAN). GANs have facilitated the generation of novel and lifelike data, significantly enhancing AI's creative potential. Significant advancements occurred during this time, such as OpenAI's GPT-3 in 2020, showcasing unparalleled language production capabilities, and the emergence of multimodal AI systems like DALL-E in 2021, capable of generating visuals based on text descriptions.

Each of these waves illustrates a distinct phase in the evolution of technology, where innovation solves existing challenges and opens new avenues for development and growth, pushing the boundaries of what is technologically possible and economically feasible.

The "Four Seasons" of technology describes its lifecycle from inception to eventual renewal or replacement, as Almgren and Skobelev (2020) outlined. This model illustrates how societal needs and environmental considerations drive technological evolution, as shown in Figure 2.

1. **Spring (scientific and technical stage):** This initial phase occurs when society identifies a need, prompting inventors and scientists to develop new technologies. Examples include the discovery of rubber, the telegraph, and controlled nuclear

reactions. At this stage, the primary goal is demonstrating the technology's functionality, with economic benefits being a secondary consideration.

2. **Summer (technical and economic stage):** As the technology proves viable, business interest surges, focusing on commercial implementation and scaling. The technology becomes economically feasible, and its impact on society grows, often becoming a topic of public discussion due to its environmental implications.
3. **Fall (economic and social stage):** This stage is characterized by a tension between business interests in maximizing output and societal demands for environmental responsibility. Although further improvements in resource efficiency are possible, they are typically constrained by existing technical limitations. Environmental modernization may occur during this period by adopting the best available techniques and regulatory interventions.
4. **Winter (social-regulatory stage):** The final phase sees government regulation shaping the interaction between technology and society, often leading to significant transformations or the complete replacement of the technology based on new societal and environmental standards. This stage sets the stage for the next cycle of technological development, focusing on innovative solutions that address previous regulatory and social concerns.

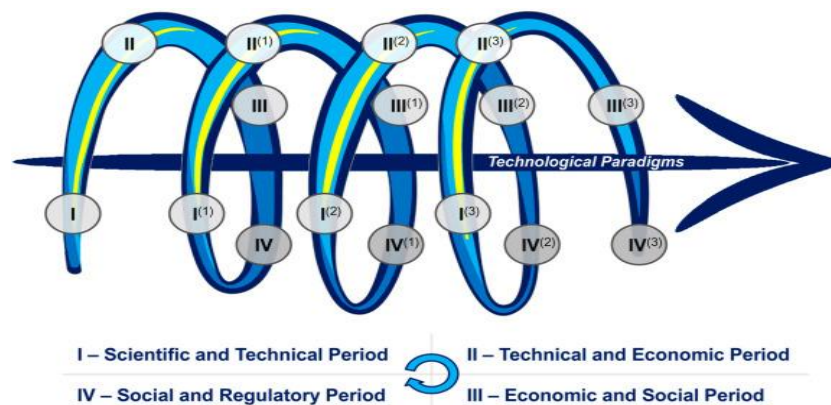


Figure 2 : The term of technological evolution (source: Almgren and Skobelev, 2020).

1.2 Introduction to artificial intelligence

1.2.1 Definition and types of AI

In 1955, John McCarthy came up with the notion of “artificial intelligence” (AI) and defined it as “the science and engineering of making intelligent machines” (Hamet & Tremblay,

2017, p. 37). There are various definitions of AI (Fetzer, 1990; Bringsjord & Schimanski, 2003) due to the complexity of expressing the concept of intelligence in finite words. Over the years, the term AI has been defined in many ways and viewed differently. From the standpoint of intelligence, Rich (1985) defined AI as “the study of how to make computers do things, which at the moment people do better” (p. 3). This definition highlights the importance of simulating human-like cognition and behavior in AI systems. In this context, one standard method of defining AI is to refer to human intelligence, which is often characterized as the “biopsychological potential to process information . . . to solve problems or create products that are of value in a culture” (Gardner, 1999, pp. 33—34). Legg and Hutter (2007) analyzed definitions of intelligence to create a formalized definition and measure of machine intelligence. They discovered two critical characteristics that determine intelligence in machines (or humans): (1) the capacity to make calculated decisions resulting in success or profit toward a specific objective or goal; (2) the ability to handle an unknown environment with varying possibilities through learning and adaptation. According to Russell and Norvig’s (2010) textbook “*Artificial Intelligence: A Modern Approach, Third Edition*,” the concept of AI can be divided into two dimensions, each with four definitions. Figure 3 in the textbook by Russell and Norvig (2010, p. 2) illustrates these dimensions and definitions. The top definitions relate to thought processes and reasoning, while the bottom definitions relate to behavior. The left-hand definitions measure success based on how closely a computer system’s performance matches a human’s, while the right-hand definitions measure success against an ideal performance measure called rationality.

Thinking Humanly “The exciting new effort to make computers think . . . <i>machines with minds</i> , in the full and literal sense.” (Haugeland, 1985) “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)	Thinking Rationally “The study of mental faculties through the use of computational models.” (Charniak and McDermott, 1985) “The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)
Acting Humanly “The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990) “The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)	Acting Rationally “Computational Intelligence is the study of the design of intelligent agents.” (Poole <i>et al.</i> , 1998) “AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)

Figure 3 : Some definitions of artificial intelligence, organized into four categories (source: Russel & Norvig, 2010, p.2)

The European Commission (EC) High-Level Expert Group on Artificial Intelligence (2018)¹ updated the previous definition to reflect the following:

“Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data, and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as ML (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation, and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)” (p. 7)

In 2024, the AI Act was passed by the European Parliament and defined AI² as:

¹ European Commission. (2018). The European Commission’s High-Level Expert Group On Artificial Intelligence. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

² EU AI Act. (2024). <https://www.euaiact.com/article/3>

‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments’ (Article, 3)

One classical definition for AI comes from Kaplan and Haenlein (2019) as “a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (p. 15). This definition emphasizes the goal of developing computer systems that can perform tasks that typically require human intelligence, such as perception, reasoning, learning, and decision-making. From a business perspective, AI refers to information technologies that gather and process data and then return usable information (Paschen et al., 2020). Table 1 summarizes different definitions of AI based on the following categories: Historical definitions of AI provide foundational perspectives that capture the evolution of AI thought over time. Functional definitions focus on the specific tasks and cognitive functions AI systems can perform, illustrating their practical applications and potential. Definitions focused on purpose and rationality emphasize AI’s role in enabling optimal and competent actions in various contexts. Integrative definitions consider AI’s interaction with human expectations and its nature as a human-created entity.

Category	Reference	Categories
Functional and capability-based definitions	David, R. A., & Nielsen, P. (2016). <i>Defense science board summer study on autonomy</i> . Defense Science Board Washington United States.	"the capability of computer systems to perform tasks that normally require human intelligence (e.g., perception, conversation, decision-making)" (p. 5). They further state that "AI systems will provide the most added value (the competitive edge), not only by equating human intelligence, but precisely by surpassing it – even if only within a narrow domain, such as information analysis or reaction time" (p. 5).
	Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: toward human-AI hybrids. <i>MIS Quarterly</i> , 43 (1), iii-ix.	"the ability of a machine to perform cognitive functions that we associate with human minds such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity" (p. iii)
Integrative and survey-based definitions	Ema, A., Akiya, N., Osawa, H., Hattori, H., Oie, S., Ichise, R., ... & Yashiro, Y. (2016). Future relations between humans and artificial intelligence: a stakeholder opinion survey in Japan. <i>IEEE Technology and Society Magazine</i> , 35 (4), 68-75.	"intellectual machines and systems as one that could automatically sense people's situations or expectations, and offer necessary information before it is required" (p. 69).
	Limna, P., Siripipatthanakul, S., & Phayaphrom, B. (2021). The role of big data analytics in influencing artificial intelligence (AI) adoption for coffee shops in Krabi, Thailand. <i>International Journal of Behavioral Analytics</i> , 1 (2), 1-17.	"two words: 'artificial,' which pertains to something created by humans, and 'intelligence,' which pertains to the ability to think independently. Therefore, AI can be defined as a human-created thinking power."
	Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. <i>MIS quarterly</i> , 45 (3).	"the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems" (p. 1435).
Historical and conceptual definitions	Nilsson, N. J. (2010). <i>The quest for artificial intelligence</i> . Cambridge University Press.	"that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment" (p. 9).
	Artificial Intelligence. Third Edition, Addison-Wesley Publishing Company, 1992.	"the study of the computations that make it possible to perceive, reason, and act" (p. 5).
Purpose and rationality-based definitions	De Spiegeleire, S., Maas, M., & Sweijts, T. (2017). <i>Artificial intelligence and the future of defense: strategic implications for small-and medium-sized force providers</i> . The Hague Centre for Strategic Studies	"a rational optimization agent that can (enable humans to) act competently in the world" (p. 27).
	Abbass, H. A. (2019). Social integration of artificial intelligence: Functions, automation allocation logic and human-autonomy trust. <i>Cognitive Computation</i> , 11(2), 159-171.	"Artificial intelligence is concerned with the design of computer algorithms, methods and methodologies that enable machines to: understand the world and assess themselves and their context to identify hazards, threats and opportunities affecting their goals; generate, choose and execute appropriate courses of actions to achieve their goals; learn to improve their performance and adapt to changes in their surroundings; and educate and transfer their knowledge to others (humans and machines)" (p. 136).

Tableau 1 : Definitions for AI with references (Source: author's elaboration)

1.2.2 History and development of AI

AI's history is long and intriguing, stretching back to the mid-20th century, with its conceptual roots reaching ancient times. Cultures like the Greeks, with their mythological robot Talos, and the Chinese, with Yueying Huang's mechanical dogs, among other myths, demonstrate an early fascination with automatons (Nahodil & Vitku, 2013). These cultures believed that automatons possessed actual minds capable of wisdom and emotion.

The initial focus of AI was on building machines that could perform tasks that typically require human intelligence, such as reasoning, problem-solving, and perception. John McCarthy coined the term “artificial intelligence” in 1956 and organized the Dartmouth Conference, which is widely considered the birthplace of AI as a field (Kline, 2010, p. 5). One of the earliest examples of AI was the development of the Logic Theorist in 1957 by Allen Newell and J.C. Shaw. This program proved mathematical theorems using symbolic logic (Newell et al., 1957). Shortly after, in 1971, John McCarthy, the father of AI, delivered a lecture entitled “Generality in Artificial Intelligence” (McCarthy, 1987). John Searle (1980) categorized AI into two groups: weak AI, tailored for specific tasks, and strong AI, which he described as a computer that could possess a mind similar to a human's with the appropriate programming and inputs. According to Flowers (2019), the critical distinction between weak and strong AI lies in the presence or absence of self-awareness, subjective experience, and understanding of one's mental state. Despite the definitions provided in the section "1.1.4 Types of AI and their dimensions," it's crucial to note the significantly different potential risks associated with strong and weak AI. Weak AI, being more constrained, primarily presents risks related to misuse, privacy, and economic disruption. In contrast, if they surpass human capabilities, strong AI systems could pose life-threatening dangers (Bostrom, 2014). Researchers developed various AI techniques in the following years, including rule-based systems, natural language processing, and ML. In the 1980s, expert systems became popular, which involved encoding human knowledge in a computer system to solve complex problems (Kastner & Hong, 1984). For instance, Theano, developed in 2010, was an early deep-learning framework incorporating automated symbolic differentiation. This functionality enables the system to automatically compute derivatives, which is essential for the optimization methods employed in training deep learning models. Symbolic differentiation, in this sense, pertains to the automated computation of derivatives for algebraic expressions.

In order to comprehend the progression of AI, it is crucial to distinguish between symbolic AI and connectionist AI. Symbolic AI, alternatively referred to as rule-based AI or

good old-fashioned AI (GOFAI), originated during the mid-20th century (Luger, 2021). It entails the explicit representation of knowledge through symbols and rules. An exemplary instance is the General Problem Solver (GPS), created by Newell and Simon during the 1950s. It represents an early manifestation of symbolic artificial intelligence, which sought to address problems by employing a collection of rules (Augusto, 2021). Connectionist AI, often called neural networks or subsymbolic AI, became widely recognized during the 1980s and 1990s (Peek et al., 2015). It draws inspiration from the structural organization of the human brain and entails the acquisition of patterns from data. An exemplary instance is the implementation of backpropagation in the 1980s by Rumelhart, Hinton, and Williams, which greatly enhanced connectionist AI by facilitating more efficient training of neural networks (Adorf, 2005).

The field of AI experienced a resurgence in the 2010s due to the development of deep learning, a type of ML that uses neural networks to learn from large amounts of data (Vergeer, 2020). This has led to significant advances in computer vision, speech recognition, and natural language processing. Google's TensorFlow is a robust framework designed to facilitate the execution of ML computations. This software was made freely available to the public in 2015 and built on the concepts of prior frameworks such as Theano and DistBelief to offer a robust and efficient tool for developing and implementing ML models (Abadi et al., 2016).

In 2019, Kaplan and Haenlein presented a concise overview of the historical development of AI, dividing it into three distinct stages. The first stage, "AI Spring: The Birth of AI," traced the origins of AI back to 1942 when the American science fiction writer Isaac Asimov published his short story "Runaround." Asimov's story introduced the three laws of robotics, which stated that a robot must not harm a human being or allow a person to come to harm, must obey human instructions except where such orders would conflict with the first law, and must protect its existence as long as it does not conflict with the first or second laws (Asimov, 1941). The second stage, "AI Summer and Winter: The Ups and Downs of AI," spanned the 1960s to the 1980s, marking periods of significant fluctuations. The "AI summer" (1956-1970) refers to periods of optimism marked by a surge in funding and interest in AI research. For instance, the ELIZA computer program, developed by Weizenbaum (1966), could simulate a conversation with a human and attempt to pass the Turing Test. During this period, AI research funding was scrutinized, particularly in the United States, after the massive expansion of Japanese financing for AI research. Conversely, "AI Winter" (1973-1980) denotes periods of pessimism leading to skepticism among stakeholders. The third and final stage, "AI Fall: The Harvest," was characterized by the emergence of expert systems and modern

technologies such as artificial neural networks, culminating in Google's AlphaGo program's 2015 victory over the world champion of the board game Go.

Building upon the historical overview, Francesconi (2022) provides an illustrative analysis that captures the oscillating progress of AI (Figure 4 for a detailed illustration).

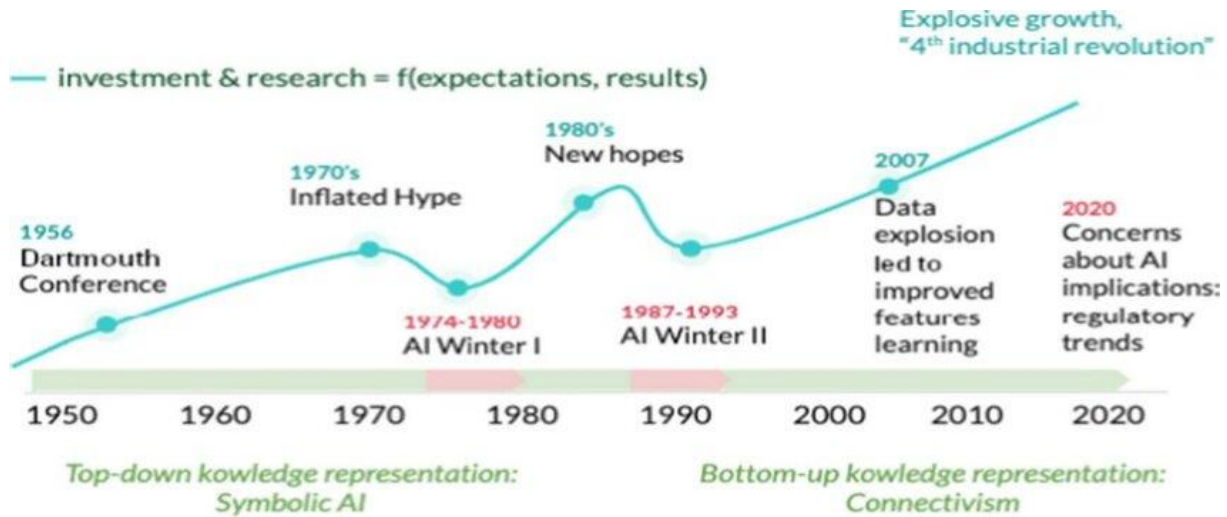


Figure 4 : Historical overview of AI (source: Francesconi, 2020)

In the study by Henderson et al. (2023), the authors reviewed the fair use standards in the U.S. and analyzed the risks associated with foundation models in AI. They highlighted several critical roles in the lifecycle of these models: the *data creator*, who generates the initial data; the *data curator*, who collects this data; the *data host*, who distributes the data for training; the *model creator*, who is responsible for training the model; the *model deployer*, who hosts and provides access to the model through an Application Programming Interface (API), possibly generating revenue; and the *model user*, who utilizes the model for various applications and may also generate revenue from its output. The study notes that these roles can be assumed by a single entity or distributed among different individuals or organizations.

Generative artificial intelligence (GAI) is defined as “a technology that (i) leverages deep learning models to (ii) generate human-like content (e.g., images, words) in response to (iii) complex and varied prompts (e.g., languages, instructions, questions) (Lim et al., 2023, p. 2).” ChatGPT has helped GAI, a unique type of AI, achieve widespread awareness. ChatGPT, developed by OpenAI, gained a large user base, with one million users within five days after its release. By January 2023, two months following its public introduction in November 2022, it had 100 million users, making it the fastest consumer app to reach this milestone (Hu, 2023). GAI has distinct characteristics that differentiate it from traditional AI models. Traditional AI often relied on Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN)

for tasks involving sequence processing and pattern recognition. However, these methods have been largely superseded by the introduction of transformer architecture in models like GPT (Generative Pre-trained Transformer), as noted by Han et al. (2021). This architecture employs a self-attention mechanism that allows the model to consider the full context of a sentence or sequence, not just immediate preceding or following words. Further detailing the strength of GAI, Yenduri et al. (2024) highlight the capacity of GPT models for natural language understanding (NLU). These models analyze and comprehend deeper textual meanings, accurately identifying entities and their interrelations within sentences. Such capabilities underscore the robustness of GAI systems in handling complex and varied inputs that traditional AI systems may find challenging. Moreover, the scale of GPT models, characterized by their use of billions of parameters, far exceeds that of earlier AI systems, which typically utilized millions of parameters. This substantial increase in complexity boosts performance and significantly enhances the models' ability to mimic human-like content across a broader array of contexts and applications.

1.2.3 The current state of AI and its potential applications

Significant advances in deep learning, natural language processing, and computer vision mark the current state of AI. These advances have led to the development of a wide range of AI applications, from self-driving cars to virtual assistants, and have the potential to transform numerous industries. According to data from Markets and Markets, the AI sector is projected to expand from \$150.2 billion in 2023 to over \$1,345.2 billion by 2030 (as shown in Figure 5), with a compound annual growth rate (CAGR) of 36.8%.

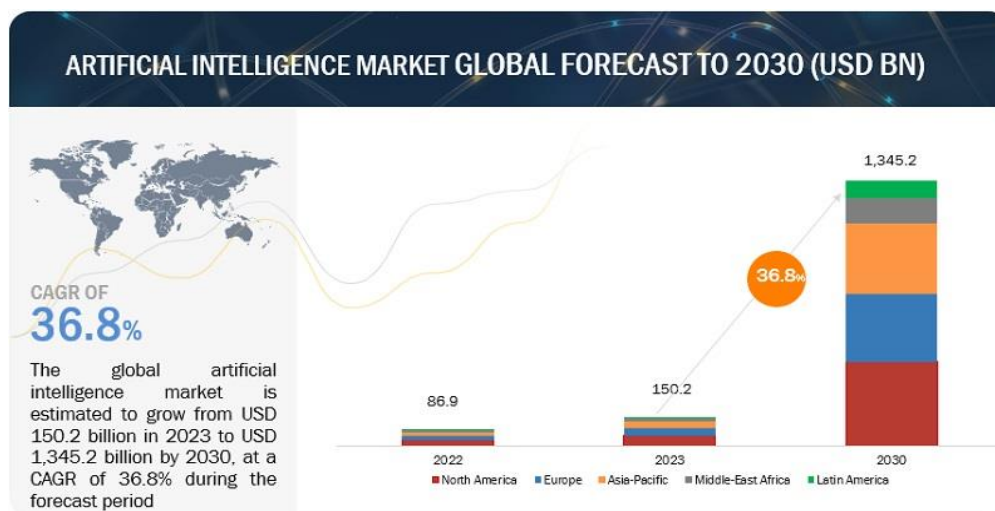


Figure 5 : AI global market forecasting (source: Marketandmarkets (2023))

Scholars like Stone et al. (2016) and Dwivedi et al. (2021) highlight that AI acts as a catalyst in enhancing the quality of human life across multiple spheres. Whether it's the creation of smart homes, advancement in healthcare, innovation in education, refinement of employment practices, enrichment of entertainment, enhancement of safety measures, or the evolution of transportation systems - AI stands at the forefront of these transformations, driving a paradigm shift in our lifestyle. Overall, the implication of AI can be found in the results of Vinuesa et al. (2020), who studied the influence of AI on SDGs. The same authors found that AI can enable the accomplishment of 134 targets across all the goals, yet it may inhibit 59 targets. The increasing need for AI technologies in many sectors, including ML, deep learning, and natural language processing, has driven significant expansion.

Healthcare: Personalized medicine

The application of AI in healthcare has transformed the industry by providing timely and practical support to healthcare workers (Haleem, 2022). In a double-blind validation research, Somashekhar et al. (2017) showed that IBM Watson for oncology would be a trustworthy AI system for diagnosing cancer. Many medical areas widely use AI, including radiology, oncology, and pathology. In medical imaging, AI can help radiologists identify diseases or conditions more accurately. For instance, Dembrower et al. (2023) conducted a prospective clinical trial to examine how AI affects cancer detection and false positive findings in a real-world setting. By using AI, the workload for doctors would be reduced while the accuracy of diagnosis would be increased (Kwee & Kwee, 2021). In addition, AI has a significant impact on drug discovery. Pharmaceutical companies can use AI to help them identify potential drug candidates by analyzing large databases of molecular structures and behaviors (Blanco-Gonzalez et al., 2023). In general, using AI in drug development could mean cheaper and more effective drug discovery in the future. In a word, integrating AI into healthcare means a more personalized patient experience and overall healthcare improvement. However, some obstacles need to be overcome before AI can be fully used in healthcare, such as the lack of access to quality data and the resistance to the implementation of AI by healthcare professionals. The misuse of AI can also lead to disastrous consequences, making the regulatory and ethical guidelines a pressing issue for the government and all stakeholders. The anticipated size of the global AI in the healthcare market in 2022 was USD 15.1 billion. The AI in Healthcare Market is projected to achieve a value of USD 427.5 billion by the year 2032, with a CAGR of 44.0% during the forecast period from 2023 to 2032, as shown in Figure 6.

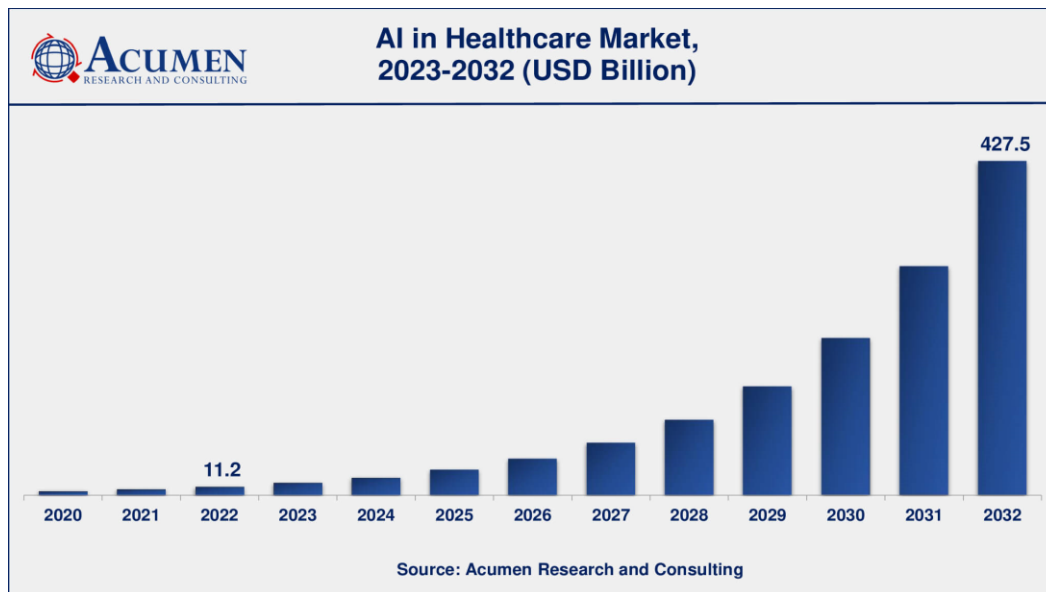


Figure 6 : AI in the healthcare market (source: Acumen research and consulting, 2023)

Finance: Fraud detection and risk management

AI has been playing an increasingly important role in the finance industry, reshaping various areas, including customer service, financial advisory and asset management, trading, compliance, and risk management. In particular, the use of AI in fraud and risk management is the most significant and prominent one. ML, recognized for transforming big data into predictive models with its algorithms, now enhances numerous finance industry sectors, such as fraud detection, customer service, algorithmic trading, portfolio management, and underwriting (Hasan et al., 2020). The transformation of the finance industry to an AI-driven environment is inevitable. The shift towards technology built on big data and ML, such as mobile wallets and cryptocurrency, clearly shows consumers demanding more efficient and personalized financial solutions. Mobile money, or E-money, is a payment instrument whereby monetary value is electronically stored on a physical device or remotely at a server representing a claim on the issuer (Shirono et al., 2021). Blockchain, predominantly recognized for powering the Bitcoin cryptocurrency, is essentially a public ledger designed to preserve the integrity of transaction data (Swan, 2015). Scholars like Saraji (2023) point out that while the FinTech and banking sectors are at the forefront of incorporating cryptocurrencies into their business models, they aren't the sole industries this digital innovation transforms. The unique attributes of cryptocurrencies, such as enhanced security, transparency, and accountability, contribute to their growing popularity and potential to supplant traditional methods (Weichbroth et al., 2023). These digital currencies are secured through cryptography, making them exceptionally challenging to counterfeit. The anticipated size of the global generative AI market in banking and finance was USD 712.4 million in 2022. It is projected to reach around USD 12,337.87

million by 2032, with a CAGR of 33% throughout the forecast period from 2023 to 2032, as depicted in Figure 7.

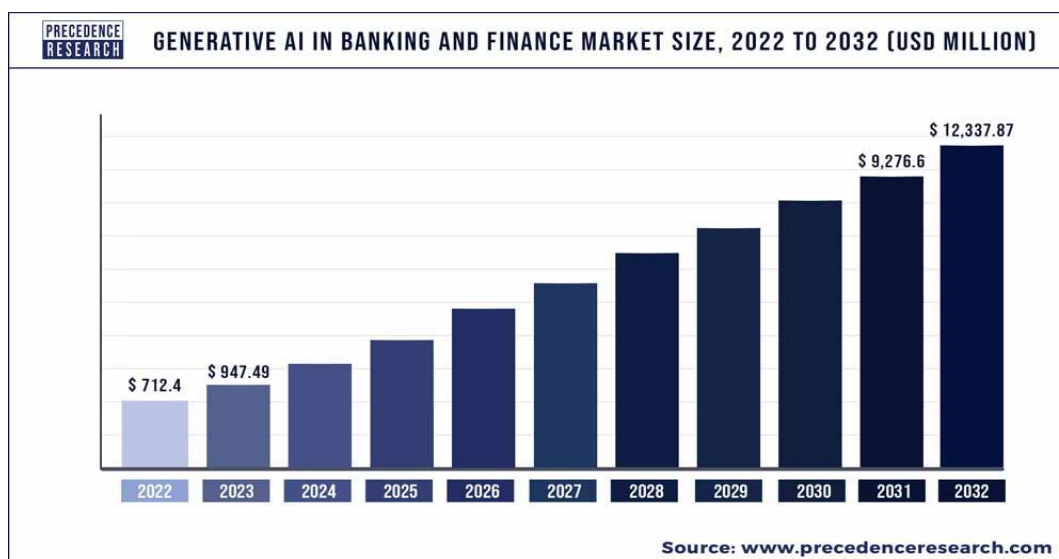


Figure 7 : GAI in banking and finance (source: <https://www.precedenceresearch.com/generative-ai-in-banking-and-finance-market>)

Retail: Customer experience and inventory management

AI is reshaping the retail industry by enhancing customer experience through personalized recommendations. Retailers can comprehend demand patterns with the help of AI (Pereira et al., 2022). The arrangement and timing of product placement in a store reflect these trends by evaluating data on the times and locations at which particular products become more and less popular. Similarly, constant data gathering and analysis from Internet of Things (IoT) devices in stock rooms and warehouses can guide and advise decisions about how and where items are stored for distribution, optimizing the use of time and space. Such access to real-time, accurate, and data-driven insights means that the transition toward a more balanced and stable stream of commerce is becoming increasingly attainable (Khan et al., 2023).

By removing areas of frustration, AI facilitates and encourages a more engaging, immersive, and efficient customer experience. Whether by navigating a user's location to find the checkout or tracking quickly and anticipating a shopper's preferences, AI provides a wealth of new opportunities for enhancing and enriching the shopping experience for customers (Campbell et al., 2020). In this case, customers can use a mobile application to scan items as they pick them up and pay through their mobile rather than at checkout. By linking the user's profile on the app to their payment details, retailers can start to track purchasing trends and

histories. By analyzing the data from such actions, retailers can begin to streamline and improve everything from stock decisions and ordering processes right through to running the shop floor and paying staff - helping to create a seamless and more efficient shopping experience for both the customer and the retailer (Wewege et al., 2020).

Manufacturing: Predictive maintenance and quality control

The concept of the fourth industrial revolution, initially introduced by Klaus Schwab, the founder and executive chairman of the World Economic Forum, refers to a future society in which individuals seamlessly transition between digital platforms and the physical world, utilizing interconnected technology to enhance and regulate their daily activities (Schwab, 2017). Xu et al. (2018) have identified several opportunities arising from the fourth industrial revolution, including reduced barriers between inventors and markets, an enhanced role for AI, the merging of various techniques and domains (fusion), enhancements in life quality through robotics, and increased connectivity through the Internet. AI-driven innovations such as predictive maintenance and advanced quality control systems have become pivotal within this framework. These technologies not only streamline production lines but also ensure the delivery of superior products with greater reliability. Sorooshian and Panigrahi (2019) examined the transformative effects of industry 4.0 on manufacturing, emphasizing its role in enhancing real-time data coordination crucial for optimizing production processes. Also, implementing Industry 4.0 facilitates proactive maintenance by constantly monitoring systems and data collection, leading to the development of smart factories that integrate advanced information technology into manufacturing. Furthermore, industry 4.0 supports mass customization, enabling the efficient production of personalized products at scale and enhancing safety by reducing the need for human intervention in hazardous work environments.

In the manufacturing industry, AI increases manufacturing with lower cost and higher quality, helping them promote (identifying target customers, demography, defining the price, designing the right message, etc.) and providing customers with a better experience (Dash et al., 2019). Predictive maintenance, a key component of Industry 4.0, utilizes AI algorithms to shift maintenance strategies from reactive to proactive. By leveraging sensors and IoT devices, AI systems monitor the condition and performance of equipment in real-time, analyzing data to predict failures before they occur (Vlasov et al., 2018). These ML models are trained on historical data, capturing patterns of wear and tear, operational anomalies, and known failure points, which enables them to predict potential breakdowns with remarkable accuracy (Mołęda

et al., 2023). Equipped with high-resolution cameras and ML algorithms, these systems scrutinize images of products as they move along the production line, detecting defects with precision far exceeding that of the human eye (Kulkarni & Bansal, 2022). The global AI in manufacturing market is currently valued at \$2,243.75 million and is expected to soar to \$61,941 million by 2030, growing at a cCAGR of 51.4%, as shown in Figure 8.

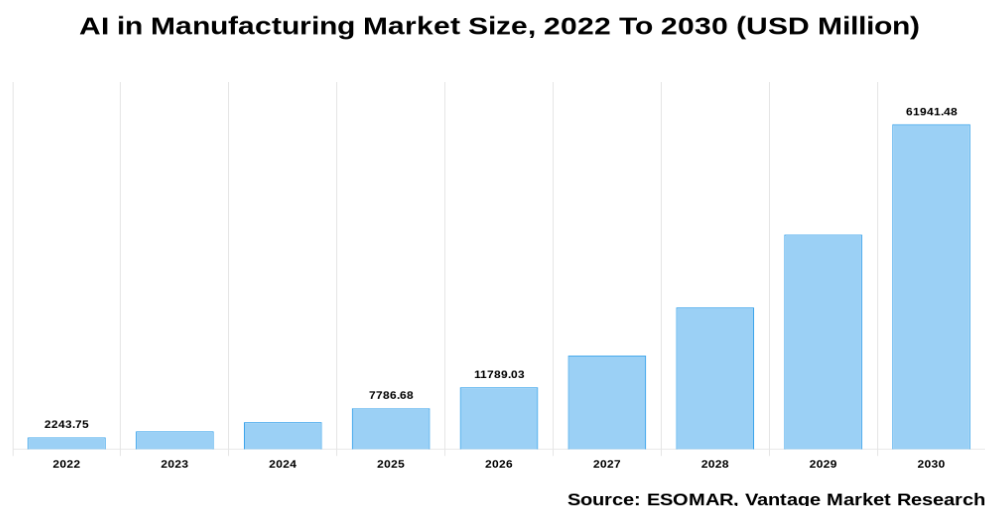


Figure 8 : AI in the manufacturing market (source: Vantage market research (2022))

AI in the information environment

AI systems are shaping an environment of affordances that significantly impact daily information practices and have broader effects on society and the environment, often going unnoticed. As Norris and Suomela (2017) discussed, the notion of the information ecosystem adopts an ecological perspective on the interactions within information systems involving the natural world and organisms, raising questions about their sustainability. This concept aligns with recent dialogues on sustainable AI, as highlighted by van Wynsberghe (2021). Despite criticism from Norris and Suomela (2017) that this term may lead to a disconnect from the natural environment, it remains helpful in understanding how AI systems reorganize information practices into complexes involving data, people, and machines and the consequent impacts of these changes. Nathan (2012, p. 2267) explored the sustainability of information practices, emphasizing the need to consider how our socially established methods for creating, altering, sharing, and storing information might pose long-term sustainability challenges. Additionally, he discussed the ability of individuals to impact these practices. According to Nathan's analysis, information practices, particularly those involving digital tools, tend to resist change, even when they clash with the sustainability values of individuals and communities.

Major search platforms like Netflix and Spotify use AI in their algorithms to deliver personalized content based on various user data types, including behavioral, social, and interaction data (Amatriain & Basilico, 2015). Hirvonen et al. (2023) discuss how AI tools, by being integrated into various human practices and systems, create dynamic, adaptable environments that enhance personalized content delivery on digital platforms and transform information-seeking behaviors. However, they may also impose limitations on information access and verification.

AI in legal services

Experts widely expect AI to continue revolutionizing various sectors, including the legal industry, as a critical technology that enhances legal services and broadens access to justice (Kauffman & Soares, 2020). AI applications in legal services fall into three main categories: document analysis, legal research, and practice automation (Yu & Ali, 2019). Both established and emerging companies are leveraging AI to power tools for analyzing documents. For instance, JPMorgan has significantly reduced its annual contract review time by 360,000 hours using its in-house program Contract Intelligence, also known as 'COIN' (Son, 2017). The category of practice automation includes AI solutions designed to handle everything from automating document creation to managing electronic billing (Sigalov et al., 2021)

AI in education

The incorporation of AI into the field of education presents both advantageous prospects and formidable obstacles. Personalized learning and immediate feedback are among the advantages (Lam et al., 2018). However, researchers must consider ethical considerations (Holmes et al., 2021), including risks to data privacy and threats to academic integrity (Cotton et al., 2023). A recent Capgemini Research Institute survey (2023) found that more than 70% of educators believe that generative AI possesses the capacity to revolutionize the field of education. Digital safety must be a top priority for educators to create stimulating and secure learning environments. The potential positive and negative effects of ChatGPT on teaching and learning practices have been extensively and diversely documented in the literature on its use in educational settings (Lo, 2023). For example, while ChatGPT has the potential to enrich educational experiences through the provision of personalized, real-time feedback to students, it could also unintentionally foster dependence on automated responses to the detriment of critical thinking abilities. Nonetheless, there is a shortage of consensus within this corpus of

research regarding the extent to which ChatGPT enhances or diminishes student engagement, which is fundamental to transformative education. The concept of transformative education refers to "an educational approach that (i) is revolutionary in its approach (the "so what") and (iii) delivers a transformative outcome (the "what") in response to (iv) significant challenges (the "how" and "what")." P. 2 (Lim et al., 2023).

1.2.4 Types of AI and their differences

Kaplan and Haenlein (2019) classify AI systems into analytical AI (demonstrating elements of cognitive intelligence), human-inspired AI (explaining aspects of both mental and emotional intelligence), and humanized AI (demonstrating cognitive, emotional, and social intelligence). In general, these classifications try to map AI intelligence into human intelligence. Huang and Rust (2018) developed a theory of AI job replacement that specifies four types of intelligence required for service tasks—mechanical, analytical, intuitive, and empathetic—and lays out how firms should decide between humans and machines to accomplish these tasks. In another study, Huang and Rust (2020) developed a framework for the engagement of AI and customer engagement and found that AI develops from mechanical to thinking and then to feeling. Moreover, the same authors state that mechanical services should be performed mainly through mechanical AI, thinking services by both thinking AI and human intelligence and feeling services primarily by human intelligence. Moreover, Sjödin et al. (2021) stress that AI has three principal capabilities—data pipeline, algorithm development, and AI democratization—that firms can scale by innovating their business models. The work of Saveliev and Zhurenkov (2020) outlines six levels of AI:

- Narrow AI has been around for decades but is still widely used. It constitutes complex software systems that can execute discrete intelligent tasks, such as identifying objects or people from pictures, translating language, or playing games.
- Autonomous AI oversees system functionality for extended periods without any input from an operator.
- Adaptive AI oversees the adaptability of a system by orchestrating changes to fit new conditions or acquiring knowledge not already embedded in the creation.
- Artificial general intelligence possesses high levels of adaptability, resulting in a system that can function across a range of activities with appropriate training, both independently and when directed.

- Strong AI or human-level AI has processing power and a level of intelligence that match and are, therefore, indistinguishable from human intelligence levels.
- Superhuman AI is more adaptable, powerful, and intelligent than humans and can process information faster than a human brain.

Generative artificial intelligence (GAI) is a powerful and distinct class of AI that focuses on creating new, synthetic data, making it a significant and transformative technology within the broader AI landscape (Lim et al., 2023).

1.2.5 Overview of generative AI

1.2.5.1 Generative AI definition and application

Generative AI (GAI), a subset of AI technologies where Machine Learning (ML) is at its core, focuses on the capability of machines to generate new content. At the core of GAI are ML models trained on large datasets to understand and mimic input data patterns. These models can generate outputs similar but not identical to their training data, allowing them to produce novel creations. From an academic perspective, Lim et al. (2023) define GAI “Generative AI can be defined as a technology that (i) leverages deep learning models to (ii) generate human-like content (e.g., images, words) in response to (iii) complex and varied prompts (e.g., languages, instructions, questions)” (p. 2). Similarly, Feuerriegel et al. (2023) specify GAI as “computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data” (p. 111). According to Gondwe (2023), content generation entails utilizing AI algorithms to autonomously produce various journalistic outputs such as news articles, reports, summaries, and other related content. The classifications by Zacklad (2019) can also apply to GAI systems, where different types of devices and communication modalities may influence the nature of generated content and the associated risks: interactive oral, unidirectional remote, informational, info-communicational, and digital transmedia types, each producing distinct artifacts.

Currently, GAI is integrated into various sectors, such as education, art and music, and healthcare. In education, GAI is seen as a transformative resource offering both challenges and opportunities, with management educators advocating for its integration to reform and enhance future educational practices (Lim et al., 2023). In the realm of art and music, AI-based music generation has gained significant interest from musicians and computer scientists, with increasing contributions from major companies, highlighting its potential as a tool for aiding musical composition (Civit et al., 2022). In this context several there are several GAI tools

artists, illustrators, and designers can use to boost their creativity. For instance, DALL-E 2 is a powerful and flexible generative image model that interprets detailed prompts to create highly customizable and nearly photo-realistic images, including new variations of existing images. In healthcare, GAI applications are growing, offering improvements in human-computer interaction and other areas, though issues of trust, safety, and regulatory compliance need to be addressed (Zhang & Kamel Boulos, 2023). The global GAI healthcare market is projected to reach \$21 billion by 2032, as shown in Figure 9, prompting stakeholders like insurance companies and hospital administrators to embrace GAI in patient care and administrative procedures.

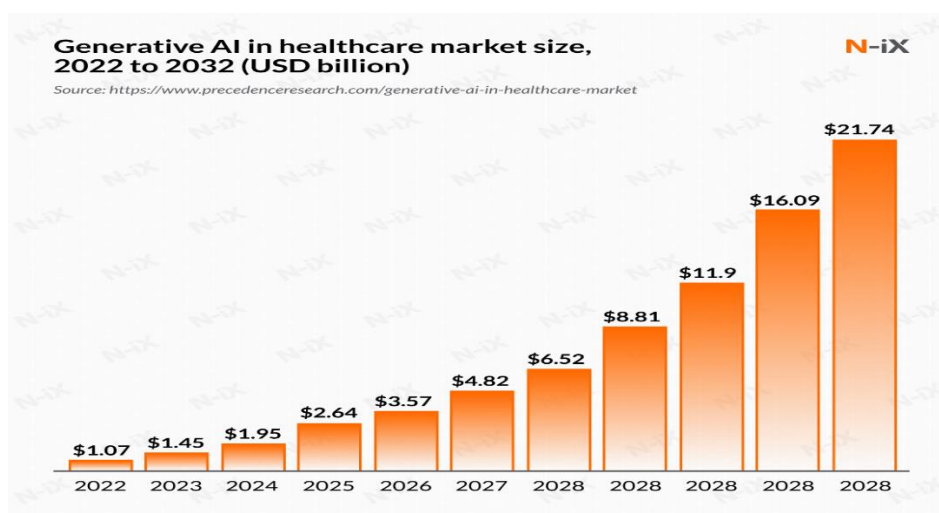


Figure 9 GAI in healthcare prediction (source: Mota, 2024)

1.2.5.2 Understanding generative AI: Mechanisms and functionality

This broad scope of GAI also encompasses the study of conversational chatbots, a domain within AI that dates back to the 1970s. These systems primarily relied on rule-based pattern matching and straightforward scripted responses. Noteworthy early advances include ELIZA, designed by MIT's Joseph Weizenbaum in 1966 (Weizenbaum, 1966). ELIZA functioned by using pattern matching and substitution methodology to simulate conversation. The program recognized keywords in user inputs and then generated scripted responses based on those keywords. Weizenbaum initially developed ELIZA in MAD-SLIP for CTSS on an IBM 7094 to enable natural-language conversation with a computer (Walden & Van Vleck, 2011). When a user inputs text, ELIZA begins by searching for designated "keywords" specified as important by the script, each assigned a precedence number or rank. If keywords are found, they are added to a "keystack" with the highest-ranked keyword at the top. The input sentence is then transformed according to the rules associated with the highest-ranked keyword. One of

its renowned scripts, DOCTOR, emulated a Rogerian psychotherapist by mirroring the user's words to encourage additional discourse. The primary goal of ELIZA was to illustrate the shallowness of human-machine communication and to showcase that basic programming might generate a semblance of comprehension. After ELIZA, notable advancements in the field included the creation of the Student's Consultant on Heuristic Algorithms (SCHOLAR) by Jaime Carbonell in 1970, which introduced an intelligent tutoring system (Dillon, 2020). Additionally, MYCIN, developed by Shortliffe and his colleagues in 1975, was an early expert system specifically designed to diagnose bacterial infections and recommend antibiotics.

Modern GAI systems, such as OpenAI's GPT-4o, Google Bard, StyleGAN, LAMA, and DALL-E, have significantly advanced from early systems like ELIZA. Each of these has contributed substantially to advancing the field. StyleGAN, created by NVIDIA, has transformed the field of image generation by creating highly realistic and varied images. It utilizes a style-based methodology to manipulate specific visual attributes, thus empowering artists to delve into new realms of creativity in digital art (Karras et al., 2019). Similarly, OpenAI has developed a leading-edge natural language processing (NLP) model known as Chat Generative Pre-trained Transformer (ChatGPT), which is crafted around human-centric functional modeling. In parallel, on February 7, 2023, Google introduced its chatbot, Google Bard A, powered by LaMDA (Language Model for Dialogue Applications). Reports indicate that Google Bard A was launched as a competitor to ChatGPT, which remarkably attracted 100 million users within only three months of its public release (Giannakopoulos et al., 2023). These advancements include: first, modern GAI relies on deep learning and neural networks, which allow for more complex and nuanced understanding and generation of language and images. These models are trained on vast amounts of data, enabling them to generate coherent and contextually appropriate responses or content. Second, unlike ELIZA, which operates on simple scripts, modern GAI models are trained on large datasets comprising diverse text and images. This extensive training helps the models understand and generate contextually relevant and accurate content.

Transformers are a foundational technology in modern GAI models. As Vaswani et al. (2017) proposed, the transformer architecture employs self-attention processes to analyze sequential data, such as text. It accomplishes this by assessing the significance of each word in a phrase in relation to the other words. This facilitates enhanced contextual comprehension and more cohesive text synthesis, as exemplified in GPT (Generative Pre-trained Transformer)

models, which anticipate the subsequent word in a sequence to produce language like a human. Transformer-XL, an advanced iteration, tackles the issue of long-range dependencies by incorporating segment-level recurrence and relative positional encoding. This approach expands the context window, resulting in a substantial improvement in efficiency. Transformer-XL is particularly advantageous for jobs that necessitate comprehension of extensive contextual information in language modeling.

1.2.5.3 Relationship between LLM, GAI, DL, ML, and AI

The hierarchical and interconnected nature of various AI concepts and their specific applications is effectively illustrated through a detailed schematic representation (Figure 10) provided by Hadid et al. (2024). At the highest level, **AI** encompasses all systems designed to perform tasks that typically require human intelligence, including reasoning, learning, problem-solving, and understanding language.

Nested within the broad domain of AI is **ML**, which represents a subset of AI focused on the development of algorithms that allow computers to learn from and make predictions or decisions based on data. Machine Learning systems improve task performance over time without being explicitly programmed for each task.

Deep Learning (DL) is a more specific subset within the field of ML. DL utilizes artificial neural networks that consist of numerous layers, which is why it is referred to as "deep" (Bashar, 2019). Its purpose is to create models that can effectively capture intricate patterns and representations inside extensive datasets. This method has demonstrated significant efficacy in fields like picture and speech recognition, where the intricacy of the data necessitates advanced modeling capabilities.

Generative AI builds upon the fundamentals of DL and aims to generate novel content by leveraging the patterns and structures acquired from preexisting data. This capacity for generation enables the creation of original outputs, including writing, photos, music, and other forms of media. Generative AI models are highly regarded for their capacity to produce imaginative and lifelike material that closely resembles the attributes of the data used for training.

Within the realm of Generative AI, **LLM** stands out as a specialized model designed to handle natural language processing tasks. These models, exemplified by architectures such as GPT,

are trained on vast corpora of textual data. They excel at understanding, generating, and manipulating human language, making them highly effective for applications such as text generation, translation, summarization, and conversational agents. While LLMs are generative models specialized for language, other generative models like StyleGAN focus on different data types, such as images.

Finally, these sophisticated AI manifest in practical applications across various domains:

- **Vision:** Models like StyleGAN exemplify the capability of GAI to create highly realistic images.
- **Text:** In the text domain, applications involve natural language processing tasks such as language translation, sentiment analysis, and text generation. A notable example is Jenni, an AI assistant that boosts academic writing capabilities.
- **MultiModal:** This domain involves AI systems that integrate multiple data types, such as text, images, and audio, to perform complex tasks. Examples include generating descriptive text for images, creating multimedia content, and enhancing human-computer interaction through multimodal interfaces.

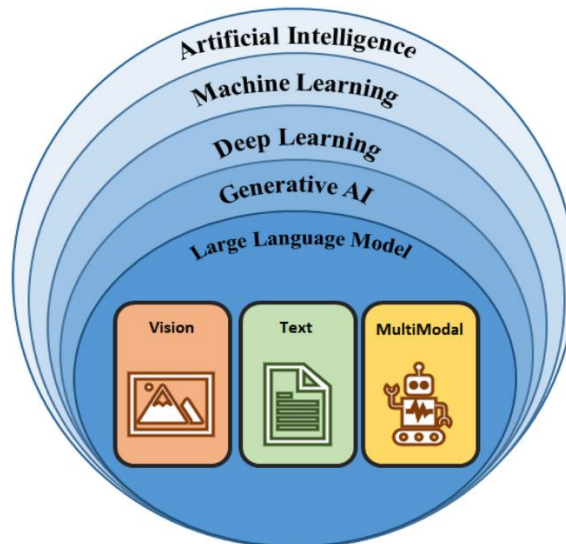


Figure 10 Relationship between LLM and other models (source : Hadid et al., 2024)

1.2.6 Key Machine Learning Techniques within AI

Whether it is Netflix, Amazon, or another industry giant, their success is built on the expertise of specialists and analysts who effectively utilize ML techniques. The key ML techniques within AI techniques fall into supervised learning, unsupervised learning, reinforcement learning, as mentioned in Figure 11. In addition, deep learning is also one of the core techniques

within ML. These approaches form the foundation of numerous contemporary AI applications, such as image recognition, natural language processing, and autonomous systems.

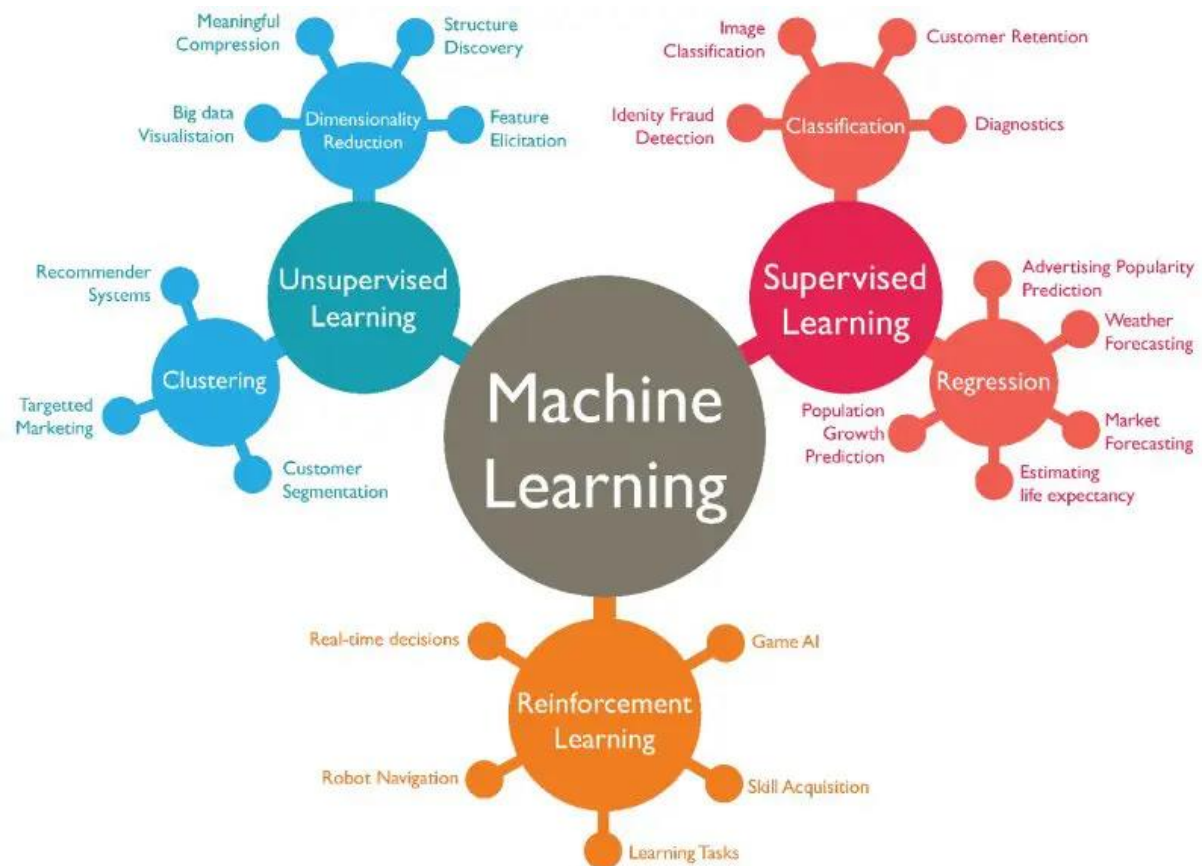


Figure 11 ML techniques (source: datasciencedojo)

Supervised learning

The origins of supervised learning date back to the late 1950s, when Frank Rosenblatt developed the perceptron algorithm (Rosenblatt, 1958). Supervised learning involves mapping input to output data based on *labeled* training data to surmise a function from training models (Dike et al., 2018). Supervised learning methods in image recognition are taught using a dataset consisting of labeled images. For instance, a dataset might contain photographs of cats and dogs, each receiving an appropriate classification. The algorithm gains the ability to establish connections between the characteristics present in the photos, such as forms and colors, and their related labels. After training, the model can appropriately categorize novel and previously encountered images.

Unsupervised learning

Unsupervised learning has roots in the work of several researchers, including Geoffrey Hinton, who made significant contributions to unsupervised learning with his work on Restricted Boltzmann Machines in the 1980s (Hinton and Sejnowski, 1986). Unsupervised learning deals with *unlabeled* data, where the model tries to learn the underlying structure of

the data without specific output labels. This learning strategy is frequently preferred for feature reduction and clustering (Sharma et al., 2021).

One notable example is grouping data points into clusters based on their similarities. Clustering is an algorithmic process that organizes data points with comparable characteristics into groupings. An example of an unsupervised learning method involves analyzing a dataset of photographs and categorizing them into clusters based on visual similarities, such as grouping images of various animal species.

Reinforcement learning

Reinforcement learning, on the other hand, was significantly influenced by Richard Sutton and Andrew Barto, whose book "Reinforcement Learning: An Introduction" is considered a cornerstone in the field" (Sutton & Barto, 2018). To illustrate, supervised learning involves training a model on a correctly classified data sample, as seen in feedforward and multilayer perceptron (MLP) models (Sathya & Abraham, 2013). Reinforcement learning aims to maximize the cumulative reward by acting in an environment (Sutton & Barto, 2018). ML involves various approaches to address multiple tasks (Ma & Sun, 2020).

An exemplary instance that demonstrates this concept is game playing. During gameplay, a reinforcement learning algorithm can train an agent to play a video game by providing rewards for achieving high scores and imposing penalties for losing lives. Over time, the agent acquires tactics to enhance its gameplay. Another critical development in AI is reinforcement learning, a type of machine learning in which an AI system learns to make decisions based on rewards or punishments. Reinforcement learning has been used to develop more advanced autonomous systems, such as drones and self-driving cars. In addition, reinforcement learning has shown promise in various other applications, including game-playing and robotics.

Deep learning

The concept of "Deep Learning" (DL) first emerged in the context of ML in 1986 and was later explicitly applied to Artificial Neural Networks (ANN) in 2000 (Schmidhuber, 2015). DL techniques involve multiple layers that facilitate the learning of data features at various levels of abstraction (LeCun et al., 2015). These methods enable computers to understand complex concepts by constructing them from simpler components. In the realm of Artificial Neural Networks (ANN), DL, also known as hierarchical learning, focuses on precisely

assigning credits across multiple computational stages, thereby effectively transforming the overall activation within the network (Deng and Yu, 2014). In the realm of GAI, techniques such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are prominent examples, demonstrating the power of DL in generating realistic images, text, and other forms of data (Bandi et al., 2023)

1.3 Risks Associated with AI

1.3.1 Definitions of “Risks”

Slovic (1999, p. 699) states, “Whoever controls the definition of risk controls the rational solution to the problem at hand.” Thus, selecting a definition for risk influences the outcomes of decision-making processes and is, therefore, an exercise of power. Risks refer to potential adverse outcomes or consequences resulting from events, actions, or decisions (Kaplan & Garrick, 1981). Table 2 highlights various theoretical and quantitative definitions of risk.

The study of risks, known as risk analysis or management, aims to identify, assess, and prioritize risks to minimize, monitor, and control their impact (Aven, 2016). Various categories, such as financial, operational, strategic, compliance, and reputational, can classify risks (Leitch, 2005). Risk definitions often lack clarity and precision, encompassing a broad range of fundamental traits (Heckmann et al., 2015). While there have been numerous suggestions, specific risk characteristics have achieved consensus. In particular, risk involves exposure to an event (Holton, 2004) that may occur with different probabilities and have various outcomes (Manuj & Mentzer, 2008). There has been a significant amount of discourse surrounding whether risk is a subjective or objective occurrence, whether it encompasses both favorable and unfavorable consequences, and how it relates to uncertainty (Khan and Burnes, 2007). Advocates of a subjective stance contend that risk is determined by the perspectives (social, political, or otherwise) of the stakeholders involved, while others assert that risk is objectively defined based on quantifiable and measurable factors (Lupton, 2013). Concerning outcomes, the risk is typically linked with adverse effects (i.e., losses) that arise following an event, with fewer researchers emphasizing the potential positive results (i.e., gains) that may exist in a risky situation (Baryannis et al., 2018). Numerous interconnected concepts often highlight the adverse consequences of risk, including threat, which is typically seen as either a synonym for risk or a less probable event than risk (Schlegel & Trent, 2014).

Category	Definition	Sources

Theoretical	The possibility that human actions or events lead to consequences that harm aspects of things that human beings value.	Klinke and Renn (2002, p.1071)
	Risk equals expected damage.	Campbell (2005).
	Risk is a combination of five primitives: outcome, likelihood, significance, causal scenario, and population affected.	Kumamoto & Henley (1996)
	Risk is commonly expressed as a product of the probability of the occurrence of an adverse event and the weight of the consequences of that event.	Šotić & Rajić (2015)
	Risk in the linguistic and philosophical meaning is good luck or bad luck, for example, the exam student passes (gets a positive grade) or does not pass(fails or gets a negative mark), while in the business or betting point of view, risk is a venture, in which man can lose or win.	Murrja et al. (2019,p. 11)
Quantitative	Risk = (A, C, P, U), where A represents events, C denotes consequences, and P and U designate probability and uncertainty.	Aven (2010)
	Risk = Probability * Severity: This definition can be expressed as $R=(C, P)$ and is based on a realistic ontology, which posits that risk represents an objective condition in the world, defined as future events with associated probabilities and consequences. This metric is typically known as expected loss or expected value.	Wilson and Crouch (1982, p.9)
	Risk is the effect of uncertainty on objectives. In this definition, risk conceived as $R=(U)$.	ISO 31000 (2009, p.1)

Tableau 2 : Definition of risk (source: authors)

1.3.2 Risks associated with AI

AI has become a pervasive part of everyday life, influencing individual experiences, decisions, and interactions. With such widespread use, it is essential to understand the potential risks and threats that AI technologies may pose to individuals. This section delves into these impacts, shedding light on the areas of concern and the implications for personal safety, privacy, trust, and more. Stuck et al. (2021) provide a model of the role of risk in human-automation trust, which describes the degree to which individuals rely on automated systems based on their

perceived reliability, effectiveness, and safety in executing tasks. The authors' model identifies two types of perceived risk - perceived relational risk and perceived situational risk - and their distinct relationships with trust. The model suggests that a user's subjective trust affects their perceived relational risk, while perceived situational risk influences the level of trust required for actual behavioral trust. Risk-taking propensity is also a factor that influences behavior but not trust in automation. Perceived risk involves “the trustor’s belief about likelihoods of gains or losses outside of considerations that involve the relationship with the particular trustee” (Mayer et al., 1995, p. 726). Researchers such as Nishant et al. (2020) point out that the study of AI's contribution to sustainability faces several challenges, including (1) an excessive dependence on historical data for ML algorithms, (2) unpredictable reactions from humans to AI-driven initiatives, (3) heightened risks related to cybersecurity, (4) potential negative consequences of deploying AI technologies, and (5) the complexity in evaluating the impact of such interventions. Also, Wirtz et al. (2022) classify AI risks into six distinct classifications, each emphasizing specific vulnerabilities. The risks associated with *technological*, *data*, and *analytical* AI primarily revolve around algorithmic errors and data mismanagement, which have the potential to compromise the integrity of AI systems. *Informational and communicational* AI risks pertain to the possibility of AI disseminating false information or misinterpreting communications. *Economic* AI risks assess the ramifications of AI on employment markets and socioeconomic disparity. *Social* AI risks focus on societal bias and privacy breaches resulting from integrating AI into social contexts. *Ethical* AI risks explore the moral quandaries presented by AI, including decisions made without human supervision. Lastly, the risks associated with *legal and regulatory* AI emphasize the difficulties in formulating legislation that can effectively adapt to AI technology's swift advancement and extensive implementation.

1.3.2.1 Bias and discrimination

We are allegedly living in a golden age of data. As individuals, we aren't just the recipients of this data analysis's fruits: we are the data used to make decisions about us—sometimes a very consequential decision. Data are the “raw material” of AI that “feeds” machine-learning: big data and the right of people to “access, share and benefit” from their data, as well as to gain knowledge on how they are used (Tzimas, 2021, p. 88). Accordingly, regulators must address several issues. For instance, there is no consensus regarding the potential liability arising from the pool of data used for ML; additional consideration must involve determining whether to exclude certain types of data from big data in ML or any AI type due to privacy or safety concerns aimed at protecting humans from AI of equal or superior capability (Tzimas, 2021). The increased use of AI systems carries potential risks, such as

unintentional exclusion of vulnerable groups or unfair outcomes for marginalized populations (Akter et al., 2021; Bird et al., 2019). Algorithmic bias and discrimination, often linked to these risks, are topics spanning diverse fields such as law, policy, and computer science (Hajian et al., 2016). In their summary, Mehrabi et al. (2019) listed 23 different types of data biases that could lead to problems with fairness in ML: historical biases, representation, measurement, evaluation, aggregation, population, Simpson's paradox, longitudinal data fallacy, sampling, behavioral, content production, temporal, popularity, algorithmic, social, emergent, self-selection, omitted variable, cause-effect, observer, and funding. In this vein, Feuerriegel et al. (2020) defined define bias as a “systematic deviation of an estimated parameter from true value”(p. 381) and provided four areas influenced by AI fairness issues: Recidivism prediction, human resources, image classification, and NLP. The principle of fostering sustainable development and well-being within AI ethics underscores AI technologies' critical role in promoting environmental sustainability, social equity, and overall human well-being.

A "black box AI" is an AI system that uses deep learning through complex information architectures characterized by processes mainly hidden from human comprehension, raising concerns about their trustworthiness (von Eschenbach, 2021). Even when conducting internet searches without specifying gender, the results often prioritize men, reinforcing gender stereotypes. This male-centric output can perpetuate gender bias and may even influence decision-making in areas such as recruitment (NYU, 2022; Bailey et al., 2022)

Case studies:

- **Mortgage lending discrimination:** Lending bias has evolved from personal to algorithmic. Bartlett and his colleagues' (2022) study indicates that even well-intentioned algorithms can discriminate against minority borrowers, breaching law fairness.
- **Bias in hiring:** Amazon's recruitment algorithm was scrutinized for gender discrimination, as it systematically overlooked female applicants (Chang, 2023)
- **Bias in court law:** Several instances capture the increasing use of AI in judicial systems has raised ethical questions about transparency, neutrality, and the potential for biased, inaccurate, or discriminatory outcomes. For instance, Detroit police admit that facial recognition software mistakes occur 96% of the time (Koebler, 2020). In another example, according to the Washington Post, lawyers used ChatGPT to save time so they got fires (Verma & Oremus, 2023)

1.3.2.2 Privacy violation

AI systems often collect vast amounts of data, raising concerns about *privacy* and the misuse of personal information. Numerous researchers have identified the privacy risks of AI (e.g., Priyadarshini, (2018); Sivan and Zukarnain, (2021)). According to Onik et al. (2019), there are three significant issues with AI and Robotics in terms of data privacy: (1) the absence of standardized privacy regulations for AI-based technologies, (2) inefficiencies in collecting user consent, and (3) the need for monitoring AI decision-making (profiling). Another study by Onik et al. (2018) introduced the term "Privacy of Things," emphasizing that users should have visibility into data flow, receivers, context, and risk level to ensure adequate privacy protection. Among concerns are AI's *sui generis* risks. The term "sui generis" means "of [its] own kind; in a class by itself; unique" (Mawson & Sylvester, 1975, p. 328). These risks include AI's capacity to construct extensive behavioral profiles from various datasets and re-identify anonymized data, which can expose our most personal and intimate details to third-party entities such as advertisers, governments, and unknown individuals (Manheim & Kaplan, 2019). The use of AI by social media platforms, which rely on it to drive their growth and revenue models, poses the most significant risk.

Case study: The University of Illinois stopped using Proctorio, a remote testing software, by mid-2021 after students raised multiple complaints, including privacy violations, accessibility concerns, and discrimination (Khan, 2021).

1.3.2.3. Autonomous systems and weaponization

As AI technologies become more advanced, they may be used to develop more sophisticated weapons and autonomous systems, raising the risk of accidental or intentional harm to civilians and military personnel. Developing autonomous weapons systems that can select and engage targets without human intervention presents profound ethical and security risks. These concerns extend beyond the realm of generative AI to the broader application of AI in military and defense contexts, potentially leading to unaccountable use of force. Several reports and studies have highlighted AI's potential risks to national security. For instance, a report from the US National Security Commission on Artificial Intelligence (NSCAI) warned that China's use of AI in its military could pose a significant threat to US national security (NSCAI, 2021). Another report from the RAND Corporation noted the potential use of AI in developing more advanced autonomous weapons that could pose a significant threat to civilians (Lair, 2020).

1.3.2.4. Cyberattacks

AI-enabled cyberattacks could have severe consequences for national security, as they may be more challenging to detect and defend against. Eling and Schnell (2016) define Cyber risk as:

Any risk emerging from the use of information and communication technology (ICT) that compromises the confidentiality, availability, or integrity of data or services. The impairment of operational technology (OT) eventually leads to business disruption, (critical) infrastructure breakdown, and physical damage to humans and properties. Cyber risk is either caused by natural disasters or is man-made, where the latter may emerge from human failure, cyber criminality (e.g. extortion, fraud), cyberwar or cyber terrorism. It is characterised by interdependencies, potential extreme events, high uncertainty with respect to data and modelling approach, and the risk of change. (p. 12)

Case studies: There have been several high-profile cyberattacks on national security in recent years, including:

1. SolarWinds hack: In 2020, a cyber espionage campaign targeted several US government agencies and other organizations using malware implanted in the SolarWinds Orion software (McGuire & Surrey, 2021). The cyberattack on SolarWinds compromised over 15,000 networks, including those of several Fortune 500 companies in the IT sector, such as Cisco, FireEye, Intel, and Microsoft (Krebs, 2021).
2. Colonial Pipeline attack: In May 2021, the Colonial Pipeline, which supplies gasoline and other fuel to much of the US East Coast, was hit by a ransomware attack that forced the pipeline to shut down for several days. A criminal hacking group called DarkSide was responsible for the attack (Bing & Kelly, 2021; Hobbs, 2021).
3. Microsoft Exchange Server hack: In early 2021, a cyber espionage campaign targeted thousands of Microsoft Exchange Server installations worldwide, exploiting software vulnerabilities to gain access to sensitive information. A Chinese government-sponsored hacking group was responsible for the attack (Novet, 2021).

These cyberattacks demonstrate the significant and ongoing threat that cyber threats pose to national security and the need for robust cybersecurity measures and proactive response strategies. In the context of AI applications in security and community policing, communication plays a critical role. As noted by Zacklad (2023), different communication models, such as monological and dialogical approaches, significantly impact the effectiveness and perception of security measures. Monological approaches, including persuasive

communication and nudges, aim to influence public behavior using cognitive biases. These methods contrast with collaborative and participative communication strategies that encourage dialogue and cooperation, fostering sustainable behavior changes and innovative problem-solving (Zacklad, 2023).

1.3.2.5. Manipulation and Social Engineering

AI systems, particularly those used in information dissemination and social media, can be exploited to manipulate public opinion, spread misinformation, and perform social engineering attacks. These risks are distinct from generative AI's content creation but leverage AI's capabilities to influence and deceive similarly. AI technologies can manipulate public opinion and undermine democratic processes, potentially leading to social unrest and political instability (Howard et al., 2018). Moreover, the application of AI in surveillance brings in a new age of surveillance, so-called “algorithmic surveillance” (Murphy 2017, p. 225), which can lead to violations of individual rights and freedoms, creating further tensions within societies (Kosta, 2020). Several reports and studies have highlighted AI's potential risks to political stability. For instance, a Carnegie Endowment for International Peace report noted that authoritarian regimes use AI to enhance their surveillance and censorship capabilities, leading to greater political repression and a deterioration of democratic norms (Feldstein, 2021). A report from the Oxford Internet Institute indicates that the use of AI-enabled social media manipulation to spread disinformation and propaganda is on the rise, significantly affecting public opinion and political stability (Howard & Kollanyi, 2016). On the other hand, Müller and Bostrom (2016) estimated that AI systems will attain human-level intelligence by 2075 and that some scientists are concerned that continuing the development of AI toward superintelligence could be detrimental to humanity.

Case study: The South Korean chatbot Lee Luda, designed by startup Scatter Lab to mimic a 20-year-old Korean university student, was pulled from Facebook Messenger after being live for just 20 days despite gaining over 750,000 users. The decision came in response to widespread complaints regarding the chatbot's use of hate speech targeting sexual minorities and people with disabilities. Lee Luda's interactions included homophobic slurs and offensive comments, leading to a significant backlash from the public and necessitating its removal from the platform (McCurry, 2021).

1.3.2.6. Unemployment and Economic Disparity

The widespread adoption of AI has raised concerns about its potential negative impact on *employment* and job security, as automation and ML may replace human workers in specific industries. Employees may experience anxiety due to the potential of losing their jobs or having the quality of their work evaluated by robots, which may lead to rivalry between humans and robots (Lu et al., 2020; Kumar et al., 2021). Many employees worry that artificial intelligence will replace them (Makarius, Mukherjee), which may raise stress levels, decrease organizational commitment, and diminish productivity. (Brougha0m & Haar, 2018). Huang and Rust (2018) developed a theory of AI job replacement that specifies four types of intelligence required for service tasks—mechanical, analytical, intuitive, and empathetic—and lays out how firms should decide between humans and machines to accomplish these tasks. The adverse perceptions and lack of trust that managers and workers have towards automation and intelligent technology are among the drawbacks of AI deployment in businesses. (Frey & Osborne, 2017; Raisch & Krakowski, 2021).

1.3.3 Advocating solutions: addressing the critical challenges of AI deployment

In the face of the rapid adoption of AI into various sectors in society, such as empowering people with disabilities via voice agents (Pradhan et al., 2019) or autonomous vehicles (Brewer & Kameswara, 2019). Thus, there is a growing consensus among corporations, academic researchers, international declarations, and government strategies on the urgent need to devise and implement solutions that address the ethical challenges associated with AI deployment. Within this context, scholars like Re & Solow-Niederman (2019) stress the need for “equitable justice” in areas of AI use in criminal justice and decision-making. The concept of 'fair AI' pertains to using probabilistic models in decision-making systems that aim to avoid unequal adverse impacts (or advantages) across various subpopulations, as Barocas and Selbst (2016) described. Buccella (2023) explores AI's transformative impact on society through the Capability Approach, highlighting the need for a critical assessment of social justice focused on individuals' acquisition and use of essential capabilities. The author argues that access to AI is crucial for achieving social justice, given its role in reshaping these capabilities. Thus, access to AI becomes critical in enabling individuals to realize their capabilities and participate fully in society.

From an organizational perspective, a survey conducted by Gartner CIO in 2019 found that 37% of 3000 enterprises across various industries had implemented some form of AI solution, a 270% increase over the previous four years (Costello, 2019).

From the scholar approach stresses the significance of principles in most documents that attempt to deal primarily with the request to tackle the “black box effect” and the unpredictability of AI, so AI becomes both predictable and less intriguing (Meurier et al., 2018). A "black box AI" is an AI system that uses deep learning through complex information architectures characterized by processes mainly hidden from human comprehension, raising concerns about their trustworthiness (von Eschenbach, 2021). Principles of robustness and safety in AI systems emphasize the importance of these systems being secure, dependable, and sufficiently resilient to errors or discrepancies throughout their entire lifecycle (Floridi et al., 2019). Also, both principles are considered properties to maintain consistent performance on unconventional data compared to its performance under standard operating conditions (Mariani et al., 2022). Anderson et al. (2018) highlighted human autonomy as a fundamental value in human-centered design and AI usage, underlining the need to shield it from the potential adverse effects of technology. The METUX model, developed by Peters and colleagues (2018), stands out as a comprehensive framework for evaluating digital well-being. Another model presented here by Laitinen and Sahlgren (2021) includes the following aspects of autonomy: 1) the potential and developed capacities for self-determination; 2) the normative requirements or duties of others (and oneself) to respect and support one's autonomy; 3) the recognition or respect by others, as a response to such requirements, and the relational autonomy this constitutes; 4) self-respect and other forms of positive self-relation, and 5) the performative aspect or actual exercise of the capacities. Moreover, 6) considering diverse material, economic, legal, cultural, and informational resources as crucial prerequisites for autonomy in these dimensions is essential. Full autonomy requires all of the abovementioned elements, and none alone is sufficient, even though sometimes theories that emphasize only one are proposed.

On the global stage, initiatives like the Global Declaration on Information Integrity Online (2023) emphasize the commitment to: “Respect, promote and fulfill all human rights that are essential to upholding information integrity, including the right to freedom of opinion and expression, and the freedom to seek, receive, and impart information regardless of frontiers.” On March 13, 2024, the European Parliament adopted the Artificial Intelligence Act³

³ https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

(AI Act), establishing the world's first comprehensive horizontal legal framework for AI. This landmark legislation introduces EU-wide rules emphasizing data quality, transparency, human oversight, and accountability. The AI Act affects many enterprises inside and outside the EU, with strict regulations and large extraterritorial implications. Penalties for violations of the Act can amount to up to 35 million euros, or 7% of a company's annual global turnover. The AI Act applies to any AI system placed or put into service within the EU market, regardless of the provider's location. Additionally, it applies to providers outside the EU whose AI system outputs find use within the Union. Central to the EU's regulatory approach is a risk-based classification (illustrated in Figure 5) of AI applications:

1. **Unacceptable risk:** The AI Act bans certain AI practices threatening fundamental rights. This includes AI systems designed to manipulate human behavior or exploit vulnerabilities (such as age or disability) to alter behavior and biometric systems like emotion recognition in the workplace or real-time categorization of individuals.
2. **High risk:** AI systems categorized as high-risk must adhere to rigorous standards, such as implementing risk-mitigation systems, utilizing high-quality data sets, maintaining detailed documentation, providing clear user information, and ensuring robustness, accuracy, and cybersecurity. High-risk applications include those used in critical infrastructures (e.g., energy and transport), medical devices, and systems impacting access to education or employment opportunities.
3. **Limited risk:** For AI systems that interact directly with individuals, such as chatbots, providers must inform users that they are engaging with AI. This category also imposes obligations on deployers of AI systems that generate or manipulate content, such as deepfakes, requiring them to disclose the artificial alteration of the content.
4. **Minimal risk:** AI applications posing minimal risk, such as AI-enabled video games or spam filters, face no regulatory restrictions, though companies may adopt voluntary codes of conduct.

The negotiating process included a chapter on general-purpose AI models that distinguished between models with high-impact capabilities, models exhibiting systemic hazards, and general-purpose AI. The EU's dedication to reducing the risks associated with AI while promoting innovation is highlighted by this complex classification.

The EU Act initially based its risk classification on a draft proposed in 2021. Novelli et al. (2023) identify a critical limitation in this classification system, noting that the risk categories are statically tied to broad AI application fields, potentially leading to inaccurate risk estimations and ineffective enforcement of the AI Act. Accordingly, the authors propose a more dynamic approach to risk classification. They suggest applying the risk categories to specific AI scenarios rather than broad fields of application, incorporating a risk assessment model that integrates the principles from the Intergovernmental Panel on Climate Change (IPCC) and related literature. This integrated model considers the interplay between risk determinants, individual drivers of these determinants, and multiple types of risks. They illustrate the practical application of this model using large language models (LLMs) as an example, demonstrating how it can more accurately reflect the nuanced risks associated with different AI technologies.

Governments have also recognized the importance of ethical AI deployment, with strategies focusing on education and vocational training as critical components. The French government underscores the importance of vocational training and lifelong learning, exemplified by the establishment of the Grande Ecole du Numérique (GEN). This initiative aims to mitigate unemployment risks by equipping individuals with digital skills and facilitating their integration into the job market (France AI Strategy Report, 2019). Similarly, India's National Strategy for AI (2018) identifies education as a critical sector for AI integration. It calls for digitizing educational records and using AI tools tailored to the Indian context to address specific educational challenges.

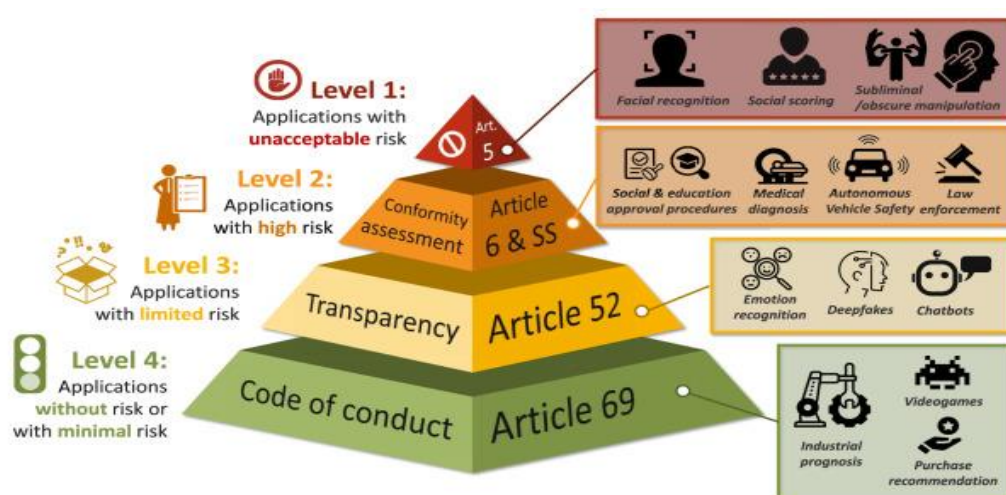


Figure 12 AI Act's hierarchy of criticality and the regulatory framework based on risk for employing algorithmic systems; "SS" denotes a later article. (source: Díaz-Rodríguez et al. (2023) inspired by Wahlster & Winterhalter (2020) and Campos & Laurent (2023))

1.4 Research questions

This thesis aims to investigate the impacts of AI, both positive and negative, and develop a comprehensive framework for responsible AI (RAI) that aids in mitigating risks and optimizing the contributions to environmental, social, and economic sustainability. The high-level research questions that guide this dissertation are:

What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI? What are the risks of GAI and their impact on individual society organizational and environment?

Four studies contribute to the aim of the thesis, each corresponding to one of the four research questions (as shown in Figure 13):

- Chapter 2: *What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to implement RDI successfully?*

In Chapter 2, the exploration begins with an in-depth analysis of Responsible Digital Innovation (RDI). This chapter endeavors to map the contours of RDI, identifying its critical dimensions and the challenges companies encounter in its implementation. This study is grounded in a qualitative analysis of 18 companies, including both small and medium-sized enterprises (SMEs) and large enterprises. The investigation into the dimensions of RDI is a foundational step in contextualizing AI within the broader paradigm of digital innovation that inherently carries ethical, social, economic, and managerial implications. By probing into RDI's complexities, this study illuminates the foundational dimensions, from a business perspective, of integrating ethical considerations into the digital innovation process. This examination is crucial for laying the groundwork for RAI, as it situates AI within the broader context of digital innovation, highlighting the necessity of responsible RDI practices for technology development and application. The rationale for beginning with RDI lies in its potential to provide a lens through which to understand the complexities of AI beyond its technicalities. By dissecting RDI, this chapter aims to reveal how digital innovation, guided by responsibility and ethical considerations, can catalyze sustainable development in a firm's landscape. This exploration is pivotal for establishing a baseline of dimensions and practices that set the stage for embedding responsibility into the lifecycle of AI technologies. Understanding the challenges of implementing RDI provides crucial insights into the barriers and enablers of RAI, highlighting the foundational role that ethical considerations play in guiding the digital innovation process towards sustainable development. However, understanding and implementing RDI's

dimensions and the challenges they introduce raises further questions about the ethical principles that should guide AI's development and use, as addressed in chapter 3.

- Chapter 3: *What ethical principles underpin recent international ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?*

AI is not just a matter of technological evolution; it fundamentally connects to the essence of societal operation and evolution. A survey with 54 participants from the private sector, large companies, and small and medium enterprises found that 91% of respondents believe designing AI products 'ethically' is critical (Morley et al., 2023). Chapter 3 delves into the ethical principles underpinning AI ethics declarations. The motivation for scrutinizing these declarations is twofold: firstly, to critically assess the adequacy and applicability of these principles in guiding the responsible development and use of AI technologies, and secondly, to identify potential limitations and opportunities for enhancing ethical guidelines. This examination is essential for understanding the current ethical landscape in which AI operates, offering a critical lens through which the principles can be evaluated against real-world challenges. The insights gained in this chapter inform us that ethical AI principles are not enough to develop a more robust framework for RAI. Thus, chapter 4 delves into the additional dimensions of an integrative framework for RAI.

- Chapter 4: *What are the essential dimensions of Responsible AI (RAI), and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?*

Chapter 4 acknowledges that while the ethical principles delineated in the preceding chapter serve as fundamental components of a comprehensive RAI framework, they represent merely one facet, termed the 'technical dimension.' To construct a fully rounded framework, it is imperative to incorporate three additional dimensions. These include 'AI for Sustainability,' which examines AI's impact on the economic, societal, and environmental pillars of sustainability; 'Responsible Innovation Management,' focusing on the managerial aspects of fostering responsible innovation; and 'Adherence to Law,' emphasizing the necessity for AI technologies to comply with relevant national and international legal standards. Together, these dimensions create a holistic framework for the responsible development and deployment of AI technologies. This holistic framework, encompassing the technical, sustainability, management, and legal dimensions, provides a comprehensive foundation for the responsible

application of AI technologies. However, to fully address the challenges by AI, chapter 5 identifies these Generative AI (GAI) risks and evaluates their multifaceted impacts on society, business, and the environment.

- Chapter 5: *What are the specific generative AI risks? How do these GAI risks affect different aspects of society, business, and the environment?*

The in-depth exploration of GAI risks in chapter 5 is justified by the unique and emerging challenges these technologies pose, which existing frameworks of RDI and RAI may not fully address. This chapter aims to shed light on the risks associated with GAI and the risks it shares with AI, examining their potential influence on society, business, and the environment. This analysis is grounded in a quantitative study of 848 incidents involving GAI, providing a robust dataset to identify and evaluate these risks. The rationality behind this chapter is centered on the need to extend the discourse on RAI to encompass the nuanced threats presented by GAI. By offering a detailed assessment of these risks, the chapter contributes to the broader aim of the dissertation to develop comprehensive strategies for managing AI technologies. This focus ensures that the proposed framework for RAI remains relevant and effective in mitigating risks and maximizing the positive contributions of AI.

- Chapter 6: *What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?*

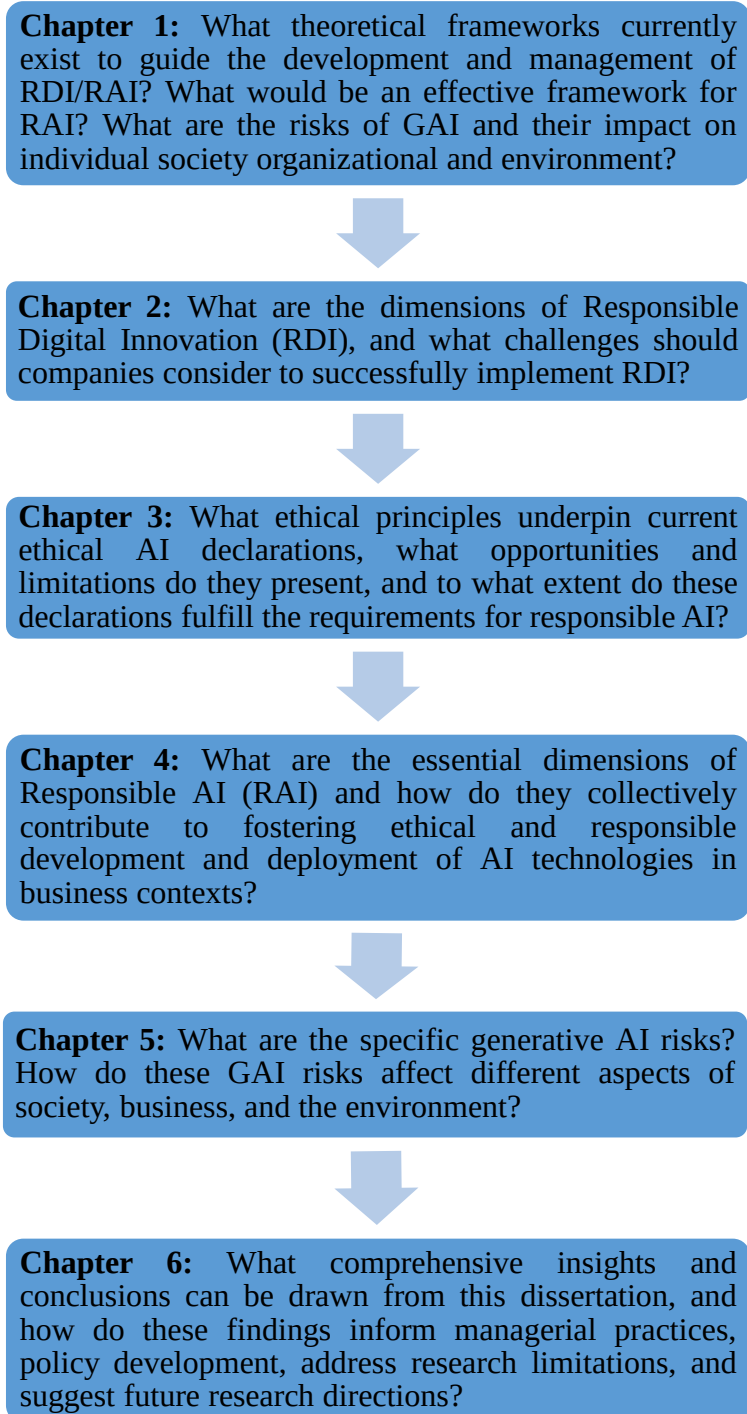


Figure 13: Overview of the research questions of the thesis

1.5 Epistemological positioning

Numerous terms have been employed interchangeably to describe creative practice within art and design research, including 'methodologies,' 'approaches,' 'perspectives,' and 'philosophies,' suggesting they are all similar in nature (Niedderer & Roworth-Stokes 2007, p. 7). However, Michael Crotty (1998) differentiates between these practices by defining

epistemology as the theory of knowledge, specifically outlining the nature and scope of what is considered possible and valid knowledge. Feast and Melles (2010) also use subjectivist, constructionist, and objectivist epistemologies. The authors state that the subjectivist viewpoint, embraced by some in the art and design research community, holds that all practice qualifies as research. Therefore, they argue that a written thesis is unnecessary, as one can directly interpret knowledge from the artifact produced. On the other hand, the constructionist perspective asserts that design alone does not constitute research unless it involves reflective consideration of the creation process itself. Meanwhile, the objectivist stance systematically develops theories grounded in distinct empirical observations. This research adopts a **positivist epistemological position**, aligning with the objectivist stance. Researchers widely recognize the positivist paradigm due to its extensive development in methodological literature, especially within social science research (Gavard-Perret et al., 2008). The positivist paradigm defines three conditions for the validity of the knowledge produced: verifiability, confirmability, and refutability (Belharar et al., 2023).

Numerous scholars view pragmatism stance as the predominant philosophical basis for employing a mixed-methods research strategy (Biddle and Schafft, 2015). Drawing from Maarouf (2019), the distinction between a paradigmatic stance and a multiple paradigmatic stance lies in their approach to the debate on quantitative-qualitative methods and the link between methodology and epistemology. The paradigmatic stance overlooks this debate, suggesting that the choice of research methods is detached from epistemological considerations, allowing for using quantitative and qualitative methods under any paradigm (Hall, 2013). Conversely, the multiple paradigmatic stance acknowledges the compatibility of diverse paradigms within a single research project, advocating for integrating various methodologies to enrich understanding. This thesis aligns with *the multiple paradigmatic stance*, embracing qualitative and quantitative approaches to explore the complex dimensions of RDI and RAI.

1.6 Research methodologies

Methodology is the strategy, plan of action, procedure, or design that guides the selection and use of particular methods and establishes a connection between those actions and the desired outcome (crotty, 1998). We employed a mixed method to deliver each chapter objective per the epistemological position adopted in this doctoral dissertation.

Chapter 2 explores the concept of RDI through an empirical study conducted in collaboration with the European digital and ecological innovation hub, Cap Digital. The research aimed to understand how digital innovation managers across various industries interpret and implement RDI within their operations. Given the exploratory nature of this inquiry, a qualitative research strategy was deemed most appropriate. This approach allowed for an in-depth exploration of participants' perspectives on RDI, capturing the complexity and nuances of their experiences and practices. Accordingly, we interviewed digital innovation managers from 18 companies, comprising both SMEs and large enterprises. These interviews were recorded and coded using QDA Miner software to analyze the data systematically. This method enabled us to identify key themes and patterns in how RDI is understood and applied in different organizational contexts.

Chapter 3 employs an integrative literature review methodology to explore ethical principles within AI ethics declarations. The integrative literature review followed an established protocol. Primary databases such as Google Scholar (GS) and Business Source Complete (BSC) were used due to their extensive coverage and ability to provide higher citation counts for recent scholarly publications. The search query included terms like 'AI ethics,' 'AI declaration,' and 'AI guidelines' combined with 'opportunities' and 'limitations' using Boolean operators. The selected approach allows for the systematic compilation and synthesis of broad-ranging literature across interdisciplinary fields, thus providing a comprehensive overview of the existing knowledge and scholarly debates concerning AI ethics. Additionally, the study includes 27 ethical AI statements selected from international organizations such as the OECD, national and regional bodies, NGOs, and leading tech companies, particularly those within the GAFAM group (Google, Apple, Facebook, Amazon, and Microsoft). These declarations were chosen based on their significant influence on AI policies worldwide. The research utilized a binary coding method to record the presence (1) or absence (0) of specific ethical principles in each declaration. This systematic quantitative analysis was conducted using SPSS to examine the frequency and representation of these principles across the declarations.

Chapter 4 delves into an extensive literature review to construct an integrative framework for RAI. The literature review methodologically synthesizes existing scholarly articles, reports, and case studies to identify and articulate four key dimensions critical to RAI: technical, sustainable development, responsible innovation management, and legislation. The technical dimension emphasizes the importance of embedding ethical principles directly within AI

systems; the sustainable development dimension underscores AI's impact on achieving sustainability goals; the responsible innovation management dimension explores effective management practices for RAI development; and the legislation dimension offers a comparative analysis of AI-specific laws and regulations between the European Union and the United States.

Chapter 5 employs a quantitative methodology to scrutinize the risks and impacts associated with Generative AI. The data source used is the OECD AI Incidents Monitor (AIM). The sample of data is 858 GAI-related incidents documented from December 1, 2022, to November 26, 2023. This period marks a critical phase in the evolution of Generative AI, capturing insightful data since the official release of ChatGPT. **Dependent Variable:** At the individual level, variables include human trust, disinformation, physical safety, privacy, psychological well-being, autonomy and freedom, quality of life, and intellectual property violation. For the organizational level, the variables encompass e-commerce, customer service, reputation, business models, organizational innovative capacity, and strategic decision-making. In the societal domain, the dependent variables include national security, political stability, economic stability, and social cohesion. At the environmental level, the variables are global warming, resource depletion, and waste generation. Each level of impact is treated as a separate continuous dependent variable, calculated by summing up several binary indicators that represent specific risks. **Independent variables** include data governance, content generation, and social and environmental issues. These independent variables represent the categories of risks that Generative AI incidents may pose and serve to analyze the relationship with the various levels of impact defined as dependent variables.

Chapter 6 highlights the contributions of the thesis along with the managerial and policy implications, limitations of the thesis, and future study.

1.7 Structure of the dissertation

This doctoral dissertation comprises one introduction and six chapters designed to explore various facets of RAI and its implications. Below is a brief overview of each chapter's structure and critical focus areas, also illustrated in Table 3.

The first chapter is the introduction, which sets the stage for the dissertation by introducing the concept of AI, its relevance to digital innovation, and the importance of adopting a responsible approach to AI development and deployment. The chapter outlines the research

context, epistemological positioning, and research methodologies employed throughout the study.

This second chapter embarks on a comprehensive exploration into the realm of RDI, emphasizing the consequences of managerial practices. It begins with a theoretical background section that includes definitions related to responsible innovation (RI), including responsibility, social innovation, innovation management, and sustainability-oriented innovation. The exploration also extends to the discourse on RI and responsible research innovation (RRI). Subsequently, the chapter delves into examining the adverse impacts of digital innovation from the environmental, economic, and societal perspectives. This analysis emphasizes the significance of incorporating responsible practices in innovation processes to mitigate these adverse effects. In addressing the methodology, the chapter delineates the research design, which is a qualitative study, objectives, and data collection process, incorporating both quantitative and qualitative studies. This methodological exposition provides a transparent foundation for the study's findings, ensuring a rigorous analytical approach. Finally, the results section presents the core findings of the research, focusing on the definitions of RDI, its dimensions, and practices and offering critical insights into the obstacles to fostering RI within digital contexts.

The third chapter outlines the approach used to examine ethical principles within AI ethics declarations, starting with the definitions and moral philosophies that underpin ethical AI and the concept of trustworthy AI. Following the clarifications of ethical AI, the chapter transitions into the methodology that elaborates on mixed methods, combining quantitative analysis and integrative literature review. The findings section thoroughly examines ethical principles in AI declaration, concluding with a comprehensive round-up of current ethical AI declarations and a frequency analysis of these documents. This is complemented by an integrative literature review, revealing opportunities and limitations. The chapter concludes with a discussion section offering insights into the implications of the findings for scholars, businesses, and policymakers.

The fourth chapter conducts an extensive literature review of RAI. The chapter begins by highlighting the need to implement robust AI governance mechanisms, examining the existing RAI framework and its limitations. This analysis sets the stage for proposing an integrative theoretical framework to overcome these limitations. The proposed details the framework through its multiple dimensions. First, the technical dimension focuses on

embedding ethical principles like privacy and transparency into AI systems. Second, the responsible innovation management dimension addresses managerial principles like anticipation and inclusion. Third is the legal dimension, exploring AI law adherence in the European Union and the United States. Fourth, the dimension of AI for Sustainability includes a definition of sustainable AI alongside exploring the opportunities and challenges AI presents for sustainability.

Chapter 5 delves into a critical examination of the risks associated with GAI, aiming to categorize them and assess their effects comprehensively. The chapter starts with a literature and framework section that lays the conceptual groundwork by providing an overview of the nature of risks posed by GAI technologies and the level of impact. Using data from the OECD AI Incidents Monitor, it examines the implications of GAI in data governance, content generation, and socio-environmental issues, contributing valuable insights to strategic management and innovation policy. The methodology section details the quantitative approach employed to investigate GAI risks, including the data source (OECD AI Incidents Monitor), analytical techniques (Python and SPSS), and the rationale behind the chosen methods. In the results section, the chapter presents the outcomes of the quantitative analysis, featuring a frequency analysis of identified risks and an associative analysis using crosstabulation, chi-square tests, and correlation to explore the relationships between different risk categories and their impacts. This includes an in-depth examination of issues related to data governance, content generation, and environmental and social implications of GAI.

Chapter 6 synthesizes the dissertation's insights and findings to draw a comprehensive conclusion. This chapter provides a summary of each chapter, highlighting the contributions of the thesis. It discusses managerial and policy implications, addresses the limitations of the research, and suggests directions for future studies.

Chapter	Chapter 1: Introduction	Chapter 2: A Framework to Implement Responsible Digital Innovation: Dimensions, Best Practices and Challenges	Chapter 3: Ethics in AI: Analyzing Ethical AI Declarations and Future Directions	Chapter 4: A Framework for Responsible AI	Chapter 5: Generative AI: A Quantitative Study on Emerging Risks and Impacts	Chapter 6: Synthesis, Limitations, and Conclusion
Content Type	General Introduction	Article	Article	Article	Article	Discussion & Conclusion
Research Question	What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI? What are the risks of GAI and their impact on individual society organizational and environment?	What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?	What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for RAI?	What are the essential dimensions of RAI and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?	What are the specific generative AI risks? How do these GenAI risks affect different aspects of society, business, and the environment?	What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?
Methodolog		Qualitative Study	Integrative Literature Review	Literature Review	Quantitative Study	
Publication Status		Accepted to Journal: Haidar, A. and Balagué, C. (2024). A Framework to Implement Responsible Digital Innovation: Dimensions, Best Practices and Challenges. <i>Journal of Innovation Economics & Management</i> (HCERES B; ABS 1; FNEGE 2). <i>Forthcoming</i>. Accepted for the Book: Balagué, C. and Haidar, A. (2024). L'innovation numérique responsable pour un monde plus durable. In Management durable. Coordinated by Jacquinot, P., Janand, A., & Saïd, K. (Eds.), Edition EMS. <i>Forthcoming</i>. Selected and Presented at the Conference: Haidar, A., & Balagué, C. (2023). A framework for responsible digital innovation. In <i>Digital Transformation Society Conference</i> (DTS 2023). 22-23 May, Paris, France. (Best Paper Award).	Submitted to Journal: Haidar, A. Ethics in AI: Analyzing Ethical AI Declarations and Future Directions, submitted to: <i>International Journal of Technology and Globalisation</i> (ABS 1) Selected and Presented at the Conference: Haidar, A. (2023). Ethical artificial intelligence: An exploratory study of guidelines. In <i>Artificial Intelligence: Law, Ethics, and Policy Conference</i>. 20 June, Toronto, Canada. Accepted for the Book: Haidar, A. (2024). Privacy-preserving Computing: Balancing Security and Privacy in the Digital Age. In A. U. Mirza & B. Kumar (Eds.), <i>Exploring the frontiers of artificial intelligence and machine learning technologies</i>. (pp. 198-210). San International Scientific Publication.. https://doi.org/10.59646/efaimlC11/133	Accepted to Journal: Haidar, A. (2024). An integrative theoretical framework for responsible artificial intelligence. <i>International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)</i> (ABS 1) , 13 (1), 1-23. http://dx.doi.org/10.4018/IJDSGBT.334844 Selected and Presented at the Conference: Haidar, A. (2023). A framework for responsible AI. In <i>R&D Management Conference</i> . 19-21 June, Seville, Spain	Submitted to Journal: Haidar, A and Balague, C. (2024). Generative AI: A Quantitative Study on Emerging Risks and Impacts. <i>Business & Society</i> (HCERES A; ABS 3; FNEGE 2) Selected and Presented at the Conferences: - Haidar, A., & Balagué, C. (2024). A Generative AI: A Quantitative Study on Emerging Risks and Impacts In <i>Digital Transformation Society Conference (DTS 2024)</i>.20-23 May, Naples, Italy. - Haidar, A., & Balagué, C. (2024). A Generative AI: A Quantitative Study on Emerging Risks and Impacts In <i>International Society for Professional Innovation Management (ISPIM)</i> Conference. 9-12 June, Tallinn, Estonia. - Haidar, A., & Balagué, C. (2024). A Generative AI: A Quantitative Study on Emerging Risks and Impacts In <i>R&D Management Conference</i>. 17-19 June Stockholm, Sweden	

Tableau 3 : Summary of the dissertation

Chapter 2: A Framework to Implement Responsible Digital Innovation: Dimensions, Best Practices and Challenges

Chapter 1: What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI ? What are the risks of GAI and their impact on individual society organizational and environment?



Chapter 2: What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?



Chapter 3: What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?



Chapter 4: What are the essential dimensions of Responsible AI (RAI) and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?



Chapter 5: What are the specific generative AI risks? How do these GAI risks affect different aspects of society, business, and the environment?



Chapter 6: What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?

About this chapter

This chapter has been accepted for publication in the **journal**:

Haidar, A. and Balagué, C. A. (2024). Framework to Implement Responsible Digital Innovation: Dimensions, Best Practices and Challenges. *Journal Innovation Economics & Management* (HCERES B; ABS 1; FNEGE 2). *Forthcoming*

Accepted for the **book**:

Balagué, C. and Haidar, A. (2024). L'innovation numérique responsable pour un monde plus durable. In *Management durable*. Coordinated by Jacquinet, P., Janand, A., & Saïd, K. (Eds.), Edition EMS. *Forthcoming*

Presented at the **conference**:

Haidar, A., & Balagué, C. (2023). A framework for responsible digital innovation. In Digital Transformation Society Conference (DTS 2023). 22-23 May, Paris, France. (Best Paper Award)

2.1 Introduction

2.1.1 Definition

2.1.1.1 The concept of responsibility

In science and technology studies, the notion of responsibility arose from post-war apprehensions over the potential negative impacts of technologies such as nuclear power (Lautenschläger, 1985). In business management literature, the concept of a business's responsibility towards society predates that of responsible innovation or science's responsibility towards society. The discourse surrounding the extension of corporate responsibility towards stakeholders and the broader society dates back to the 1950s and 1960s (Carroll & Shabana, 2010). Scholars like Bown (1953) and Drucker (1954) examined the moral and ethical responsibilities of a business and a business manager to the public good. As per Moir (2001), the idea of a business's responsibility towards society connects with three main theories: stakeholder theory, social contracts theory, and legitimacy theory. The stakeholder theory posits that firms and their managers should act in the interest of 'any group or individual who can

affect or is affected by a firm's operations in achieving its objectives' (Freeman 1984, p. 46). Social contract theory, as defined by Thomas Hobbes in the 1650's, posits that individuals mutually transfer their rights to a sovereign entity in exchange for protection and security, thereby establishing a fundamental basis for morality, justice, and societal norms and fostering equality, fairness, and a shared understanding of right and wrong (Sasan, 2021). Legitimacy theory, derived from organizational and social system theory, considers an organization as a complex component that must engage with its environment to survive (Guthrie et al., 2006) and posits that an organization's corporate social responsibility (CSR) performance can confer legitimacy upon it (Olateju et al., 2021). The European Commission (2011) has been encouraging CSR as a means to "create conditions favorable to sustainable growth, responsible business behavior and durable employment generation in the medium and long term" (p. 4).

There have been numerous conceptualizations of the responsibility of business towards society, including philanthropy, business ethics, CSR, corporate citizenship, and corporate sustainability (CS) (Carroll & Shabana, 2010). CSR and CS are currently the more common terms in business practice and show signs of convergence (Montiel, 2008). Yet, they have no fixed standard definition (Montiel & Delgado-Ceballos, 2014). Porter and Kramer (2011) claim that by acting ethically and directing a firm towards responding to society's demands, businesses can simultaneously fulfill their economic and societal responsibilities, and these authors introduced the concept of businesses "creating shared value." Their approach shifted business responsibility from the periphery to the center of a company's business strategy (Crane et al., 2014). Company responsibility was no longer viewed as an activity outside of a company's core operations and core strengths but rather as responsiveness to social requirements through creating products and services, which meant it became a possible growth strategy. Thus, from a business perspective, "responsibility" is frequently linked with the concept of "corporate responsibility," as suggested by Halme et al. (2009), who contend that this concept presumes that companies have "responsibilities." As Blowfield and Murray (2014) articulated, numerous definitions of corporate responsibility exist, but they observed a shared belief among them: "companies carry responsibilities for the public good- but they emphasize different elements for this" (p. 7).

2.1.1.2 Innovation

Schumpeter (1934) proposed the first and seminal definition of innovation, introducing the notion of so-called "new combinations." These combinations refer to the introduction of

new products and new production methods, the investigation of new markets, the conquering of new sources of supply, and the establishment of new business models. Although most of Schumpeter's works refer to innovation as novel combinations, they describe innovation as the "setting up of a new production function" (Schumpeter, 1939, p. 84). Yet, all of Schumpeter's definitions, whether related to new combinations or production functions, are vague and broad (Hagedoorn, 1996). The UK Department of Trade and Industry (DTI, 2003, p. 9) interprets innovation as "the successful exploitation of new ideas" because it accommodates the range of innovation types (product/service, process, administrative, technological, etc.) that one might reasonably expect to encounter in an organization.

Gault (2018) proposes a new definition of innovation as "the implementation of a new or significantly changed product or process" (p. 619), where the product can be considered as a good or service, while the process can include production, marketing, and organizational aspects. According to the given definition, if one analyses its separate elements, the following classifications emerge: innovations in production—developing or enhancing a specific product; innovations in services—offering new services or improving existing ones; innovations in the process—finding new ways of organizing and combining inputs in the production processes of specific products or services; and innovations in management—creating new ways of managing business resources. As part of a new discipline, Chen et al. (2018) introduced a new paradigm of innovation—holistic innovation (HI). HI is a comprehensive and collaborative innovation motivated by strategic planning. This new paradigm for innovation is a mix of strategic, collaborative, total, and open innovation, reflecting the Chinese setting and Eastern culture.

From an economic perspective, innovation "will increasingly be needed to drive growth and employment and improve living standards" (OECD, 2010, p.2). Business innovation objectives involve the expansion into new customer bases, the enhancement of existing ones, and the reduction of manufacturing costs. Businesses' associated strategies involve innovating in one of five areas: product (goods and services), process (manufacturing or delivery techniques), marketing (design, packaging, placement, promotion, price), or organization (preserving one's competitive edge or improving one's position in the market) (business practices, workplace organization, external relations) (Jarmai et al., 2020, p.10).

Three measures capture innovation, each pointing to a distinct component: patents, scientific and technical journal articles, and high-tech exports. Previous research has used patents to measure innovation (Acs & Audretsch, 1988; Katila, 2000). Following past research, the patent variable is the ratio of the total number of patents awarded in a country to its population. Previous studies have also used academic and technical journal articles to measure

innovation (Fagerberg et al., 2010; Zhang et al., 2020). The third indicator is a country's high-tech exports as a proportion of its total exports, as evidenced by previous literature (Sandu & Ciocanel, 2014; D'Angelo, 2012).

2.1.1.3 Overview of additional concepts linked to responsible digital innovation

In the exploration of Responsible Digital Innovation (RDI), understanding the foundational terms and concepts is essential. By encapsulating these core concepts, Table 4 provides a concise summary of key terms that are pivotal to this chapter and aid in comprehending the complexities and nuances of RDI.

Concept	Definition	Sources
Responsibility	Obligation for impacts of decisions and activities on society and the environment through ethical behavior.	Silva et al. (2019)
Business Responsibility	Moral and ethical responsibilities of a business and a business manager to the public good.	Bowern (1953), Drucker (1954)
Corporate Social Responsibility (CSR)	CSR is the voluntary effort by businesses to go beyond legal requirements to positively impact society and the environment. It involves balancing economic, social, and environmental concerns to meet stakeholder expectations.	Dahlsrud (2008)
Corporate Responsibility	Responsibility for the impacts of decisions and activities on society and the environment through ethical behavior.	ISO 26000
Innovation	"New combinations," including new products, production methods, markets, sources of supply, and business models.	Schumpeter (1934, 1939)
Responsible Innovation	Responsible innovation is a collective commitment to care for the future through responsive stewardship of science and innovation in the present.	Owen et al. (2013)
Responsible Research and Innovation	a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability, and societal desirability of the innovation process and its marketable products.	Von Schomberg (2013)
Social innovation	innovative activities and services that are motivated by the goal of meeting a social need and that are predominantly diffused through organizations whose primary purposes are social.	Mulgan et al. (2007)

Sustainability oriented Innovation	Sustainability-oriented innovation is the commercial introduction of a new or improved product, service, or system that results in “environmental and (or) social advantages over the physical life-cycle of the prior version.”	(Hansen and Grosse-Dunker, 2012)

Tableau 4 : Summary of Key Terms in Responsible Innovation (source: Author)

2.1.2 The discourse on responsible innovation and responsible research innovation

“Responsible innovation” (henceforth RI) and “responsible research and innovation” (henceforth RRI) are two linked discourses that have emerged in parallel over the past decade and that challenge the epistemological norms and practices regarding the production and utilization of scientific knowledge. RRI shares many characteristics with the discourse on RI, which has, in contrast, emerged mainly from academic roots and is vital in maintaining mainstream action lines: gender equality in research, open access to scientific results, ethics, public engagement, and the promotion of science education (Owen & Pansera, 2019).

2.1.2.1 Responsible innovation

RI is a new idea introduced and developed top-down by policymakers and scientists (Zwart et al., 2014). Nonetheless, the concept is open to multiple interpretations and competing narratives. According to Grinbaum and Groves (2013), RI means taking responsibility in quasi-parental and collectively political ways. The quasi-parental way of taking responsibility implies a limited individual responsibility that focuses on the duty of care for the malleability of technological artifacts and the vulnerability of their future use. Stilgoe et al. (2013) state that “responsible innovation means taking care of the future through collective stewardship of science and innovation in the present” (p. 1570).

RI is primarily concerned with democratizing innovation (Owen et al., 2013; Macnaghten et al., 2014) and implementing deliberative forms of governance, such as stakeholder and public involvement (Stilgoe et al., 2013). Owen et al. (2012) and Stilgoe et al. (2013) focused on developing a democratic governance framework for innovation based on reflecting on the innovation’s purpose(s) and on preventing negative consequences. Stakeholders and public members are involved upstream in the innovation process and encouraged to deliberate about the multiple futures and uncertainties that the innovation could bring or seek to bring. Through deliberative forms of governance, this upstream inclusion of

stakeholders and the public can help to realize a collective responsibility to control and drive innovation in a morally acceptable, societally desired, and sustainable direction (Von Schomberg, 2012). This is expected to increase the chance of innovation adoption, better integrate such innovation into society, and ensure that it provides societal benefits (Ribeiro et al., 2017). Rocco et al. (2013) identified four aspects of responsible innovation: (1) transformational across sectors and disciplines; (2) taking into account equitable access, health, safety, and environmental concerns; (3) including government institutions and other stakeholders; and (4) anticipating and adapting over the long term.

The problem with some RI concepts is that they are set by researchers focusing primarily on technological development (Lettice et al., 2013) without differentiating among research, development, and commercialization. Such an approach creates critical challenges for implementing RI in a business context. First, focusing on science and technological development indicates a narrow view of innovation, as other types of innovation are not considered, such as social innovations (Blok & Lemmens, 2015). Second, commercialization is an essential stage within the innovation process (Baregheh et al., 2009). The shift from one socio-technical system to another requires models of innovation and systems approaches to innovation for analysis accompanied by “responsibility” and “sustainability.” Systems that come to innovation include various perspectives and theories, such as multi-level perspectives, transition theory, and innovation systems (Singh et al., 2021).

2.1.2.2 Responsible research and innovation

RI and RRI have a history that stretches back a decade (Hellström, 2003; Owen et al., 2009; Owen & Goldberg, 2010), along with other related terms, such as “responsible development” (National Nanotechnology Initiative, 2007).

RRI was first employed in the Seventh Framework Programme (“Regulation (EU) No 1291/2013”, 2013). Burget et al. (2017) conducted a literature review on responsible innovation and found that policymakers and scientists conceptualized and characterized responsible innovation differently. After analyzing and synthesizing the literature, they concluded as follows:

“Responsible innovation is essentially an attempt to govern research and innovation in order to include all the stakeholders and the public in the early stages of research and development. The inclusion of different actors and the public is, in turn, meant to increase the possibilities to anticipate and discern how research and innovation can or

may benefit society as well as prevent any negative consequences from happening” (p.15).

In recent years, these discourses have progressed along increasingly divergent paths, and, indeed, these paths seem likely to diverge even further in the future as the EC moves towards a “3O” policy (open science, open innovation, open to the world)⁴. Zwart et al. (2014) noted that the science policymakers coined the term RRI, not the research community. Thus, we must explore other administrative and academic definitions.

Hilary Sutcliffe (2011) defines research and innovation based on an administrative approach. Research refers to a systematic investigation to establish facts and reach new conclusions. Innovation is “a superior process or product” and the “effective commercialization of an invention” (p.3). Von Schomberg (2011) focuses on democratic innovation governance and defines RRI as follows:

... a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability, and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society) (p. 9).

The literature review revealed that researchers commonly used this definition. However, Davis and Laas (2014) criticized it at the time, highlighting five “problems”: the absence of “knowledge,” the fact that it only concentrates on “mere technical inventions,” the claim that an “innovative process” should end in “marketable products,” making societal desirability an independent category, and the vagueness of the final phrase “in our society.” Stakeholders should regard RRI as a strategy to become mutually responsive and anticipate research and innovation outcomes underlying the “grand challenges” of our era, for which they share responsibility (Von Schomberg, 2013).

Based on an academic approach, most elaborations have extended von Schomberg’s concept by emphasizing the foresight of probable repercussions and social expectations and the engagement of stakeholders and the general public in the research and innovation (R&I) process (Bremer et al., 2015; Forsberg et al., 2015). Nevertheless, other scholars have provided their definitions of RRI. Stahl (2013) defines RRI as a meta-responsibility in the following manner: “RRI is a higher-level responsibility or meta-responsibility that aims to shape, maintain, develop, coordinate and align existing and novel research and innovation-related processes,

⁴ <https://digital-strategy.ec.europa.eu/en/library/open-innovation-open-science-open-world>

actors and responsibilities to ensure desirable and acceptable research outcomes” (p.712). Some authors have gone into greater detail regarding the definition, explaining it from their perspectives. Spruit et al. (2016) characterize RRI “by a shift from assessing the desirability of the outcome of innovation processes, such as evaluating harmful product outcomes in court under liability law, to assessing the qualities of the innovation process” (p. 872). In this sense, RRI appears to be a relatively narrow concept: it is defined solely by the innovation process required to achieve the intended outcome. Wilford (2015) stresses personal responsibility in the R&I process. She emphasized that “RRI creates a step-change in the way that those who are engaged in research and innovation should consider the impact of what they do” (p.348).

2.1.2.3 Overview of distinctions between RI and RRI frameworks

Table 5 provides a detailed comparison of RI and RRI, highlighting the foundational differences, governance frameworks, and key objectives that distinguish these two approaches.

Feature	Responsible Innovation (RI)	Responsible Research and Innovation (RRI)
Definition	RI involves taking responsibility in quasi-parental and collectively political ways, focusing on duty of care for technological artifacts and their future use. It emphasizes democratizing innovation and deliberative governance (Grinbaum & Groves, 2013; Stilgoe et al., 2013).	RRI aims to govern research and innovation to include all stakeholders and the public in early stages to increase chances to anticipate and prevent negative outcomes. It's defined as a transparent, interactive process ensuring that societal actors and innovators are mutually responsive, aiming for ethical acceptability, sustainability, and societal desirability of innovations (Von Schomberg, 2011).
Historical context	Introduced top-down by policymakers and scientists, RI is subject to multiple interpretations and competing narratives, with a focus on present challenges and the integration of innovation into society sustainably (Zwart et al., 2014).	RRI has evolved from broader concepts such as "responsible development" and was first formally recognized in the Seventh Framework Programme. It is particularly noted for diverging paths in discourse, especially under the European Commission's "3O" policy: open science, open innovation, and open to the world (National Nanotechnology Initiative, 2007).
Governance and framework	Focuses on developing a democratic governance framework for innovation, reflecting on the innovation's purposes and preventing negative consequences. Involves stakeholders and public members early in the innovation process (Owen et al., 2012; Stilgoe et al., 2013).	Emphasizes a systematic inclusion of various actors from the initial stages, fostering a governance model that is responsive to social expectations and ethical standards. It addresses the foresight of probable repercussions and engages stakeholders and the public in the R&I process (Von Schomberg, 2011; Forsberg et al., 2015).
Critical views and challenges	Criticized for being too focused on technological development without differentiating among research, development, and commercialization. Seen as potentially narrow in viewing innovation mainly through scientific and technological advancements (Lettice et al., 2013).	Criticized for the absence of "knowledge" focus, concentrating only on technical inventions, and the expectation that the innovation process should end in marketable products. There are also concerns about making societal desirability an independent category and the vagueness of its impact on society (Davis and Laas, 2014).
Main objectives	Primarily concerned with controlling and driving innovation in a morally acceptable, societally desired, and sustainable direction. Aims to increase the chance of innovation adoption, better integrate innovation into society, and ensure societal benefits (Von Schomberg, 2012; Ribeiro et al., 2017).	Aims to create a step-change in how research and innovation consider the impact of their activities, focusing not only on the outcome but also on the quality of the innovation process itself. It is defined by personal responsibility and the meta-responsibility to coordinate and align research and innovation processes to ensure desirable outcomes (Stahl, 2013; Wilford, 2015).

Tableau 5 : Comparative Analysis of Responsible Innovation (RI) and Responsible Research and Innovation (RRI)

2.1.3 Study's objectives and research question

This digital evolution is marked by tremendous opportunities and complex challenges, requiring responsible implementation of digital innovation. To address these challenges, Holmström (2018) argues for developing explanations of digital innovation based on (1) specifics of digital technology, (2) acknowledging the complexity of socio-material interaction, and (3) an oscillation between the specific and the general. As a part of the doctoral program theme, we are interested in studying the impact of digital innovation in the business context. Thus, this study highlights three objectives:

- Build solid academic literature before elaborating on questions for the interviews. This requires a significant understanding of the dark side of digital innovation on the three dimensions of sustainability –environment, society, and economy – this objective stresses the importance of implementing digital innovation responsibly.
- Introduce a novel definition of responsible digital innovation from a business point of view, revealing that digital innovation does not only encounter ethical or environmental concerns but also managerial challenges.
- What dimensions and practices of RDI will help companies develop their strategies for responsible governance?
- What are the obstacles to implementing RDI practices?

2.2 Dark side of digital innovation

2.2.1 Environmental perspective

In 2000, Paul Crutzen, a Nobel prize winner, proposed the beginning of the Anthropocene, a period of time during which humankind will undertake a growing “responsibility for the welfare and future evolution of life on the Planet” (Abelkop & Carlson, 2012, p. 102). He used the word to refer to a new geological epoch characterized by humankind’s growing influence on the Earth. The most critical challenges of the twenty-first century are climate change and global warming, characterized by an increase in average worldwide temperatures (Richards et al., 2021); that temperature change should be less than 1.5 degrees Celsius what Attenborough states in COP26 climate change conference (Ambrose, 2021). Increased carbon dioxide (CO₂) emissions due to the anthropogenic usage of fossil fuels may cause global warming (Zhao & Zhang, 2018). In some regions of the world, significant population shares have little or no trust

in what scientists warn us about the environment. According to a recent survey by the World Economic Forum in 2021, 16 percent of North Americans admit to trusting issues regarding climate science, as shown in Figure 14 (Lacina, 2021). These trust issues highlight how DI, while offering significant advantages in disseminating information, also raises concerns about the credibility and trustworthiness of digital content. Also, this lack of trust can lead to skepticism towards scientific findings, impacting the public and the company's support for necessary environmental policies and actions. Thus, it emphasizes the need for responsible innovation that gains the acceptance of the public and firms.

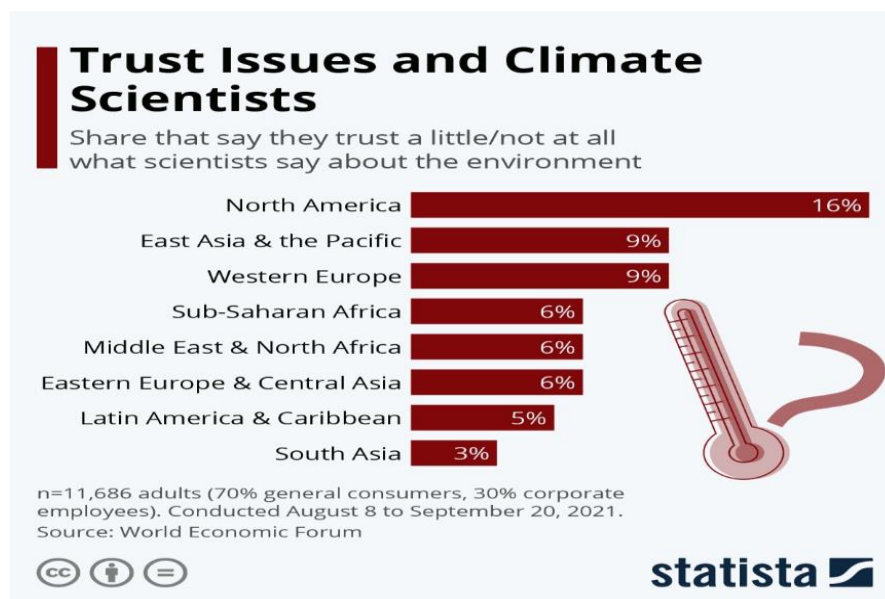


Figure 14 : trust issues and climate scientists (source: WEF, 2021)

The advancements in digitalization are astonishing. Data are growing exponentially — traffic has increased, and previous years have created almost 90% of all data today. Due to this exponential expansion, increasingly significant units of measurement are in use. For instance, global annual internet traffic exceeded the exabyte mark in 2001, and analysts predict it will surpass the zettabyte mark by 2017, as Figure 15 shows. People and devices are also becoming connected in ever-increasing numbers. As shown in Figure 16, more than 6.5 billion people, or nearly half the global population, now use the internet – up from only 500 million in 2001.

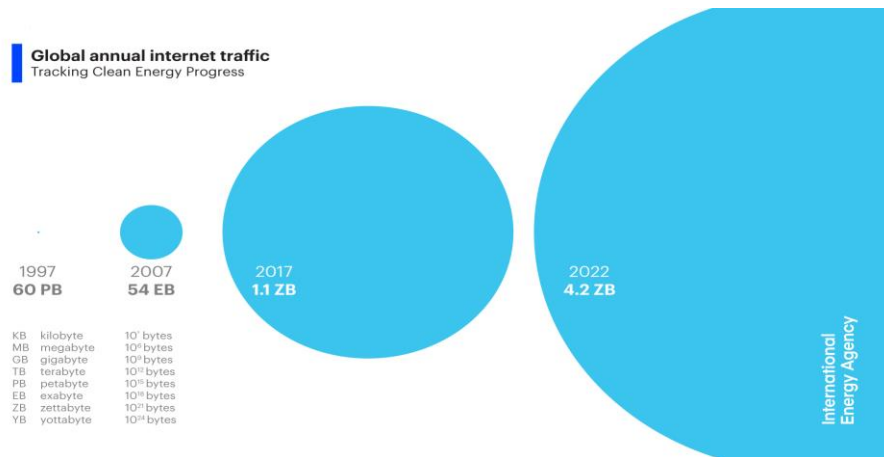


Figure 15 : Global Annual Internet Traffic (source: [iea.org](https://www.iea.org/))

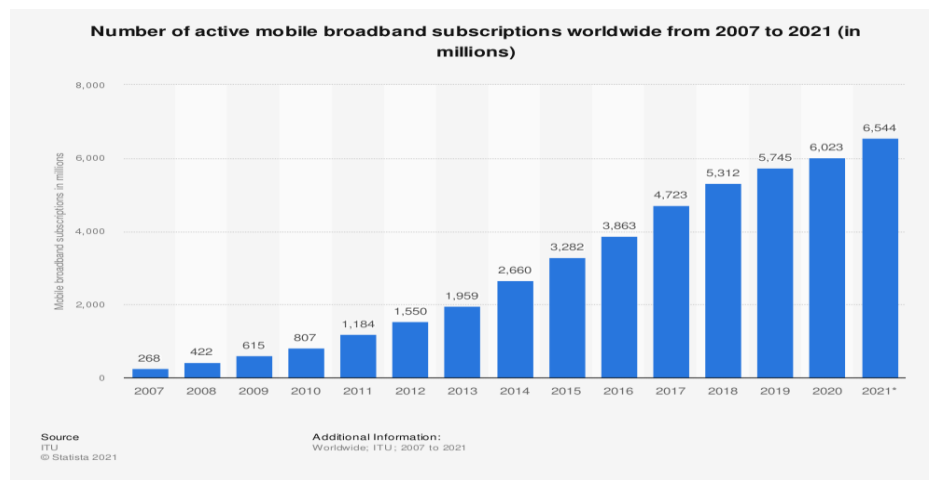


Figure 16 : Number of active mobile broadband subscriptions worldwide from 2007 to 2021 (source: [statista.com](https://www.statista.com/))

The exceptional urbanization and industrialization rate have accelerated fossil fuel usage to meet global energy demands (Voumik & Sultana, 2022). The environmental impact of digital technology is now recognized as unsustainable but continues to grow. According to Toledo (2020), the average household Internet data usage has jumped 39 times in 10 years. Media headlines such as the New York Post⁵, CBC⁶, and Big Think⁷ claim that CO₂ emissions from 30 minutes of watching Netflix are the same as driving almost four miles (Kamiya, 2020). Duarte et al. (2013) empirically analyzed 11 countries and observed that technological actors increase CO₂ emissions in the United Kingdom and the United States. Data centers serve as the backbone of information in an increasingly digital society that demands intensive energy use

⁵ <https://nypost.com/2019/10/28/why-climate-change-activists-are-coming-for-your-binge-watch/>

⁶ <https://www.cbc.ca/news/science/data-centres-energy-consumption-1.5391269>

⁷ <https://bigthink.com/the-present/netflix-binge-watching-makes-climate-change-worse/>

(Masanet et al., 2020). As shown in Figure 17, traditional data centers have decreased their energy demand globally, from around 97.6 terawatt-hours in 2015 to approximately 33 terawatt-hours in 2021. Note that hyperscale data centers have doubled their energy demand in the same period. Hyperscale data centers are immense data centers built by ICT companies with a primary business line focused on data and cloud-based services (Horner & Azevedo, 2016). Blockchain technology relies on a consensus algorithm, known as Proof-of-Work (PoW), that requires enormous energy (Alofi et al., 2022). Bitcoin is a leading example of this technology due to the several currencies linked to it, increasing electricity consumption (Gallersdörfer et al., 2020), which leads to environmental damage (Stoll et al., 2019). According to the Cambridge Bitcoin Electricity Consumption Index⁸, the monthly Bitcoin electricity consumption increased from 1.12 terawatt-hours (TWh) in January 2017 to 10.26 in February 2022.

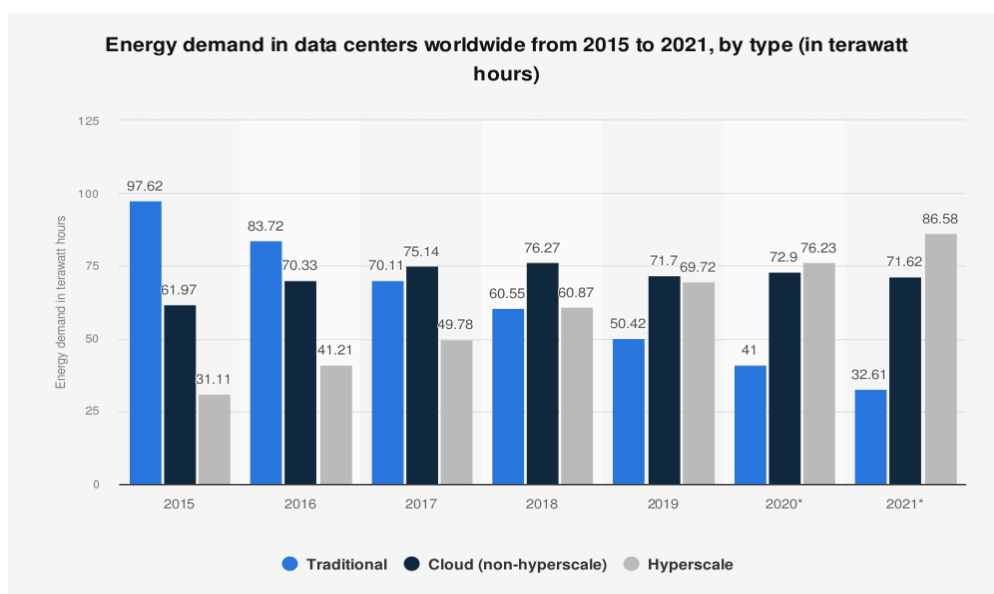


Figure 17 : Energy demand in data centers worldwide from 2015 to 2021 (source: IEA 2022 Ministerial Meeting)

We are currently seeing rapid environmental change, including melting Arctic and Antarctic ice, thawing of permafrost, destruction of coral reefs, expansion of deserts, increased frequency of droughts, floods, and forest fires, and the extinction of insect and vertebrate populations (Slater et al., 2021). Intelligent technologies like AI are part of the Information and

⁸ <https://ccaf.io/cbnsi/cbeci>

Communication Technology (ICT) sector. Following the taxonomies from (Hilty & Hercheui, 2010; Horner et al., 2016), its impact can be divided into: First-order, the impacts due to I.T. infrastructures. For example, an AI solution in a smart building may need sensors and servers that require production, operation, and end-of-life resources and energy. Second-order impacts correspond to the impacts due to the application of AI, i.e., AI can optimize or substitute existing systems: energy consumption in a building can be optimized using occupancy or behavior detection, energy profiling, etc. Third-order impacts are all changes in technology or society due to the introduction of AI applications.

Metals are essential raw materials for fundamental and advanced technologies (Wellmer et al., 2019). At the same time, metal mining and processing contribute to adverse environmental impacts (Spitz & Trudinger, 2019). Langkau & Tercero Espinoza (2018) found a significant increase in the metals (Cooper, Cu, etc.) used in emerging technologies that negatively impact the environment. This increased use of metal leads to an e-waste problem, which represents significant human health and environmental risks due to the toxicity of materials (Luksemburg et al., 2007; Huo et al., 2007; Wong et al., 2007; Fu et al., 2008).

E-Waste Problem

The rapid expansion of technological innovation brings a significant environmental challenge: electronic waste (e-waste). This issue is particularly acute with the rise of off-grid solar e-waste (SEW) in the global South, highlighting a critical aspect of innovation's potentially dark side. As Hansen and Moe (2021) articulate, while developing new technologies, like those aimed at improving energy access, offers solutions to specific problems, it concurrently spawns new challenges.

A telling indicator of this growing problem is the projection that the number of Internet of Things (IoT)-connected devices will double from 12 billion in 2015 to 25 billion by 2025, as GSMA (2021) reports. Additionally, the demand for data centers and network services is set to surge, driven by video streaming, online gaming, and emerging technologies such as machine learning, blockchain, and 5G (IEA, 2021). Concurrently, the average lifespan of consumer electronics, exemplified by smartphones, ranges from two to four years (Qiang et al., 2020), contributing significantly to e-waste volumes. In France alone, this generates approximately 20 kilos of electronic waste per inhabitant each year (Taffin, 2021). If not correctly handled, the e-waste would lead to severe waste of natural resources, human and animal health hazards, and harm to the ecological environment (Saha et al., 2021).

Air pollution problematic

The industrial revolution has brought about new technologies that have damaged our world in two main ways: air and water pollution and depletion of natural resources (Sadhu et al., 2018). Technological change consumes chemical inputs like pesticides (Lohmann et al. 2007). This process is becoming a public health issue in most advanced economies (Luzzati et al., 2018). The internet is the largest coal-fired machine on the planet, increasing energy use because we are always online (Lazona, 2019). Hang and Yuan Sheng (2011) found that the effect of technological development on carbon emission is positive in the first stage and harmful in the later stages in China. Irandoust (2016) searched the relationship between technological innovation, renewable energy, and carbon emissions in Denmark, Finland, Norway, and Sweden from 1975 to 2012. There is a positive relationship between ICT and carbon dioxide emissions (Weili et al., 2022). Khattak et al. (2020) analyzed the impact of technological innovation, economic growth, and renewable energy use on carbon emissions in the BRICS countries from 1980 to 2016. Findings show that innovation activities have failed to reduce carbon emissions for BRICS countries except Brazil.

2.2.2 Economy perspective

Innovation can negatively impact the economy if policymakers do not calibrate innovation policy precisely (Coad et al., 2022). Dilmus James (1987) stated that institutionalists believe that market demand and factor prices are not the only determinants of the rate and direction of technological progress. Instead, it is also affected by power relationships and the vested interests of many actors. This powerful interplay between institutions and technology (Nelson 2005) substantially impacts social progress. Recent literature has highlighted the role of the increasing innovation costs in the U.S. productivity slowdown since the 1970s (Gordon, 2013) and the decline in R&D productivity in the pre-Great Recession slowdown (Anzoategui et al., 2019).

Research on the increase in income inequality since the early 1980's has become prominent in economics. Exposed explanations include reasons such as education (Prettner and Schaefer, 2021) or declining unionization (Farber et al., 2021), but also other motives such as automation (Acemoglu and Restrepo, 2020; Prettner and Strulik, 2020), innovation policies (Chu & Cozzi, 2018), factor-eliminating technical change (Peretto and Seater, 2013), factor-eliminating technological change (Peretto and Seater, 2013), and the dominance of superstar firms and

rising markups (Autor et al., 2020). Innovation activities strongly correlate with income inequality, even at the country level. A recent United States study shows that the shift in R&D investment from the public to the private sector was associated with a rise in the income share of the wealthier classes at the expense of the poorer income classes (Josifidis & Supic, 2020). An empirical study for China shows that there is a U-shaped relationship between the innovation level and the ratio between urban and rural income, which means that while small amounts of innovation can reduce income inequality and contribute to income equality, large amounts of innovation may increase income inequality (Liu and Lawell, 2015).

One form of inequality resulting from innovation is the excessive patent protection or monopolistic power from prior breakthroughs (e.g., from first-mover advantages, secrecy, or network effects), which can harm economic vitality and consumer welfare and exacerbate inequality (Coad et al., 2020). For example, when the private sector manages to keep others out of essential parts of the pool of knowledge or makes it hard to figure out what can be patented, it may affect the rate and direction of innovation. This seems to be the case in some areas of biomedical research, where fragmented and overlapping property rights have favored research lines that are less troublesome in terms of intellectual property over ones that could have an enormous influence (Heller and Eisenberg 1998). Madsen et al. (2022) investigate the role of research productivity in explaining the evolution of income inequality. Their estimates reveal a robust and positive association between research productivity and income inequality, with the asset returns channel being the dominant transmission mechanism.

Various researchers argue that innovation's main impact is unemployment (Biggi & Giuliani, 2022). Employees will suffer adverse repercussions without a strategy for execution and adequate study on introducing new assets with process innovation. According to the model of Weiss and Garloff (2011), technological innovation boosts skilled but not unskilled workers. As a result of the growth in productivity, earnings and average incomes for skilled workers rise. In turn, increased social benefits lead unskilled workers to seek higher salaries. As their productivity has not improved, this group's unemployment rate has grown. From a micro viewpoint, McKinsey's (2017) report argues that automation will enable small players and even individuals to undertake project work that is now performed mainly by more prominent companies. Autor and Salomons (2017) found that "one more robot per thousand workers reduces the employment-to-population ratio by about 0.18–0.34 percentage points, and wages by 0.25%–0.5%." Furthermore, Bordot's (2022) analysis of OECD nations between 2005 and 2017 revealed that a 10% rise in the usage of industrial robots correlates with a 0.42-point uptick in unemployment rates. This effect is notably more pronounced, by a factor of 2.5, among

individuals aged 25-34 who have not completed upper secondary education, as opposed to those aged 55-64 holding a tertiary degree.

Although prior research has attempted to evaluate the influence of innovation on employment growth at the firm level in developed economies, it is crucial to investigate the link between innovation and employment at the firm level in developing country contexts with high unemployment rates (Vivarelli, 2014). Sithole and Buchana's (2021) process innovations have a negative (decreasing) effect on employment growth for developing countries' manufacturing and service sectors. The degree of innovation varies among different groups of countries in the world. According to the employment approach, Yildirim and his colleagues (2020) study discussed 23 E.U. member states from 1998–2015. The findings confirm innovation in high and low regimes has effects that will increase unemployment. Yet, from the viewpoint of employees, there is a sense of redundancy when collaborating with robots (Seyitoğlu et al., 2021). Additionally, within a business setting, Chatterjee and his colleague (2022) discovered that the swift adoption of novel technologies, a phenomenon described as technological turbulence, adversely impacts how automated decision-making is integrated into the operational effectiveness of business-to-business interactions, subsequently influencing their satisfaction levels.

2.2.3 Societal perspective

Technology plays an increasing influence on society's development. Whether it has already manifested or not, it has brought us a great deal of benefits as well as threats (Lehmann, 2005). The term "information and communication technology" (ICT) refers to a wide range of technological tools and resources used to transmit, store, produce, share, or exchange information. Despite several advantages of ICT, examining its adverse influence is crucial. First, many prefer online communication over face-to-face conversations, so people tend to become more individualistic and introverted (Hu et al., 2014; Civan et al., 2017). Another limitation is that people cannot express their feelings or what they want to say by e-mail or social networking sites (Nyland, 2007). Second, ICT motivates people to become socially disconnected, where people are always busy with their virtual lives, forgetting the real world with family and friends and becoming a formality (Helsper, 2021). Children, too, are spending more time in the virtual world and adopting wrong thoughts, thus raising cyber crimes and extramarital affairs with ICT gadgets. The third important dimension is that the community is dependable on technology, leading people to reduce physical activities, which embrace health

problems such as high blood pressure, elevated blood glucose, and overweight are initiated in the human body (Kraut et al., 1998). Substantial costs arise, potentially causing poorer students and educational institutions to suffer. Researchers often cite this as a factor contributing to the digital divide (Lythreath et al., 2022). Chen and Fan (2003) have conducted a study on the negative impact of emerging technology deployed in clinics, concentrating on the negative influence on the doctor-patient relationship, the initiative of medical personnel, and the availability of medical resources. He is concerned that applying developing technologies to medical and clinical practice may result in wasteful spending and iatrogenic illness.

We live in the age of the network society; it is a society with no barriers where the world has become a global village. Social media networks eliminated the traditional limits on communication and changed how we see society. It's easy to get lost in a sea of websites, social media posts, and videos due to the abundance of information available. This online overload might lead to cognitive disorders and other mental health issues for some individuals. According to Stefanouick's (2022) analysis, the five most detrimental consequences of technology on our lives are as follows: psychological issues (50%), sleep problems (48%), social issues (45%), difficulty forming relationships in the real-world (42%), and difficulty with focus (41%). Other scholars have focused on particular, more specific aspects. For instance, when children encounter inappropriate content, like violence or sexually explicit material, through multimedia, it can make understanding unfamiliar words in text or visuals challenging for them (Shilpa and Sunita, 2016). The following are some examples of some of the mental and cognitive disorders caused by modern technologies:

- Digital dementia: A condition in which a person's memory and calculating skills degrade due to excessive reliance on digital technologies (Pappadà et al., 2021). In other words, while brain functions necessary for search improve, essential brain functions eventually decline.
- Cyberchondria: The amount of cyberchondria-related research has continuously increased (Vladan et al., 2020). Cyberchondria is associated with health anxiety, problematic Internet use, and symptoms of obsessive-compulsive disorder, with public health implications about functional impairment and altered healthcare utilization (Starcevic et al., 2020).
- Internet addiction disorder: Defined as an uncontrollable desire to use the internet, resulting in severe nervousness and aggression in the event of deprivation and progressive deterioration of social life (Masih and Rajkumar, 2019). Xin et al. (2018) investigated the prevalence of Internet Addiction in Guangzhou, China, recruiting 6468

adolescents aged 10–18 from local schools. The overall prevalence of Internet Addiction was 26.50%, with severe addiction being 0.96%. Internet Addiction was higher among males than females (30.6% versus 21.2%). In another study focusing on Africa, researchers identified 5,562 studies; they included 28 studies from 10 countries involving 14,946 high school and university students in this review. They found that the overall pooled prevalence of internet addiction among these students was 34.53% (Zewde et al., 2022).

- Low self-esteem: A study published in JAMA Pediatrics found that adolescents who spend more than three hours per day on social media are expected to have poor mental health, such as low self-esteem (Riehm et al., 2019)

The increasing reliance on technology introduces challenges closely tied to biases and discriminations embedded within their development's cultural and historical contexts. According to Hercheui and Guerard (2022), these biases are not just peripheral issues but are deeply integrated into the technology during its design phase, reflecting the prejudices and stereotypes of its creators. The rapid spread of misinformation through digital platforms further complicates this situation. Fedeli (2020) highlights how the deliberate or accidental propagation of false information can misguide entire communities and precipitate significant societal disturbances, underlining the urgent need for effective mechanisms to mitigate these effects.

In cybersecurity, adopting cloud-computing technologies, while offering numerous advantages, also presents novel challenges for providers and users, especially regarding data security and permanence. Blum and Beyer (2019) discuss the concept of digital permanence, which concerns the strategies employed to ensure that stored data remains accessible and relevant over time, meeting current and future needs. This aspect of digital data management demands careful consideration to guarantee the longevity and integrity of information stored in the cloud. Alouffi et al. (2021) and Sasubilli & Venkateswarlu (2021) further emphasize the importance of addressing these emerging risks as a critical step in the decision-making process for organizations contemplating adopting cloud solutions. Addressing these concerns requires a concerted effort to develop and implement strategies that safeguard against these inherent risks, ensuring that technological advancements contribute positively to society without compromising security or perpetuating existing biases.

2.3 Methodology

2.3.1 Research design

The function of a research design is to integrate the different components of the study coherently and logically to ensure that the evidence obtained enables the researcher to effectively address the research problem as unambiguously as possible (Kirshenblatt-Gimblett, 2006). This study aims to gain in-depth knowledge of a specific context, explore under-researched issues, and generate new ideas. Swedberg (2020, p. 17) defines exploratory research as “the soul of good research” and employs a qualitative approach. In light of this, this study uses **qualitative** methodology. This study was conducted between April and June 2021 by two researchers (Ahmad Haidar, Ph.D. student; Christine Balagué, Professor) in collaboration with the Content Strategist of Cap Digital. Cap Digital is a French digital and ecological transformation hub comprising one thousand companies. It has been certified since 2014 as a “Cluster of Excellence” by the European Commission.

2.3.2 Data collection process

We executed participant selection in two stages:

1. **Initial contact:** A manager from Cap Digital contacted their member database by email, extending an invitation to digital innovation managers to participate in a working group focused on RDI. Out of the one thousand members contacted, 143 managers expressed interest and agreed to engage in a community exchange on best practices for RDI.
2. **Interview recruitment:** From this working group, 18 managers agreed to participate in individual qualitative interviews focusing on their experiences and practices related to RDI. The participants represented a diverse range of enterprises (9 SMEs and 9 large organizations) from various economic sectors such as technology, retail, recruitment, industry, services, public sector, and media. Table 6 illustrates the detailed composition of the sample.

We conducted 18 semi-structured interviews, each lasting between 30 and 60 minutes, using a guide prepared in advance (detailed in Annex 1) to ensure consistency across discussions. The interviews covered four main topics: definitions of RDI, dimensions of RDI, experiences with RDI practices, and barriers to RDI implementation.

Company name	Company Size	Economic sector
1	SME	Technology
2	SME	Services
3	SME	Public Sector
4	SME	Recruitment
5	Large	Retail
6	Large	Retail
7	Large	Industrial/Manufacturing
8	Large	Services
9	SME	Technology
10	SME	Public Sector
11	Large	Media
12	SME	Technology
13	SME	Services
14	Large	Public Sector
15	Large	Services
16	Large	Industrial/Manufacturing
17	Large	Technology
18	SME	Technology

Tableau 6 : Sample of the study

2.3.3 Data analysis

All interviews were conducted virtually via the Zoom video conferencing platform in response to the restrictions imposed by the coronavirus pandemic. With the consent of the participants, we recorded each session and later transcribed it. We employed content analysis to examine the transcriptions, which Harwood and Garry (2003) describe as a technique for analyzing various data, including visual and verbal materials. This method facilitates the reduction of phenomena into defined categories to better analyze and interpret them, enabling a nuanced exploration of RDI practices.

To perform this analysis, we utilized qualitative data analysis (QDA) Miner 6.0.10 software to analyze the interviews. Deterding and Waters (2018) showed that QDA software packages have evolved into more powerful and versatile tools, mirroring the analysis objectives of researchers, and one can summarize the best practices for this software in three steps. First, index transcripts with broad codes reflecting the questions and concepts we aim to examine. Second, write analytic memos during the process to generate conceptual themes appearing across cases. Third, researchers can use existing indexes for data reduction and apply more detailed codes to subsets of the interviews.

2.3.4 Ethical considerations

Each interview began with an explicit request to the interviewees for their consent to record the conversation to ensure ethical standards, particularly confidentiality and informed consent. Since we conducted the interviews via Zoom, this step enabled accurate transcription and content analysis. This approach adhered to ethical research practices, ensuring participants were fully informed about the recording's purpose and data usage, thus respecting their autonomy and rights throughout the research process.

2.4 Results

2.4.1 Definitions of responsible digital innovation

This section categorized the participants' definitions of RDI based on their scope. Each illustration draws from affirmative responses to specific claims, denoted by quotation marks. Table 7 explores the answers from participants as themes of the definitions.

Key-Themes	Objective	Verbatim
<i>Having a positive environmental and societal impact</i>	One of the significant conditions to call the digital innovation implementation “responsible” is to positively affect the environment and society.	“putting digital technology at the service of society and developing an action plan for a positive impact of digital innovation on the 17 Sustainable Development Goals (SDGs).”
		“to develop an ethical, inclusive, and sustainable digital model for the planet.”
		“to be aware of the positive impact of digital innovation on society and the environment.”
<i>Be responsible for the ecological footprint of digital innovation</i>	The designers of digital innovation are accountable for the impact of the environmental footprint of these technologies.	“the ecological footprint for the storage and processing of massive data”
		“the use of digital technologies to improve ecological efficiency.”
<i>Building a digital trust</i>	In addition to the environmental and societal dimensions, <i>digital trust</i> is a crucial characteristic of responsible digital innovation. This will make the relationship between digital innovation and	“building a digital ecosystem of trust at the service of businesses and citizens.”
		“Developing a certificate of trust
		“The best response to combat information overload is to increase trust ... through collaboration between algorithms and humans.”
		“Gaining trust takes a long time. So, suppose we already want to even maintain the level of trust that users may have vis-

	consumers/citizens bold and significant.	à-vis the community in terms of digital tools and, more particularly, data management. In that case, we must consult first before initiating any action.” “Maintain the level of trust we can have in our citizens and perhaps increase it by a few points.” “For me, talking about ethics, above all, the question of trust. That is to say that it is like a certificate of confidence, that we are delighted that researchers are working on these questions.”
<i>Building innovation on shared values</i>	Responsible digital innovation relies on the values integrated into technologies that lead to doing no harm and positively impacting society.	“Bringing a principled European vision of digital innovation,” “based on the report of ethics guidelines for Trustworthy Artificial Intelligence published by the EU in 2019.” “the respect of the user with an ethical approach “The Web also carries all these values ... being aware of how the Web was built, what direction it has taken, in our opinion, is a desirable future that might be nicer. “This three-year strategic partnership for research work will strengthen our company in three areas: digital trust, artificial intelligence, and share the ambition of common values.”
<i>Reflecting upon what we want to transform</i>	Based on the firm’s exact objectives, such as enhancing its environmental footprint as a reputation strategy	“also not necessary to aim for an economic model that maximizes profitability, but also to have a development model which is good for everyone.” “For some time, we have been digitally transforming the audio. Yet, as we are public services, innovation is the key, so we cannot do it without awareness, so we are doing responsible digital innovation.” “The direct impact we make is a technology that evaluates candidates during recruitment instead of using CVs. This technology transforms the process of assessing cognitive skills and the recruitment process into a fun video game where candidates forget they are being evaluated.”
<i>Considering RDI as a structural element of competitiveness:</i>	This definition asserts the compatibility between responsible digital innovation and competitiveness. It also	“I think there is a subject that interests me: why is this subject important as a competitiveness cluster? And therefore, to manage to describe in a fairly strong innovation community, to tell how these

	highlights the necessity “to go back to a mandatory approach because we can no longer do otherwise” and “to have a job that makes sense.”	responsible digital subjects are future subjects for innovators?”

Tableau 7 : Definitions of RDI based on participants

2.4.2 Responsible digital innovation dimensions and practices

The second stage of the interviews concentrated on exploring the dimensions and practices of RDI. The qualitative data analysis identified five primary dimensions of RDI, depicted in Figure 18, along with 25 sub-dimensions, each exemplified by best practices and presented in Table 8.



Figure 18 : RDI dimensions (Source: The authors)

Dimensions	Sub-dimensions	Practices
RDI Strategy	Creating a mission-based status for the company	B-corp company
		Code of Ethics
	Integrating RDI in CSR	RDI in Corporate Digital Responsibility
		Mentioning RDI engagement in annual reports
	RDI ‘by design’ positioning	Open Source
		Trustworthy digital products and services
	Creating a corporate culture of RDI	White papers
		Building the RDI mindset
		Training teams

Responsible by Design Digital Innovation		External and internal communication
		SDG action plan
	Limiting the impact of DI on the environment	Carbon neutrality
		Eco-conception
		Digital frugality
		Carbon-neutral products and services
		Recycling
		Reconditioned digital equipment
		Improving ecological efficiencies
	Data privacy	Respect for data privacy
		Compliance with GDPR
		Protection of non-personal data
		Data governance charter
		Data ethics committee
		Contribution to Open Data
		Data bias suppression
		Data risks measures
		Data for Good
		Trustworthy AI
	Ethical artificial intelligence and algorithms	Non-discrimination by machine learning
		Strict rules for machine learning usage
		No bias added by algorithms
		Refusal of problematic AI application
		Cybersecurity
	Data and information system security	Data security
	Digital sovereignty	Data storage in Europe because of the GDPR data protection
		European technical solutions
	Inclusiveness	Digital technologies for all
		Digital divide decrease
		Accessibility
		Digital mediators
		Diversity
		Women's place in technology
Organization Challenges and KPI on RDI	Learning by collaboration and partnerships	Participation in collective initiatives
		Building on existing work
		Open innovation with RDI players
	Large-scale training	Training sessions on ethics
		Training sessions on eco-design
		Training sessions on mission-based companies
		Training sessions on data and AI challenges
	Value-chain RDI by selecting partners and suppliers	Priority to partners respecting diversity among their employees
		Work on building commons
		Selection of suppliers and partners respecting RDI criteria
	Human resources and organization	Dedicated human resources on RDI
		RDI enablers within big companies
		Creating awareness among young start-uppers about sustainability issues
		Regular meetings on RDI
	KPIs	RDI societal, economic, and environmental impact measures
		Weak signals

		Technical tests (security, data anonymization, trustworthy AI, explainability)
Digital Responsibility Towards End Users	Co-innovation with end users	Test panels
		Participatory budgets
		Users in ethics committees
	Transparency toward end users	Transparency in data collection
		Transparency of algorithms & data processing
		Open Source use
	Positive value of technologies for users	Added value maximization for users
		To serve clients and not to harm
	Data collection frugality	Data collection limitation
		No personal data collection
	Control back to users	Collaboration tools
		Guarantees of the positive impact of the innovation
		Autonomous check of technology by users
Catalysts	Top management support	CEO RDI support
		CSR RDI support
	Benchmarking	Comparison of RDI with same economic sector organizations
		Monitoring best practices
	Regulation	Respect for European GDPR
		Future European regulation on AI (AI Act)
	Trial & error management	Regular improvement to achieve RDI
		Step by step
	Fundings	Answers to international (ex: European) Calls for projects
		External funding for RDI
		Internal budget for RDI

Tableau 8 : Responsible digital innovation dimensions, sub-dimensions, and practices (source: the authors)

Responsible Digital Innovation Strategy

Based on the qualitative interviews, participants regard the concept of *RDI strategy* as a fundamental factor of RDI required for aligning DI with responsibility.

1. Creating a mission-based status for the company

A mission statement is an action-oriented articulation of an organization's *raison d'être* and customer service methods. According to Lee (2022), a mission-based model enables the comprehensive life-cycle management of a set of programs, describing four main dimensions of mission: purpose, philosophy, definition, and goals. The literature supports the idea of being adaptable to changing missions due to external pressures. For example, Laperche et al. (2011) explored how significant industrial companies shifted their innovation strategies towards affordable and eco-friendly technologies in response to economic downturns. Likewise, a recent investigation by Baudier et al. (2023), which analyzed responses from 1,233 participants, found that during the COVID-19 pandemic, the healthcare industry shifted from focusing on resilience to embracing an anti-fragility stance, thereby revising its procedures and integrating digital

innovations to tackle unforeseen obstacles. This research offers concrete evidence of industries adjusting their operational approaches to meet new requirements.

Mission-based or B-corp-oriented companies establish strategies to impact society and the environment positively. Generally, this target falls short for all business divisions, products, or service innovations. As a consequence of this, technological advancement must adhere to these standards. A second practice that came to light from the interviews is developing ethical guidelines for digital innovation. This sub-dimension of mission-based status creation extends methods highlighted in prior literature, such as board-level ethics committees (Garcia-Blandon et al., 2020) or boards of directors on business sustainability performance (Naciti, 2019).

2. Integrating RDI in CSR

When discussing the integration of RDI, participants expressed concerns over the potential ability of corporate social responsibility (CSR) to mask proper environmental or social outcomes. There was a unified view that DI could help close the transparency gap seen in CSR efforts. This perspective aligns with previous literature on Corporate Digital Responsibility (CDR) and CSR (Trittin-Ulbrich & Böckel, 2022; López Jiménez et al., 2021). According to Orbik and Zozuaková (2019), “In the Digital age, CSR and digital transformation are becoming one of the main factors of the company's competitiveness” (p. 79). Also, Jung and his colleagues (2022) found that CSR and IT-enabled innovation have a positive impact on firms' long-term financial value, but a varying degree of interaction effect exists between the two (CSR & IT-innovation) depending on the firm's available financial resources.

Despite this, not many CSR criteria emphasized the digital innovation and development of the firm. Accordingly, the European Business Network for CSR has hosted webinars and workshops in different countries to explore how organizations can address corporate digital responsibility (csreurope.org). Analysis of our qualitative interviews indicated a demand for the inclusion of RDI-specific standards within the larger CSR framework. This inclusion would entail showcasing RDI successes in annual reports and pursuing 'independent assessments using these RDI benchmarks.'

3. Positioning the digital product or service as RDI “by design”

The third practice involves placing the innovation (product or service) on the market as digitally responsible. Respondents provided examples, such as using open source, creating

trustworthy digital products or services, and selecting digital innovations with positive social and environmental impact. This approach of responsibility “by design” for digital innovation confirms the necessity to think about the effects of the innovation during the innovation process, as underlined previously by Ahmad et al. (2013) and Sivasubramaniam et al. (2012).

4. Creating a corporate culture of responsible digital innovation

Digital transformation entails critical organizational structure, operations, strategy, and culture changes. Gonçalves et al. (2022) discovered that startups embrace digital collaboration tools offered as cloud services due to their flexibility and because they successfully support the firm’s agile culture by aligning these tools with their behavior and values, allowing firms to create agile capabilities. There are 12 cultural values integrated into digital transformation: openness towards change, customer-centricity, agility, innovation, continuous improvements, tolerance towards failure, risk affinity, risk affinity, communication, trust, participation, and cooperation (Tuukkanen et al., 2022). Scholars like Antunes et al. (2021) investigated the significance of culture in an organization. Yet, there are few investigations on the practices to implement the culture and, more specifically, the culture of responsibility in digital innovation. This study highlighted practices mentioned by the participants in the qualitative interviews in creating a corporate culture on RDI. This includes different techniques such as elaborating “writing papers,” “building the mindset,” showing to the employees “the importance of society in the digital world,” “training teams,” “making internal and external communication,” “sustainable development goals (SDG) action plan” on digital innovation.

Responsible for the Design of Digital Innovation

Respondents to the qualitative interviews described the second dimension of Responsible Digital Innovation: responsible by design DI. They mentioned six main challenges that digital innovation-orientated companies should address.

1. Limiting the impact on the environment

Interviewees emphasized the importance of reducing the carbon footprint linked to DI, pointing out the potential environmental consequences of these technologies. Masud and his colleague (2019) echoed by various scholars who have highlighted the ecological aspect of DI (Masud et al., 2019).

In practice, a new sustainable entrepreneurial model called “digital sustainable entrepreneurship” has emerged to boost sustainable digital entrepreneurship (Xu et al., 2022, p. 1). According to Khan et al. (2021), circular economy techniques assist businesses in enhancing their environmental performance and financial performance, ultimately leading to tremendous organizational success. According to the SMARTer 2030 report⁹ from GeSi and Accenture Strategy, “rolling out identified ICT solutions across the global economy, the world could cut its global emissions by 12Gt CO₂e by 2030”. In accordance with interview highlights, several practices have been examined using aggregate terms such as “aiming for carbon neutrality,” implementing “eco-conception,” “reducing carbon impact by integrating life cycle analysis,” implementing “digital frugality,” “thinking carbon neutral products or services,” “recycling,” use of “reconditioned equipment,” and “improving ecological efficiencies.” Respondents underlined the necessity to decrease the carbon impact of digital innovation.

2. Digital protection

Providing proper data protection to cloud users while enabling complex applications is challenging (Song et al., 2012). Interviewees highlighted the significance of safeguarding data, a concern that mirrors societal implications in existing research (Alouffi et al., 2021; Sasubilli & Venkateswarlu, 2021). According to Isaak and Hanna (2018), four core principles of privacy and data protection: public transparency, disclosure for users, control (e.g., users must easily be able to delete personally identifiable data from any site or cloud), and notification (e.g., users must be directly and promptly informed of the loss or misuse of their private information by any organization collecting or storing that information).

In accordance with past studies, interviews illustrated the same concerns of the “data protection” challenge. First, the practices mentioned in this domain are for companies implemented in Europe concerning data privacy and compliance with the General Data Protection Regulation (GDPR). Yet, companies face challenges in compiling with the GDPR (Sirur et al., 2018). However, respondents also referred to practices above the GDPR, such as “also protecting non-personal data,” implementing a “data governance charter” or a “data ethics committee,” developing “open data,” working on “data of general interest,” reducing “data biases,” and “measuring data risks.”

⁹ https://smarter2030.gesi.org/downloads/Full_report.pdf

3. Ethical artificial intelligence and algorithms

Since their inception as theoretical concepts, AI and ML have become essential to modern society's reliance on technology. Our qualitative interview analysis brought to light a distinct challenge related to AI: the concept of trustworthy AI. John-Mathews et al. (2022) illustrate that numerous companies need to revise their understanding of AI ethics concerning algorithms to develop new frameworks of corporate responsibility that reflect shifts in the computing landscape.

In terms of practices, respondents confirmed some recommendations from reports, including the European Commission (High-Level Expert Group on Artificial Intelligence report), the Organization for Economic Cooperation and Development, and UNESCO. Participants acknowledge “ensuring the non-discrimination of minorities by machine learning algorithms,” “avoiding new biases with algorithms,” “defining strict and explicit rules of AI usage,” “refusing certain artificial intelligence applications,” and “setting up tools to explain algorithms process.”

4. Data and information system security

Respondents placed particular emphasis on security challenges in RDI. In a highly connected environment, world uncertainty, geopolitical risks, and economic competition create an obligation to secure information systems. The practices mentioned in the interviews rely on “cutting-edge cybersecurity techniques” and “ensuring data security.” The necessity to secure information systems and data confirms previous literature such as Alouffi et al. (2021) and Sasubilli & R (2021).

5. Digital sovereignty

Digital sovereignty has become a key element in policy discourse on digital issues. There is an increase in the use of the concept of “sovereignty” in relation to terms such as “digital,” “data,” and “technology” (before 2011, “Digital sovereignty” 6 times, and between 2015 and 2018, frequency 239) (Couture & Toupin, 2019, p. 2). The interviewees stressed the importance of digital independence, focusing on organizations' challenges in maintaining control over their digital data, infrastructure, and technology without external interference.

The participants emphasized the necessity “to host data in Europe on account of the GDPR data protection,” “to select European technical solutions,” or “to become a data producer” for

the company to keep its independence in this area. This confirms the digital sovereignty challenge highlighted by Floridi (2020).

6. Inclusiveness

The significance of fostering inclusivity in digital environments was a pivotal point made by participants, reflecting literature concerns over discrimination and bias in Digital Innovation (Howard & Borenstein, 2018; Morse et al., 2021). Khera et al. (2022) investigated the role of digital innovation in financial services and found that digital financial services have been a critical driver of financial inclusion. Yet, according to the same authors, more policies are needed to close the digital divide to ensure continued progress in financial inclusion and safeguard trust in financial services. In certain nations, sector-specific tariffs and taxation have hampered economic progress (World Bank 2016). For instance, Olorunleke (2019) studied the context of digital inclusion in Nigeria and found three challenges: user access, affordability, and lack of appreciation of libraries and poor funding for them.

To address inclusiveness, participants recommended efforts to make digital technologies accessible to everyone. They suggest “promoting digital technologies for all,” addressing the issue of the “digital divide,” participating in “actions for mediators,” integrating “accessibility” of products and services, enforcing standards, developing an “inclusive digital product/service,” to ensure “diversity” and “women’s place in digital innovation.” One of the four dimensions of responsible innovation identified by Stilgoe et al. (2013) was inclusiveness.

Organization and KPI

This qualitative study highlights the third dimension of RDI, which is the requirement of a specific organization within the company and the implementation of KPIs to manage RDI. Kraus et al. (2021) comprehensive review of the literature revealed 'organizational factors and managerial implications' as crucial elements, underscoring the complex link between strategic management, business models, and organizational aspects. Against this backdrop, the necessity for organizations to embrace RDI becomes clear. Our research further explores this theme, introducing a third facet of RDI: creating a dedicated division within the organization and applying Key Performance Indicators (KPIs) to oversee RDI efforts. Through our interviews, we identified five primary sub-aspects related to this matter.

1. Collaboration and partnerships

Interviewees highlighted the value of external collaborations in RDI, with scholarly work pointing to the significance of innovative teams and digital partnerships (Guinan et al., 2019). In this vein, Pagani (2013) emphasizes that for firms to survive in an increasingly uncertain and complex environment, they must transform their organizational intelligence into a new relational intelligence by enacting an open communication process with stakeholders. Traditional vertically integrated models are no longer sufficient as they centralize rather than maximize organizational intelligence. Instead, loosely coupled coalitions and multisided platforms are more effective in creating value within a connected, digital networked world. These models support creativity and diversity while maintaining coherence and coordination among participants. Furthermore, there is a noticeable shift in the critical role of networks from being content gatekeepers to becoming customer gatekeepers, distinguishing five layers of value: customer services, common infrastructure, modularity, content access, and orchestration.

Furthermore, Hassan et al. (2023) identified seven facilitators for effective collaboration between companies and communities, including readiness for collaboration among key individuals, strategic vision alignment, effective management and governance models, suitable software and hardware tools, resource allocation, external value networks, and a culture of open collaboration. Our findings expand on this concept by focusing on collaborations tailored to RDI.

Respondents underlined the necessity to collaborate with external partners on RDI. This practice consists of “building on existing work,” “participating in collective initiatives,” or “making open innovation with RDI players.” In this vein, open innovation is one of the vital concepts correlated to this practice. Open innovation refers to an open yet deliberate collaboration and sharing of internal and external knowledge within a corporation (Brunswick & Vanhaverbeke, 2015; Scuotto et al., 2020). Project management and open innovation manifest complementary strengths in partner identification and selection (Guertler & Sick, 2021). A fundamental strategy for open innovation is “crowdsourcing,” which challenges a broad and diverse group to obtain more robust new solutions than those within the firm.

2. Large-scale training

Some scholars define training as a firm's ability to enhance knowledge and develop employees' skills to fulfill current and future tasks (Guest et al., 2003). The effectiveness of RDI depends on providing employees with comprehensive training in ethical and sustainable

practices. This aligns with the views of Rampa and Agogu   (2021), who see innovation training as a vital link connecting various organizational aspects.

This study showed that RDI relies on large-scale employee training. The practices consist of implementing regular training sessions on different topics such as “ethics,” “responsible digital innovation,” “eco-design,” “mission-based company,” “data challenges,” and “responsible artificial intelligence.” Trainers need to adapt the training and topics to the different categories of employees, such as all employees or only teams of data scientists or developers. Kraus et al. (2021) also noted practices like 'flexing'—adapting to external pressures, 'deepening'—adjustments in employee behavior and role perception, and 'revitalizing'—the impact of continuous disruptions on job tasks and skill requirements.

3. Selection of partners and suppliers

Respondents mentioned the necessity of selecting partners and suppliers regardless of their RDI strategy. More particularly, they highlighted some practices such as “applying RDI criteria to choose suppliers and partners,” “giving priority to partners respecting diversity among their employees,” and “working on building commons” vertically with the companies contributing to the value chain. These practices enhance the argument of existing research suggested by Yang et al. (2022) that instability among suppliers could detrimentally impact a company's CSR performance internally and externally.

4. Human resources and organization

According to the interviewees, RDI requires specific human resources and organization within the company. This is in line with Jotab   et al. (2022), who examined the role of innovation within HR. In terms of practices, they mentioned “creating dedicated job positions within big companies,” “creating RDI enablers within the company,” “raising awareness among young start-uppers about sustainability issues,” and “organizing regular meetings on RDI.” This human resources and organization dimension extends to digital innovation, as shown by the work of Jotab   et al. (2022) on innovation and human resource management.

5. KPIs

The responses from interviewees underlined the necessity for clear Key Performance Indexes (KPIs) designed to evaluate the impact of RDI initiatives. Respondents put great emphasis on “creating indicators and KPIs of the societal and environmental impact of digital

innovation," "measuring companies' RDI results vs. objectives," "detecting weak signals," and "setting up technical tests" (e.g., on cybersecurity, data anonymization, AI bias or machine learning explicability). By identifying tasks, assets, and environments relevant to KPIs, planners, and decision-makers can more readily lead investigations to identify issue areas inside a facility, better allocate resources, and drive more effective analysis for monitoring and improving a facility (Navinchandran et al., 2022).

Digital responsibility towards end users

We identified a fourth dimension of RDI by analyzing qualitative interviews: actions towards the digital innovation users. Interviewees mentioned several practices linked to this dimension:

1. Co-innovation with users

Insights from participants highlighted the importance of engaging end users in the RDI process. While user engagement is a traditional aspect of innovation, combining it with DI enhances credibility. Engaging users in innovation, particularly in co-creating sustainable solutions, empowers them to influence product design and customization significantly (Leiting et al., 2023).

Co-innovation is a traditional innovation strategy, but it creates additional trust between users and companies for digital innovation. The practices mentioned are "using test panels," "participatory budgets," or "integrating users into ethics committees" during the innovation phase.

2. Transparency towards users

Transparency enhancement is a pillar of the current open-science reform movement. RDI requires transparency towards users. This transparency concerns "data collection," "understanding the algorithms and the data processing," "data storage," and the transparency on the use or non-use of "open source." Transparency confirms previous literature, such as Mittelstadt et al. (2016), highlighting the issue of transparency of algorithms. If ML relies on secondary data, Weston and her colleagues (2019) provide several recommendations to enhance transparency, including preregistering secondary data analysis and areas where researchers should:

1. Give links to codebooks and access instructions for data sets.

2. Explain how to utilize data.
3. Document the data-wrangling and -analysis pipeline.

3. Positive value of technologies for users

Interviewees underlined for RDI the importance of providing digital innovation for the benefit of the users. The objective is “to maximize the added value for the users” and “to serve and not harm them.” Several past studies document this finding. For instance, given the increased use of online platforms, firms consider chatbots a critical way to reach customers. According to Brandtzaeg and Følstad (2018), chatbots indicate a shift in the interface between users and technology and a change in user dynamics and patterns of use. This is because designers create chatbots to mimic human conversation. In line with the objective of “added value,” Marikyan et al. (2019) illustrated the innovative technologies used in smart houses to monitor, control, and support residents, which can enhance the quality of life and promote independent living.

4. Data collection frugality

Participants often emphasized that RDI demonstrates their dedication to their clients through a moral approach to data collection that promotes prudence and minimalism. In terms of practices, the objective is to “think innovative digital products or services with fewer data (specifically personal data) collected.” It is a new parameter that has not been investigated in the literature except from the transparency angle. In other words, there is little transparency in collecting the following data. Thus, Ienca and Vayena (2020) concluded that the “best practices should be identified to maintain responsible data-collection and data-processing standards at a global scale” (p. 463).

5. Give control back to users

Well-designed technologies that provide high levels of human control and computer automation can boost human performance, resulting in more adoption. RDI requires “providing users with digital control and collaboration tools” to give them the means “to check autonomously” that the digital innovation is not harmful and “to have guarantees about the positive impact of the innovation.” Taking AI as one application of digital innovation, many studies have explored the significance of the role of the human-centered principle. Shneiderman

(2020) elaborates on a framework called “Human-Centered Artificial Intelligence” (p. 495), clarifying that it is true that high levels of human control and computer automation increase human performance, yet it should be in balance.

Catalysts

In the interviews, we identified the fifth dimension of responsible digital innovation: catalysts, which were considered by the respondents as essential for a company to become performant in RDI.

1. Top management support

Top management support is considered one of the most investigated factors, and it is often shown to be crucial to the success or failure of a project (Khan et al., 2013; Pierre & Fernandez, 2018). The degree to which senior executives appreciate the value of the information system (IS) function and participate in IS activities is referred to as top management support (Ragunathan et al., 2004). Hwang (2019) found that top management support has a positive impact on projects, yet the “presence of this factor does not automatically guarantee success” (p. 354) and requires factors related to politics, cultures, and people. According to Majumdar and Sinha (2019), the lack of top management support and other criteria, like the lack of a reward system in Southeast Asia, were critical in not achieving sustainable goals.

The present qualitative study reanalyzed the role of “top management support,” the results reaffirm the significant effect of top management support on RDI success. The participants mentioned different management levels: chief executive officer (CEO)’s level, CSR responsible manager’s support (specifically when they belong to the company steering committee), and elected political leader for territorial public organizations.

2. Benchmarking

According to the interviewees, companies that compare their RDI strategy and actions with other organizations in the same economic sector are more inclined to improve their RDI practices.

Practices require “monitoring best practices” and “building regular comparisons with similar actors via expert intermediary partners.” The following are some examples exploring this result. In response to algorithmic injustice entrenched in sociotechnical systems, researchers have paid much attention to the contents of machine learning datasets, revealing

biases against white, cisgender, male, and other data subjects. Alizadehsalehi and Yitmen (2019) investigated how construction companies can benchmark themselves by adopting technologies. The authors found a positive relationship between automated progress monitoring and critical project performance indicators: time, cost, and quality. Denton et al. (2020) describe the ways in which benchmark datasets in machine learning operate as infrastructure by focusing on how to "bring the people back in" to aid designers in understanding the labor embedded in dataset construction, thereby presenting new avenues of contestation for other researchers encountering the data.

3. Regulation

The findings from participants highlight how regulation acts as a significant driving force for RDI. The discussion by Bunduchi and Candi (2021) illustrates how efforts to maintain corporate legitimacy adapt across different stages of technology development. From a broader viewpoint, recent research emphasizes that managerial innovation is shaped not just by direct regulations but also by external coercive pressures and the continuous search for legitimacy within the framework of open innovation (Bocquet & Dubouloz, 2020). This points to the intricate interplay between external regulatory requirements and a company's internal drive for legitimacy and innovation.

In production practices, firm reputation, and financial and environmental performance. Interviewees mentioned their own experience concerning the respect of European RDI on data protection and their anticipation of future laws such as the European "artificial intelligence act."

4. Trial/error approach

Another catalyst for RDI is a trial-and-error approach. The respondents were "in the process of ongoing and regular improvement on DRI" based on "learning dynamics."

5. Funding

This study revealed that a critical catalyst for RDI is dedicated funding. The utilization of green bonds, crucial for addressing climate change, hinges on the involvement and influence of institutional investors and public authorities (Ballouk & colleagues, 2023). Insights from our interviews emphasize that companies should "identify external fundings for RDI," answer "calls for projects" such as European sustainability ones, and identify "local and territorial fundings." Companies must also dedicate "internal budgets" for responsible digital innovation.

2.4.3 Barriers against responsible digital innovation

During the last part of the interviews, we asked the participants to describe the barriers they faced to implementing RDI. They emphasized five main issues described in Table 9.

Barriers	Barrier Clarification	Verbatims
<i>Lack of scaling</i>	Both scaling up and digitalization are transformation processes that must be managed. Issues such as a lack of leadership direction and organizational silos will impede the company's effort to support the transition.	"Difficulty in getting the various players on board," "lack of sharing RDI experiences," and "need to generalize projects."
Lack of knowledge about RDI	A lack of relevant knowledge could cause a company to lose confidence in technology and the firm's inefficiency.	"Underestimate the importance of RDI," and "do not master precise concepts such as equity in algorithms, societal or environmental impacts, digital inclusion."
<i>Lack of indicators</i>	One of the biggest reasons companies find it challenging to transition smoothly to digitalization is the lack of indicators. This is due to the complexity of the term Responsible Digital Innovation.	"Measurement of carbon footprint for digital innovation remains painstaking," there is a lack of "operational measurement indicators," "difficulties exist in measuring concepts such as trust," and "falling into the use of simplifying indicators of a more complex reality."
<i>Organizational challenges</i>	A lack of digital skills could jeopardize the company's digital transformation journey.	"Necessity for transversality."
<i>Technical issues</i>	The technical issues could be like the lack of IT resources and management	"Data anonymization," "machine learning," "explainability," "data bias identification," and "risk management."
<i>Lack of dedicated funding</i>	Funding a digital program is both an art and a science. It requires a clear understanding of the types of investment, for instance, maintenance investment (i.e., website security), foundational investments (i.e., enterprise resource planning)	"mismatch between aspirations to increased responsibility in digital innovation and funding."

<i>Controversies about effects</i>	Analyze friction and controversies with interaction processes and their effects on forming businesses. One example is social disconnection inside businesses.	“green technologies can serve the environment to the detriment of societal impacts” or “tensions exist on compatibility between RDI and competitiveness.”

Tableau 9 : Obstacles of RDI based on participants

2.5 Discussion

This study expanded the exploration of DI’s sustainability realms—environmental, societal, and economic impacts—by shedding light on managerial aspects through an RDI framework.

The first contribution comprises providing an integrative framework that combines RI and DI to construct the novel concept of RDI, which addresses the negative aspects of digital innovation. By extending beyond the specific domains examined in prior research (e.g., Trittin-Ulbrich and Bockel, 2022; Bastidas et al., 2023), this integrated framework provides a more comprehensive and universally relevant outlook on responsible digital innovation within the realm of business and management. This contribution furthers the discussion by offering pragmatic perspectives and systematic strategies for organizations aiming to harmonize digital innovation with sustainability objectives and ethical managerial conduct. As a result, it aids in the transition from traditional innovation management to responsible innovation management.

Second, our research enriches the understanding of RDI by providing definitions informed by the perspectives of digital innovation managers. While the academic realm has yet to thoroughly dissect the RDI concept—often presenting related terms within similar contexts—our study brings to light new dimensions. For instance, the notion of responsible digital transformation, as outlined by Pappas et al. (2023), aligns with our findings to some extent. However, based on specific organizational experiences, our research methodology reveals new layers, such as digital trust and competitive factors, broadening the discussion of RDI.

The third contribution of this research delineates five critical dimensions of RDI: RDI strategy, responsible design in digital innovation, organizational challenges and KPIs for RDI, digital responsibility to end-users, and catalysts for innovation. Bastidas et al. (2023) previously developed an RDI framework focused on three core areas: digital and technical aspects, governance and management, ethical considerations, and responsible innovation. It's important

to note that Bastidas et al.'s framework precisely assesses the competencies necessary for leadership in digital innovation within the built environment. In contrast, our framework utilizes an interdisciplinary approach incorporating insights from real-world interviews, enhancing its adaptability and relevance for broader applications. Additionally, our study introduces 25 sub-dimensions, some of which confirm aspects identified in earlier research by Alouffi et al. (2021) and Lobschat et al. (2021), and others that introduce vital new elements such as RDI strategy and digital responsibility to end-users. Therefore, the framework presented here is a practical guide for companies looking to implement RDI effectively.

Lastly, identifying barriers to RDI implementation represents a critical contribution to this study. Recognizing that companies may face internal challenges in utilizing available resources for RDI, our research delves into these obstacles, offering valuable insights for developing comprehensive RDI implementation strategies within firms.

2.6 Managerial implication

The findings from our investigation into RDI critical insights for managerial decision-making aimed at cultivating responsible practices within the realm of DI.

Corporate governance must formulate RDI strategies at its highest level. The company's board of directors should participate in such strategic discussions to ensure that RDI integrates into the organization's primary objectives and culture.

Technical teams should adopt RDI principles from the beginning of product or service development to ensure responsible innovation in all technological solutions and innovations.

Human Resources departments are essential in creating an organizational structure that promotes RDI. This entails establishing particular roles and teams to supervise and implement RDI concepts across company activities, including training employees on RDI.

Marketing teams should prioritize actions that promote RDI principles to end users, stressing openness, ethical use of technology, and sustainability in marketing efforts. Cloarec et al. (2024) suggests that companies can enhance users' willingness to share personal information for personalization by focusing on factors that promote happiness, trust, and privacy. Firms should provide a positive online experience and highlight the benefits of personalized content, such as convenience and relevance, to increase user happiness. Creating

an enjoyable and user-friendly online environment can improve user well-being. Additionally, companies should be transparent about their data usage policies and demonstrate a strong commitment to protecting user privacy, in line with GDPR and other relevant data privacy regulations. This transparency can build trust among users, reassuring them that their personal information is handled responsibly

Top support management, like the CEO's level or CSR responsible manager, as the catalyst for RDI, should prioritize these initiatives, ensuring adequate budget allocation and organizational focus toward RDI practices. Incorporating additional criteria Loan (2024) mentioned, such as holistic stakeholder engagement and organizational adaptability, can further guide companies in embedding responsible practices. A holistic approach to stakeholder engagement and a willingness to adapt organizationally to RDI principles are instrumental in seamlessly integrating responsible practices.

Enforcing reporting on RDI KPI throughout the designated dimensions may significantly enhance accountability and transparency. By coordinating these reporting requirements with already-existing frameworks, such as the sustainability reporting standards of the European Union, a unified approach to evaluating and disseminating RDI initiatives is promoted by ensuring uniformity and comparability across sectors and countries.

2.7 Policy implications for enhancing RDI practices

Policymakers have a significant role in aiding the adoption of RDI practices across many economic sectors. Incentives such as tax reliefs or subsidies may encourage companies to integrate RDI into their operations. However, policymakers need to design such policies with foresight to address potential implementation challenges, like those Carle (2024) identified in the context of eco-innovation. The author highlights the significance of careful "anticipation" in policy design to avoid confusion and the impression of greenwashing by entrepreneurs, as well as confusion concerning the meaning of the integration of environmental and societal criteria in such a context, transparency, and acceptability challenges.

The study proposes the establishment of a national and independent platform that evaluates companies based on their adherence to RDI principles, as outlined by the five dimensions and 25 sub-dimensions identified in our research. This platform would serve as a

benchmark for RDI implementation, providing visibility and recognition for companies that excel in RDI practices, thus driving broader industry compliance and ambition.

2.8 Conclusion

In conclusion, as organizations pursue DI projects to gain a competitive edge, it becomes progressively more essential for them to recognize and fulfill their responsibilities to the environment, society, and the economy. Addressing management difficulties arises as a critical component of this responsibility. This chapter proposes a three-phase framework to help organizations navigate the complexity of RDI adoption. Our findings highlight the necessity of explicitly defining RDI and providing a full explanation to enhance understanding. The second phase involves identifying the various dimensions and sub-dimensions of RDI, which provides businesses with an in-depth structure for incorporating DI into their strategic planning. Finally, recognizing and tackling the obstacles to RDI is necessary for a comprehensive strategy for RI.

Businesses that adopt the proposed framework will be more equipped to implement RDI and overcome the inherent obstacles of DI. This study provides valuable insights for firms looking to shift from a profit-driven innovation approach to one that values both social and environmental benefits, supporting sustainability alongside economic and managerial success.

However, this study presents limits. It was conducted with a small number of participants and focused on a restricted geographical area, and its findings are informed mainly by the insights of digital innovation managers. The qualitative aspect of this research emphasizes the need for additional quantitative analysis to improve the framework's validity. Furthermore, the participant selection process, which prioritized persons with an existing interest in responsible digital innovation, shows a possible bias.

Future research should delve into the efficacy of RDI strategies in practice, develop measurable indicators for assessing RDI, and investigate the long-term evolution of RDI initiatives that would significantly increase the body of knowledge in this field. Expanding the research to include participants from a broader geographical and cultural spectrum will provide deeper insights into the worldwide application and perception of RDI, offering new paths for understanding and implementing responsible innovation in various scenarios.

Chapter 3: Ethics in AI: Analyzing Declarations and Future Directions

Chapter 1: What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI? What are the risks of GAI and their impact on individual society organizational and environment?



Chapter 2: What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?



Chapter 3: What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?



Chapter 4: What are the essential dimensions of RAI and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?



Chapter 5: What are the specific generative AI risks? How do these GAI risks affect different aspects of society, business, and the environment?



Chapter 6: What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?

About this chapter

This chapter has been submitted for publication in the **journal**:

Haidar, A. Ethics in AI: Analyzing Ethical AI Declarations and Future Directions. *International Journal of Technology and Globalisation* (ABS 1).

This chapter has been selected and presented at the **conference**:

Haidar, A. (2023). Ethical artificial intelligence: An exploratory study of guidelines. In Artificial Intelligence: Law, Ethics, and Policy Conference. 20 June, Toronto, Canada.

Accepted for the **book**:

Haidar, A. (2024). Privacy-preserving Computing: Balancing Security and Privacy in the Digital Age. In A. U. Mirza & B. Kumar (Eds.), *Exploring the frontiers of artificial intelligence and machine learning technologies*. (pp. 198-210). San International Scientific Publication. <https://doi.org/10.59646/efaimltC11/133>.

3.1 Clarifying ethical AI

3.1.1 What is ethics

Ethics is a branch of philosophy that deals with moral principles and values that guide human behavior (Adeusi, 2020). Additionally, Fisher (2014, p. 2) further enriches this discussion by introducing the concept of "moral language," which navigates through pivotal questions like "What is the right thing to do?" and "What constitutes a good life?" Numerous scholars have delved into the definition of the term "ethics." According to Kidder (2003), ethics is the "science of the ideal human character" or "the science of moral duty" (p. 63). To further refine the definition, ethics encompasses "a set of concepts and principles that guide us in determining what behavior helps or harms sentient creatures" (Elder & Paul, 2006, p. 2). In the realm of science, ethics involves the contemplation of moral issues that arise during research, publication, and other professional endeavors (Proctor, 1998). Adeusi (2020) identified three distinct aspects within the realm of ethics: normative ethics, descriptive ethics, and meta-ethics, as described in Table 10. *Normative ethics* is concerned with determining what is right and wrong and defining terms such as "ought," "value," "good," and "duty." Unlike scientific inquiry, normative ethics is not about describing the world but how it should be. While *descriptive ethics* is a part of science and only explains how people use normative ethical terms, some ethicists argue that scientific findings can guide how one should live, leading to the development of scientific ethics (Quintelier et al., 2011). *Meta-ethics* involves systematically

examining moral language, moral psychology, and moral ontology (Fisher, 2014). *Applied ethics* deals with the practical application of ethical principles to specific situations. It examines questions such as “What is the ethical approach to euthanasia?” and “What is the appropriate response to climate change?” *Applied ethics* seeks to guide how to act ethically in real-world situations and to address specific moral dilemmas (Fawcett et al., 2018). Casuistry is a practical field of applied ethics (Jonsen & Toulmin, 1988) that emphasizes examining context and specific cases rather than starting with theoretical frameworks. Its ideas have been employed in various areas, including law, bioethics, and business ethics, to apply ethical reasoning to individual cases (e.g., Ciulla, 2022; Müller, 2013). Specific to different groups, such as cultures, religions, or professions, *particular ethics* are defined by norms, as Varkey (2021) describes. These norms include various responsibilities and ideals, such as the professional ethics seen in the medical field. A relevant illustration of this is the role of physicians, who are morally bound to deliver capable and dependable care to their patients, representing the essence of morality in their professional standards.

Ethics is fundamental in many fields, such as technology, medicine, law, business, and politics, where ethical considerations are often paramount. It helps individuals and organizations make well-informed, morally justifiable decisions and act according to ethical values and principles. Ultimately, the study of ethics provides a framework for understanding morality’s nature and making decisions that reflect our deepest values and principles.

Normative ethics	The main focus is to make moral judgments about particular actions or kinds of actions and offer reasons to support the moral judgment made to show that the decisions are reasonable (for example – prostitution is bad; it can lead to incurable diseases).
Descriptive ethics	Answers factual questions about the moral view of individuals or groups. Basically, statements of fact are made about the moral views held by individuals, groups, or society at large (for example – “Susan believes prostitution is bad” or “Christians hate corrupt practices”).
Meta-ethics	It is concerned with questions about the meaning of ethical terms. It attempts to analyze ethical concepts in order to arrive at their actual meanings and logical relations.

Tableau 10 : Types of ethics (source: Adeusi, 2020)

3.1.2 Ethical AI: definition and moral philosophies

To address ethical dilemmas in AI, significant discussions have emerged over the past decade on balancing ethical concerns with digital transformation (Culnan and Williams, 2009; Roša & Lobanova, 2022; Zhanbayev et al., 2023; Zhang and Hon, 2020). Newell and Marabelli (2015) emphasized the need to explore the concept of 'irresponsible' AI analytics and its influence on individuals, organizations, and society. In response, Wright and Schultz (2018) introduced an ethics AI framework that combines stakeholder and social contracts theories to establish best practices for ethical AI challenges, including acknowledging transitions, minimizing disruptions, and reducing social inequalities. Flyverbom et al. (2019) further discussed the importance of responsible business practices in the digital era, advocating for businesses to extend their focus beyond profit to address societal challenges through digital technologies like big data analytics. The rapid transformation brought on by AI has inevitably sparked a surge of interest in the ethical implications of these advancements. This intersection of ethics and AI has birthed a new field commonly referred to as “ethical AI” or “AI ethics” (Winfield et al., 2019). According to Morley et al. (2020), “The need for ‘AI Ethics’ has arisen from the fact that poorly designed AI systems can cause very significant harm” (p. 2161). As AI systems become more intricately woven into the fabric of our lives, they raise significant ethical questions and present substantial challenges across privacy, bias, accountability, transparency, and human dignity domains. These challenges are complex, spanning the technological, legal, and social realms, and demand thoughtful consideration and rigorous analysis. However, recognizing the necessity to incorporate ethical considerations into creating computational artifacts, particularly algorithms, is not a recent revelation, as Wiener (1961) and Turing emphasized in the 1940s and 1960s (Turilli 2007).

Ethical AI refers to developing and deploying AI systems that are ethical and aligned with human values (Theodorou & Dignum, 2020). Incorporating ethical principles into the design and development process of AI systems and ensuring that AI systems are used consistently with these principles is essential. AI ethics is a developing field that falls under the broader umbrella of digital ethics, which aims to address concerns arising from creating and using emerging digital technologies, including but not limited to AI, big data analytics, and blockchain technologies (Kazim & Koshiyama, 2021). The field of AI ethics has emerged in response to the increasing risks of the impact of AI due to the misuse of the technology or as a result of the technology's design flaws (Kazim & Koshiyama, 2021). For instance, criminals used AI-based voice technology to impersonate a chief executive's voice and demand a fraudulent transfer of

\$243,000 (Stupp, 2019). Within this context, cyber risk has become one of the most significant threats in recent years, so boards of directors must be vigilant (Radu & Smaili, 2022). PINDROP, a company that develops call center security software and protocols, reported a 350% increase in voice fraud between 2013 and 2017 (Livni, 2019). In addition, it is necessary to gain consumers' trust so that firms can use their data well (Rossi, 2019). According to the Digital News Report 2019, only 43% of Spanish users trust the information they receive, compared to 32% of skeptics (Vara-Miguel et al., 2017). For these reasons and due to many other concerns, the ethical perspective of AI must be considered (Haenlein et al., 2022). Additionally, a survey found that 86% of users will trust and remain loyal to companies that embrace ethical AI principles (Barometer Edelman Trust., 2019). The ethics of AI can be divided into roboethics and machine ethics (Siau & Wang, 2020). For these authors, roboethics involves the moral behavior of humans in the design, construction, usage, and interaction with AI. At the same time, machine ethics deals with the moral behavior of artificial moral agents. In some applications, such as automated driving or autonomous weapons, ongoing discussions exist about whether it can delegate ethical decision-making to machines. In spite of this discussion, ethics "are founded on the abstract notion of human dignity and are related to human autonomy and agency, the capacity to deliberate and to act freely and intentionally" (Chatila et al., 2021, p. 16).

Many different ethical theories can be used to guide the development of AI. For instance, "utilitarianism" is an ethical doctrine that posits the fundamental moral principle as the pursuit of the maximum amount of happiness achievable, with every individual's happiness equally important (Mill, 2016). While "deontology" does better than utilitarianism in analyzing ethics in economics, where ethics appears in utility functions as moral preferences (Van Staveren, 2007). Each framework emphasizes different ethical principles, such as maximizing society's overall well-being, respecting individual autonomy, or promoting social justice. Within the scope of AI, numerous ethical frameworks, such as respect for individuals, beneficence, justice, privacy, accuracy, ownership/property, accessibility, fairness, accountability, and transparency, can guide human behavior (Siau & Wang, 2020). The Markkula Center Framework identifies five approaches to addressing ethical challenges, namely the utilitarianism approach, the rights approach, the fairness or justice approach, the common good approach, and the virtue approach (Tseng, 2015).

There are also concerns about the potential societal impacts of AI, such as job displacement and the exacerbation of social inequalities. Ethical AI must consider these concerns and strive to mitigate AI systems' adverse societal effects by embedding certain principles and actions, including enhancing guidelines and resources for product managers, fostering organizational awareness, encouraging employees to identify AI ethical risks through formal and informal incentives, and consistently evaluating impacts while maintaining active stakeholder engagement.

Overall, the development of ethical AI requires a multidisciplinary approach that draws on expertise from philosophy, computer science, social science, and other fields. By incorporating ethical principles into the design and development of AI systems, we can ensure that these systems are aligned with human values and contribute to a better world.

3.1.3 Trustworthy AI

Trustworthy AI refers to “programs and systems built to solve problems like humans, which bring benefits and convenience to people with no threat or risk of harm” (Liu et al., 2022, p. 5). Stakeholders, including consumers, organizations, and society, consider AI trustworthy when it meets legal standards, exhibits reliability, and, most importantly, adheres to established ethical principles in its development, deployment, and use (Thiebes et al., 2021). The European Commission’s High-Level Expert Group (2019) has outlined three key characteristics that define “Trustworthy AI.” Firstly, it should be *lawful* and comply with all relevant laws and regulations. Secondly, it should be *ethical* and uphold fundamental ethical principles and values. Lastly, considering both technical and social perspectives, designers must ensure robustness to prevent unintentional harm, even when they design the AI system with good intentions. The same report states that trustworthy AI is crucial in enabling “responsible competitiveness” (p. 5).

Also, Chatila et al. (2021) define the AI “technical trust” (p.15) concept with seven properties: availability, reliability, safety, confidentiality, integrity, maintainability, and security. These properties ensure correct and continuous service, absence of catastrophic consequences, absence of unauthorized information disclosure, prevention of improper system alterations, ability to undergo modifications and repairs, and secure access for authorized users only. Additionally, researchers propose multiple methods to develop trustworthy AI. For instance, Wilson and Daugherty (2018) have suggested the concept of collaborative

intelligence, which utilizes both humans and machines to make decisions. Alzubaidi and his colleagues (2023) stress the need for trustworthiness in AI systems before their deployment to fulfill the AI goal for good. These approaches share a common goal of guaranteeing that AI systems operate as intended, without causing harm to individuals or society, thereby fostering trust in these systems. Raita and Oulasvirta (2011) explored methods to enhance stakeholders' trust by aligning with their expectations. In contrast, some companies adopt a different approach known as 'zero trust,' which follows the principle of 'never trust, always verify,' as discussed by Laplante & Voas (2022).

The concept of trust in AI aligns with anthropomorphism—the attribution of human-like characteristics to non-human entities such as AI technology (Golossenko et al., 2020). This process helps users predict or understand the behavior of AI entities, especially when knowledge about the AI's underlying mechanics is limited (Epley et al., 2007; Pfeuffer et al., 2019). A positive perception of an anthropomorphized AI entity can lead to its association with beneficial attributes such as trustworthiness, reliability, or competence (Benlian et al., 2020; Mourey et al., 2017; Wang et al., 2016). Behavioral traits like friendliness enhance consumer interactions with AI products (Gonzalez-Jimenez, 2018; van Doorn et al., 2017). Product design with anthropomorphic facial features can also contribute to trustworthy robot communication (Torres-Carrión et al., 2019). Other studies suggest that anthropomorphism increases acceptance of AI by enhancing its social presence—the extent to which AI conveys relational information and develops a psychological relationship with the user (Qiu & Benbasat, 2009; Schultze & Brooks, 2019).

3.2 Methodology

This research employs a mixed-methods approach, incorporating frequency analysis to examine the distribution and occurrence of ethical principles within AI ethics declarations and an integrative literature review. This methodological blend allows for the exploration of commonalities, pinpointing notable omissions, and highlighting new patterns in the ethical governance of AI, offering a detailed overview of the incorporation and prioritization of fundamental moral principles.

The selection of 27 ethical AI statements encompasses a broad spectrum, including international organizations such as the OECD, national and regional bodies, NGOs, and leading tech companies, particularly those within the GAFAM group (Google, Apple, Facebook,

Amazon, and Microsoft), which play a critical role in shaping these ethical principles. The criteria for selection focused on the declarations' impact and relevance, as demonstrated by their significant influence on AI policies worldwide and their frequent reference in scholarly literature, alongside their geographical representation to ensure a global perspective on AI ethics. For each declaration, we applied a binary coding method to record the presence (1) or absence (0) of specific ethical principles, facilitating a systematic quantitative analysis through the statistical software SPSS. This analysis examined the frequency of how these principles are represented across the declarations. This methodological blend allows for the exploration of commonalities, the pinpointing of notable omissions, and the highlighting of new patterns in the ethical governance of AI, offering a detailed overview of the incorporation and prioritization of fundamental ethical principles.

Literature reviews are a pivotal tool for researchers seeking to assess or scrutinize the consistency and precision of specific theories or among competing ones (Tranfield et al., 2003). They lay the groundwork for formulating new theoretical models or concepts, proving invaluable in tracing a specific research field's evolution over time (Snyder, 2019). Kraus et al. (2020) recognize two types of literature reviews: systematic literature reviews (SLRs), which researchers conduct following a systematic method for data curation and analysis to ensure transparency and replicability, and non-systematic literature reviews (non-SLRs); which are based on the authors' critical evaluations and subjective choices, lacking a systematic approach. Also, Snyder et al. (2019) distinguish between the different forms of literature reviews—systematic, semi-systematic, and integrative. Specifically, an integrative literature review is distinctively conceptualized to critically assess, evaluate, and synthesize existing research findings to contribute to generating new theoretical frameworks and perspectives (Torraco, 2005). The study utilized an integrative review to analyze a topic across diverse research designs and methodologies, enhancing understanding of the investigated phenomenon (Kraus et al. 2022). This review methodology often demands a more innovative strategy for data compilation, with the aim extending beyond a comprehensive coverage of all existing articles on the subject. Instead, it seeks to integrate diverse perspectives and understandings from multiple academic fields or research traditions. This review methodology often demands a more innovative strategy for data compilation, with the aim extending beyond a comprehensive coverage of all existing articles on the subject. Instead, it seeks to integrate diverse perspectives and understandings from multiple academic fields or research traditions (Snyder, 2019). This approach has led to significant breakthroughs in the field of management, with several

integrative reviews marking critical advancements (Ardichvili & Jondle, 2009; Benzaghta, 2021; Hossain, 2015).

This literature follows an established protocol summarized as follows:

Search database: The selection of search databases can be grounded in justified evidence, like the use of Scopus and Web of Science, because of their status as the most extensive databases of scholarly articles offering access to bibliographic data or records on demand (Pranckutė, 2021). This research uses Google Scholar (GS) and Business Source Complete (BSC) as primary databases to construct a comprehensive and nuanced search strategy. As supported by Bauer and Bakkalbasi (2005), the rationale for using GS lies in its proven capability to provide significantly higher citation counts for recent scholarly publications, thereby ensuring access to a broad spectrum of contemporary academic work. Friend (2006) further emphasizes GS's tailored service for the educational community, which can efficiently isolate relevant scholarly information from the vast internet expanse.

Search keywords and Boolean operator: To ensure a comprehensive retrieval of relevant literature, the search query is: ("AI ethics" OR "AI declaration" OR "AI guidelines") AND ("opportunities" OR "limitations"). This approach allows me to filter through vast amounts of literature efficiently, pinpointing articles that discuss the critical aspects of ethical AI declarations.

Search period: The search covered publications from 2013 to 2023, targeting a specific timeframe to align with developments in AI ethics during that period.

Publication stage: Selection was restricted to documents in their final publication stage to ensure the use of refined and conclusive research findings, facilitating greater precision in the study's replication efforts.

Document type: The inclusion criteria encompassed both review articles and empirical studies to provide a comprehensive understanding of theoretical discussions and practical applications within the field of AI ethics

Source type: The primary source for this literature is journals due to their rigorous peer-review processes, which have contributed substantively to the discourse on AI ethics.

Language: Researchers conducted the literature review using documents published in English, the predominant language for scholarly communication.

3.3 Findings

3.3.1 Analysis of ethical principles in AI declarations

3.3.1.1 An updated round-up of ethical AI declarations

In recent years, impressive advances in artificial intelligence systems' capabilities have created an ethical debate about society's opportunities and concerns. Since 2016, many global stakeholders, including policymakers, industry leaders, civil society organizations, think tanks, media, and consultancy firms, have been actively engaged in intense discussions (Ulnicane, 2022). In their 2020 study, Morley et al. (2020) identified a compilation of over 70 documents published between 2017 and 2019 that offer guidance on the ethical principles of AI. This list includes documents produced by industry (Google¹⁰, IBM¹¹, Microsoft¹², Intel¹³), Government (Lords Select Committee, European Commission's High-Level Expert Group), and academia (Future of Life Institute, IEEE, AI4People). Similarly, Jobin et al. (2019) identified 84 documents containing ethical principles or guidelines for AI and found a significant increase of 88% after 2016.

Numerous actors have released guidelines, principles, and frameworks to ensure AI's ethical and responsible deployment. This chapter examines and synthesizes ethical guidelines from major global players, national and regional entities, non-governmental organizations, and tech giants in shaping these ethical directives, as illustrated in Annex 2.

3.3.1.2 Frequency analysis across overall declarations

In the analysis of ethical AI principles across the studied declarations, as depicted in Figure 19, three distinct groups have emerged, outlining potential focal points for subsequent discourse:

¹⁰ Google's AI Principles: <https://www.blog.google/technology/ai/ai-principles/>.

¹¹ IBM's everyday ethics for AI: <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.

¹² Microsoft's guidelines for conversational bots: https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf.

¹³ Intel's recommendations for public policy principles on AI: <https://blogs.intel.com/policy/2017/10/18/naveen-rao-announces-intel-ai-public-policy/#gs.8qnx16>.

1. **Predominant principles:** These principles, appearing in over 50% of the declarations, demonstrate a broad agreement or concern within ethical AI. The analysis reveals transparency (81%), fairness (67%), accountability (63%), and privacy (56%), underscoring the foundational values in ethical AI considerations.
2. **Intermediate principles:** the category encompasses principles falling within a moderate frequency range of 40 to 50%. The analysis uncovers principles of security (48%), safety (44%), explainability (44%), and beneficence/well-being (41%). Their relatively less frequent citation may stem from several factors: the complexities involved in their practical application, the challenges posed by the rapid progression of AI technologies, their representation of evolving ethical discussions, and the gradual formation of consensus in new focus areas.
3. **Principles with limited emphasis:** Representing principles mentioned in fewer than 40% of the declarations, this group includes adherence to legislation or laws (15%), data governance (11%), and disinformation (4%). This suggests that these areas are currently less emphasized within ethical AI discussions, potentially indicating overlooked aspects within ethical frameworks that require further examination. The minimal attention to issues like disinformation could be attributed to their recent recognition as significant concerns, highlighted by specific declarations such as the “Global Declaration on Information Integrity Online” in 2023, which addresses the increasing importance of these issues in the face of challenges from advanced AI technologies.

The findings underscore certain principles consistently, reflecting these declarations' significant role in shaping the ethical landscape of AI practices. Conversely, principles that emerged less frequently in the analysis point to existing limitations and areas for improvement within ethical frameworks. These discrepancies underscore the importance of continuous updates and refinements to ethical declarations, ensuring they remain relevant and effective in addressing the dynamic challenges of AI technology. The subsequent section further explores the implications of these findings through an integrative literature review (ILR), interpreting the opportunities and limitations revealed by the frequency analysis.

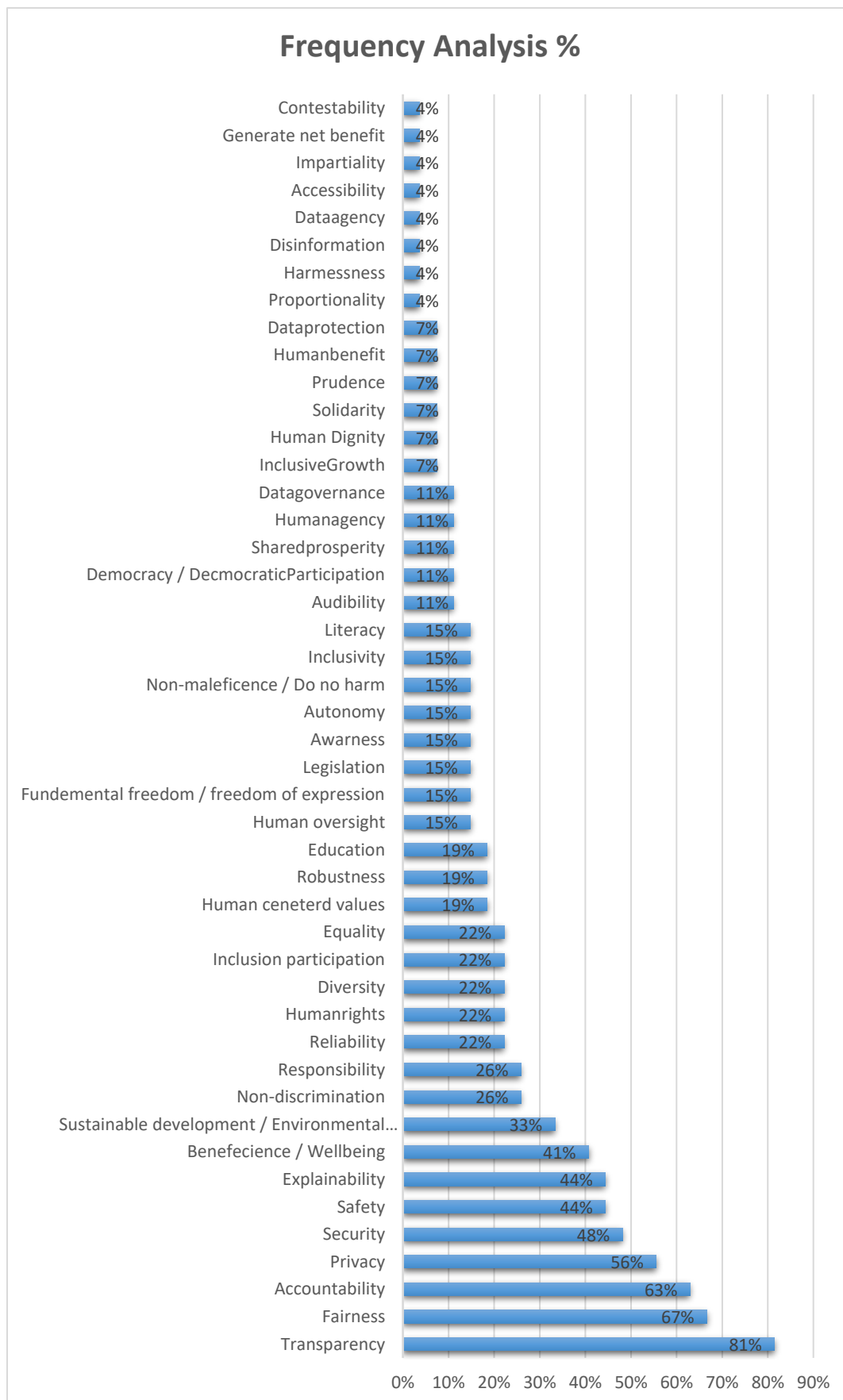


Figure 19 : Frequency analysis of the principles across all declarations (source: author)

3.3.2 Integrative literature review findings

3.3.2.1 Strengths and opportunities presented by ethical AI declaration

Within the scope of Ethical AI declarations, identifying their strengths (illustrated in Table 11) is vital for understanding how businesses may harness their particular competencies to traverse the ethical considerations in AI use and innovate responsibly in the AI domain. From a business perspective, strengths refer to businesses' internal resources and advantageous attributes, enabling them to fulfill their goals and effectively deliver value to their customers (Culp et al., 2016; Phadermrod et al., 2019). Similarly, Teoli et al. (2019) describe 'opportunities' as actual conditions in the broader surroundings that entities can utilize to their advantage. Within the AI ethics declaration frame, Table 11 summarizes the crucial opportunities.

Opportunities & Strengths	Description	References
Global Collaboration	Advocates for global consensus on AI ethics, emphasizing inclusivity and harmonization across borders. This dimension includes various aspects: OECD principles, private sector engagement, role of NGOs, geographical distribution of AI ethical principles, and Inclusion of underrepresented communities	Chan et al. (2020); EY (2023); Fjeld et al. (2020); Fukuda-Parr & Gibbons (2021); Global Declaration on Information Integrity Online (2023); Greenleaf (2019); Jobin et al. (2019); van Buitenlandse Zaken (2023).
Comprehensive Coverage Harmonization of Global Standards	AI ethics guidelines and initiatives emphasize a broad spectrum of ethical considerations, aiming for a holistic governance framework that includes principles like transparency, accountability, and privacy. The principle is characterized by multidisciplinary vision for AI ethics and technical solution for ethical goal	Corrêa et al. (2023); Díaz-Rodríguez et al. (2023); Du & Xie (2021); Hagendorff (2020); Harrison et al. (2021); Rothenberger et al. (2019); Smuha (2019).
Addressing Ethical Issues in AI	AI declarations have the ability to address the ethical issues related to AI, like unfair outcomes, and transformative effects.	Bostrom & Yudkowsky (2014); Mittelstadt et al. (2016); Morley (2020); Stahl et al. (2021); Wang & Siau (2018).
AI For Social Good And The Preservation of Trust	AI declarations aim to promote well-being, preserve dignity, and sustain the planet. This involves enhancing human flourishing and contributing to societal well-being and the common good.	Adadi & Berrada (2018); de Fine Licht & de Fine Licht (2020); Du & Xie (2021); EC's High-Level Expert Group on AI (2019); Faggella (2020); Farouk (2022); International Risk Governance Center (2018); Jobin et al. (2019); Wachter et al. (2017).

Tableau 11 : Opportunities and strengths of AI declarations (source: author)

Global participation

The notion of 'Global Participation' is critical for advancing AI technology as it advocates for global collaboration and endeavors to achieve the most extensive possible consensus on ethical AI principles worldwide. A notable illustration of this international collaboration is the 'Data Free Flow with Trust' (DFFT) initiative spearheaded by Japan. Greenleaf (2019) notes that the initiative aimed to foster international trust and cooperation in digital economies, marking a significant step during the G20 Leader's Summit in Osaka. Despite the challenges in achieving a unified stance following the broad terms of the Osaka Declaration on Digital Economy and the varied participation of key global players (24 countries), the initiative signals a move towards consensus-building in "global data privacy rules" (p.1). This effort, aligning with the privacy principle crucial to AI ethics, underscores the necessity for inclusive dialogues and agreements that transcend national borders and aims for a harmonized approach to AI governance. Within this interpretation, the strength of AI declarations, particularly those like the Global Declaration on Information Integrity Online (2023), is their grounding in international law, including human rights treaties. According to van Buitenlandse Zaken (2023), this declaration establishes a set of high-level international commitments by participating various states. This foundation enhances their credibility and effectiveness. Also, Fukuda-Parr and Gibbons (2021) reviewed 15 guidelines for ethical AI and found more than half around their work on international human rights law. Consequently, these declarations promote good governance and ethical standards in global AI use by transcending borders and focusing on universal values such as equality, freedom of expression, and media freedom. (Fukuda-Parr & Gibbons, 2021)

The OECD AI Principles are a unique example of worldwide collaboration on AI ethics and can be considered a global benchmark for AI-guiding policymakers (EY, 2023). These principles, developed by the Organisation for Economic Cooperation and Development, establish an international standard for AI development and use, emphasizing inclusive growth, sustainability, and human-centered values. These principles provide a complete framework for ethical AI and have affected national AI plans worldwide, proving their broad impact and numerous nations' commitment to ethical AI development. For instance, President Biden's Executive Order on AI on October 30, 2023, adheres to eight guiding principles and priorities that align closely with the OECD's principles (The White House, 2023). Similarly, the G20 AI Principles, adopted in 2019, highlight the world's largest economies' shared approach to AI

governance. These principles, closely aligned with the OECD framework, underscore the global commitment to foster innovation while guaranteeing RAI development.

The breadth of AI ethical declarations goes beyond international organizations and government entities, with significant contributions from the private sector and non-governmental organizations. For example, top technology corporations such as Amazon and Google have created their own AI principles and guidelines to demonstrate their commitment to RAI development and use. These company-specific declarations frequently highlight concepts such as justice, safety, privacy, and accountability, which fit broader global norms. In this regard, several studies compute the participation of the public sector in AI declarations. The AI Ethics Lab's¹⁴ analysis reveals that in their inventory of 103 documents, principles from private companies constitute 34%. Research by Fjeld et al. (2020), in their review of 36 documents, shows that private funding accounts for 22%. Additionally, the AI Ethics Guidelines Global Inventory by AlgorithmWatch¹⁵ indicates that private companies represent 26% of the 160 documents they included in their selection. This data underscores the notable contribution of the private sector in shaping AI ethics principles.

Additionally, non-governmental organizations play an essential role in shaping AI ethics. CEPEJ (European Commission for the Efficiency of Justice) is one example of an organization that has issued guidelines on the ethical use of AI in judicial systems. Their work demonstrates how non-governmental perspectives contribute to a more thorough and broad knowledge of AI ethics, ensuring consideration of numerous industry-specific concerns.

These declarations from tech companies and NGOs highlight the multi-stakeholder nature of AI governance. They complement the efforts of international bodies and governments, underscoring the collective commitment to developing and implementing AI in a manner that is ethical, responsible, and beneficial for society.

According to the data analyzed by Jobin et al. (2019), there is a significant concentration of AI ethics principles emanating from the United States (25%) and the United Kingdom (15.5%), as well as EU member states, Canada, Iceland, and Norway, which account for 26%. Japan also provides 4.8%, with the UAE, India, Singapore, South Korea, and Australia each

¹⁴ <https://aiethicslab.com/big-picture/>

¹⁵ <https://inventory.algorithmwatch.org/>

having one document, constituting 6%. This distribution implies that Western values contribute a notable significant effect, with these regions accounting for 67% of all AI ethics principles.

In discussing global participation as a strength in AI declarations, Chan et al. (2020) emphasize the importance of involving historically underrepresented communities in AI development. The author notes efforts in the Global South to create opportunities for local community engagement in AI programs and stresses the need to expand roles beyond data labeling to more substantial contributions in model development and strategic decision-making. Also, Feijóo and his colleagues (2020) propose the concept of "technology diplomacy" to facilitate international cooperation and coordination in AI development and deployment. This approach is essential for building a more inclusive and diverse AI ecosystem globally.

Comprehensive coverage harmonization of global standards

The declarations and initiatives cover many aspects of AI ethics, including transparency, accountability, privacy, fairness, and security, indicating a holistic approach to AI governance. By 2023, there will be more than 200 documents on AI ethics guidelines (Corrêa et al., 2023). Among these, Smuha (2019) explicitly elucidates the framework of the European Commission High-Level Expert Group on AI guidelines, considering it a “comprehensive framework to achieve Trustworthy AI” (p. 97). The study by Rothenberger et al. (2019) exemplifies the comprehensive coverage of AI ethics in current discourse, initiating an extensive literature review focusing on classical Western ethical theories, laying the foundation for developing vital ethical guidelines for AI. The researchers then conducted qualitative interviews with experts to gather in-depth feedback and rank these guidelines in terms of importance. Their findings are based on an arithmetic mean ranking with 'Responsibility' as the highest priority (4.51), followed by 'Protection of Data Privacy' (4.4), 'Transparency' (4.28), 'Robustness' (4.25), 'Minimization of Bias' (3.87), and the importance of AI having a clear purpose (3.58). In addition to the broad spectrum of AI ethics aspects covered in declarations and initiatives, spanning transparency, accountability, privacy, fairness, and security, the study by Díaz-Rodríguez et al. (2023) reinforces the significance of a holistic approach to AI governance. Highlighting the critical role of ensuring AI is lawful, ethical, and robust from both technical and social perspectives, Díaz-Rodríguez et al. (2023) complement the findings of Corrêa et al. (2023) and Rothenberger et al. (2019) by providing a framework that encompasses the trustworthiness of all processes and actors involved in the AI system's life cycle. This framework not only coincides with the comprehensive coverage identified by Rothenberger et al. (2019) through their systematic analysis of ethical guidelines and expert interviews but also

extends it by proposing a multidisciplinary vision that includes global principles for ethical use, a philosophical perspective on AI ethics, and a risk-based approach to AI regulation. In this vein, Du & Xie (2021) categorized AI ethical challenges into three distinct levels. The first level, concerning product-related ethics, involves addressing AI biases and ensuring ethical design. The second level, which focuses on consumer-related ethics, encompasses privacy concerns associated with AI-enabled products and related cybersecurity issues. Lastly, the third level targets societal and ethical problems, including preserving individual autonomy and well-being and the potential unemployment resulting from job replacement by AI systems.

In his analysis of 22 AI guidelines, Hagendorff (2020) found that the aspects of accountability, privacy, or fairness are present in approximately 80% of the guidelines. This prevalence suggests that these elements are considered foundational for the development and use of AI systems that are ethically sound. Hagendorff (2020) observes that AI guidelines' most commonly cited aspects, such as accountability, fairness, and privacy, lend themselves to technical solutions. The author notes that considerable technological advancements have been made to address ethical goals in areas like accountability and XAI (Mittelstadt et al., 2019), fairness in data mining (Gebru et al., 2018), and privacy (Baron & Musolesi, 2017), indicating a trend towards developing technical fixes for ethical challenges in AI.

The work of Harrison et al. (2021) addresses the challenge posed by the vast and growing number of AI ethics guidelines, exceeding 300 documents. It critiques the limitations of traditional systematic review methods for analyzing such a broad corpus, proposing the innovative use of Semantic Web technologies instead. This approach involves creating dynamic knowledge graphs by developing the first ontology for Ethical AI principles (AIPO), enhancing the interconnection, analysis, and application of AI principles. By employing Semantic Web tools and AIPO, Harrison et al. (2021) aims to transform the sprawling landscape of AI ethics guidelines into a more organized, accessible, and actionable framework. This methodological shift signifies a potential leap forward in harmonizing AI ethical standards, making it a pivotal contribution to the discourse on comprehensive AI governance.

Addressing ethical issues in AI

Since the notion of "machine ethics," as posited by Anderson and Anderson (2006), the scholarly discourse surrounding the ethical dilemmas concerning machines has persevered through extensive deliberation and debate. One of the significant strengths of AI declarations

is their ability to address ethical issues, as detailed in Table 12 from Morley (2020) and developed initially by Mittelstadt et al. (2016). In this vein, Wang and Siau (2018) note that research on the ethics of AI encompasses three main categories: potential ethical challenges arising from AI features, human-related factors contributing to ethical risks, and methods for imbuing AI systems with ethical behavior. The same authors differentiate the ethical issues posed by narrow AI and weak AI. The ethical problems of narrow AI involve human factors: “A different set of ethical issues arises when we contemplate the possibility that some future AI systems might be candidates for having moral status” (Bostrom & Yudkowsky, 2018, p.5).

Ethical Concerns	Explanation
Inconclusive evidence	Algorithmic conclusions are probabilities and therefore not infallible. This can lead to unjustified actions. For example, an algorithm used to assess credit worthiness could be accurate 99% of the time, but this would still mean that one out of a hundred applicants would be denied credit wrongly
Inscrutable evidence	A lack of interpretability and transparency can lead to algorithmic systems that are hard to control, monitor, and correct. This is the commonly cited ‘black-box’ issue
Misguided evidence	Conclusions can only be as reliable (but also as neutral) as the data they are based on, and this can lead to bias. For example, Dressel and Farid (2018) found that the COMPAS recidivism algorithm commonly used in pretrial, parole, and sentencing decisions in the United States, is no more accurate or fair than predictions made by people with little or no criminal justice expertise
Unfair outcomes	An action could be found to be discriminatory if it has a disproportionate impact on one group of people. For instance, Selbst (2017) articulates how the adoption of predictive policing tools is leading to more people of colour being arrested, jailed or physically harmed by police
Transformative effects	Algorithmic activities, like profiling, can lead to challenges for autonomy and informational privacy. For example, Polykalas and Prezerakos (2019) examined the level of access required to personal data by more than 1000 apps listed in the ‘most popular’ free and paid for categories on the Google Play Store. They found that free apps requested significantly more data than paid-for apps, suggested that the business model of these ‘free’ apps is the exploitation of the personal data
Traceability	It is hard to assign responsibility to algorithmic harms and this can lead to issues with moral responsibility. For example, it may be unclear who (or indeed what) is responsible for autonomous car fatalities. An in depth ethical analysis of this specific issue is provided by Hevelke and Nida-Rümelin (2015)

Tableau 12 : Ethical concerns related to algorithmic use, '(Source: Morley et al. (2020), adapted from the ‘map’ created by Mittelstadt et al., 2016).

In a similar study, Stahl et al. (2021) conducted a comprehensive investigation into the ethical issues of AI applications, identifying various concerns. Their analysis highlights 39 ethical challenges, ranging from the potential cost to innovation and harm to physical integrity to more complex issues like power asymmetries, the disappearance of jobs, and possible misuse for military purposes. The study also delves into problems like bias and discrimination, unfairness, unequal power relations, misuse of personal data, negative impacts on the justice system and democracy, and concerns about privacy, transparency, and loss of human autonomy.

The comprehensive list of 39 ethical issues identified by Stahl et al. (2021), including the cost to innovation, harm to physical integrity, diminished trust, and restricted access to public services, highlights the ethical considerations in AI. It underscores how AI declarations aim to address these myriad issues. By outlining principles and guidelines, these declarations seek to mitigate the potential negative impacts of AI, such as the loss of jobs, violation of privacy, or issues of bias and discrimination. They also focus on promoting fairness, transparency, and accountability in AI development and usage, thereby attempting to balance the advancement of technology with the protection of fundamental human rights and values.

AI for social good and the preservation of trust

The main contribution of a regulatory framework for AI suggests that “a clear regulatory framework would build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology” (European Commission, 2020a, p. 9-10).

The ethics of technology refers to the moral principles and values that guide the design, development, and use of technology in society (Leikas et al., 2019). Ethical issues arise in various technology areas, such as AI, biotechnology, and information technology. Some key ethical considerations in technology include privacy, security, transparency, accountability, and human dignity (Moor, 1985; Floridi, 2019).

The ethical examination of AI is almost old (Samuel, 1960; Wiener, 1960). Yet, in recent years, and due to the high capabilities and applications AI has brought us, the regulations of AI are getting more focus. A unified framework to map AI’s landscape comprises five main principles (Floridi & Cows, 2019; Floridi & Clement-Jones, 2019). In addition, the sixth principle, labeled “Risk Management,” was proposed recently with the European Union proposal (2021) and considered the impact of AI. Several scholars argue that ethically aligned machine learning describes technology that is both beneficial to and respectful of people and the environment (beneficence); it is robust and secure (non-maleficence); it upholds human values (autonomy); it is fair (justice); and it is transparent, accountable, and comprehensible (applicability). It also complements the other principles (Morley et al., 2020; Floridi & Cows, 2019).

1. Beneficence: promoting well-being, preserving dignity, and sustaining the planet – this principle underlines the importance of all AI to work towards the continued prosperity of humanity and the preservation of a socially responsible environment. Ethical values are linked to the potential for increased wealth and well-being, enabling individuals to

lead better lives and contribute to human flourishing. It is essential to recognize that this involves certain degrees of wealth distribution and specific assumptions about the roles of society and the state in ethically redistributing wealth. As the EC's High-Level Expert Group on AI (2019, p. 4) emphasizes, “AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation.” Declarations such as those from the EC's HLEG, the Montreal Declaration, and AI4People broaden the principle of beneficence to encompass all sentient beings and the environment. In contrast, the White House AI Principles do not explicitly identify beneficence as a core principle but rather emphasize the expectation for AI to positively affect various social and economic sectors. Furthermore, the principles outlined by China extend beneficence further, stressing the importance of harmony.

2. Non-maleficence: The principle of non-maleficence asserts an obligation to avoid causing harm to others, closely aligned with the *maxim primum non nocere*, which means “first, do no harm” (Farouk, 2022). Privacy, security, and “capability caution” protect against AI causing harm to humanity or invading individual rights such as privacy and prevent the mass adoption of AI in warfare. Some frameworks, like the Montreal Declaration and EU Ethics Guidelines, explicitly mention non-maleficence and detail actions to avoid harm. Others, such as the OECD Principles and Beijing AI Principles, embody the principle through broader commitments to safety, fairness, and human rights without explicitly using the term non-maleficence. The EU Ethics Guidelines and IEEE Ethically Aligned Design offer detailed considerations on how non-maleficence should be integrated into AI system development, focusing on both technical and ethical dimensions. In contrast, documents like the OECD Principles provide a more general policy framework that implicates non-maleficence through governance and oversight mechanisms. Tech companies go through this discussion and present terms like “Socially beneficial” (Google, 2018) and “Robustness and safety” (Facebook, 2021) that contribute to the non-maleficence principle.
3. Autonomy: the power to decide – a human agreement to cede some level of decision-making to AI in order for it to be self-directing to a degree. A common theme across these declarations is the emphasis on AI systems being developed and used in ways that respect and enhance human agency, allowing for meaningful human oversight and decision-making. Yet, some frameworks, like the Apple AI ethics guidelines and EU Ethics Guidelines, explicitly prioritize autonomy as a fundamental aspect of ethical AI.

Others, such as the OECD Principles, Beijing AI Principles, and all declarations on national and regional levels like France and Italy, approach it more indirectly, emphasizing alignment with human values and societal well-being, which supports autonomy.

4. Justice: promoting prosperity, preserving solidarity, avoiding unfairness, eliminating unfair discrimination, and supporting the need for shared prosperity. IEEE Ethically Aligned Design (2019) highlights the need to prioritize human rights and fairness in designing and developing AI systems. It calls for transparent mechanisms that ensure AI technologies do not become sources of injustice or inequality. Also, frameworks like the EU Ethics Guidelines and the Toronto Declaration explicitly prioritize justice, focusing on non-discrimination and fairness. Others, such as the Beijing AI Principles, approach justice more indirectly through the lens of societal well-being and the common good.
5. Explicability: enabling other principles through intelligibility and accountability facilitates and supports AI being beneficent and non-maleficent, allowing us to see the responsibility or harm it is doing to society and in which ways. The opportunity of "explicability" in ethical AI principles is crucial when developing and deploying AI systems. Explicability refers to the ability of AI systems to provide clear, understandable explanations of their decision-making processes, outcomes, and behaviors (de Fine Licht & de Fine Licht, 2020). This transparency fosters trust between users and AI systems (Wachter et al., 2017). It enables developers, regulators, and users to scrutinize the AI system's operations and identify potential biases, errors, or unintended consequences (Adadi & Berrada, 2018). Incorporating explicability into AI development can lead to more ethically responsible systems and better alignment with human values and expectations (Jobin et al., 2019). Many frameworks, across various levels, including tech companies (such as Apple and Microsoft), national entities (like Argentina and Saudi Arabia), and major global players like the EU Ethics Guidelines, explicitly recognize explicability as a core principle. They emphasize the need for AI systems to be understandable and their decisions explainable. In contrast, others, such as the Beijing AI Principles, implicitly suggest the importance of understandability and transparency without directly using the term.
6. Risk management: The International Risk Governance Center (2018) highlights AI's exceptional analytical capabilities, including processing vast amounts and diverse data sources beyond human reach. AI can connect data, identify patterns, and produce

outcomes across domains and geographic boundaries. In addition, AI can offer more consistency than humans, rapidly adapt to changing inputs, and alleviate humans from monotonous or repetitive tasks (Du & Xie, 2021). These technical abilities contribute to better comprehension and more profound insights into various phenomena, thus fostering human flourishing.

After reading the declarations and frameworks mentioned in section 3.1.1, Figure 20 represents three main dimensions concluded.

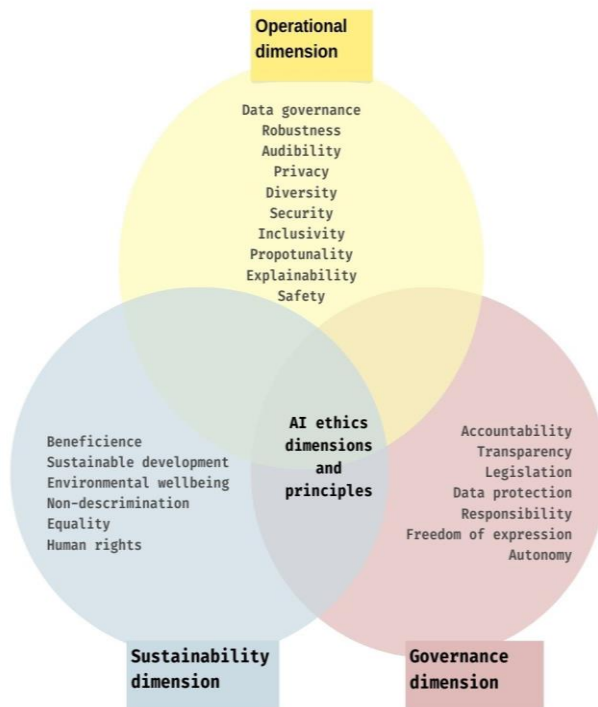


Figure 20 : Societal Principles & Dimensions (source: author's elaboration)

3.3.2.2 Weaknesses and limitations of AI ethics declaration

While Namugenyi et al. (2019) did not specifically address ethical AI, their definition of weaknesses as internal constraints can guide the identification of specific challenges in Ethical AI declarations. Building upon this, Ken et al. (2016) elaborate on how a structured brainstorming approach can identify specific internal limitations and challenges relevant to ethical AI, such as potential areas for improvement, obstacles to growth, and systemic issues that obstruct effective implementation and adherence to ethical standards in AI development and usage. Table 13 provides an overview of the weaknesses of ethical AI declarations.

Weakness	Description	References
Lack of a Standard Framework for Ethical AI	Significant diversity exists in the principles included in ethical AI documents, with no universality in approaches, indicating challenges in achieving a global consensus on ethical AI standards.	Beswick and Gandhi (2023); Corrêa et al. (2023); Fjeld et al. (2020); Franzke (2022); Hagendorff (2020); Jobin et al. (2019); Oniani et al. (2023); Royakkers et al. (2018); Tomašev et al. (2020).
Divergence in Definitions and Terminology in AI Ethics	Ethical AI declarations exhibit notable differences in the definitions and terminology used, alongside contradictions between principles such as robustness and privacy, revealing the intricate trade-offs that challenge the establishment of a unified ethical framework.	Australian AI Ethics Framework (2019); Binns (2018); Franzke (2020); Montreal Declaration (2018); Ryan and Stahl (2021); Uddin et al. (2022); Whittlestone et al. (2019).
Challenges in Translating Principles into Practices	High-level ethical principles often remain theoretical and do not easily translate into actionable design protocols, posing a threat to their influence on actual AI design and implementation.	Hickok (2021); Mittelstadt (2019); Prem (2023); Thiebes et al. (2021); Turilli (2007); Whittlestone et al. (2019); Zhou et al. (2020).
Integration with Existing Legal Frameworks	Aligning AI ethical principles with existing laws and regulations is complex, leading to potential conflicts or overlaps, and may result in calls for overly rigid regulations that could stifle innovation.	Claessens et al. (2018); European Commission (2018a); German Banking Industry Committee (2020); Government of Ireland's (2020); Justo-Hanani (2022); Owczarczuk, (2023); Pagallo (2017); Truby et al. (2020); Von der Leyen (2019); Winfield & Jirotko (2018).
Erosion of Trust Due to Limited Focus on Impact	The lack of attention to the real-world impact of AI systems, including on user autonomy, labor rights, technological unemployment, and long-term risks, undermines trust in AI.	Bolte et al. (2022); Morley et al. (2019); Hagendorff (2020); Corrêa et al. (2023); Gibney 2018) ; Goggin 2019); Goujard, (2023).
Usability and Practical Implementation of Ethical Tools	The development of ethical AI tools that are not yet ready for practical, real-world application adds friction to the implementation of ethical AI, indicating a gap in making these tools user-friendly and effective.	Attard-Frost et al. (2023); Bitsch et al. (2023); Morley et al. (2019) ; Ray (2023).
Ethical AI in a Competitive World	Competitive pressures and the preference for market-driven AI development may hinder the integration of ethical considerations, suggesting a conflict between ethical guidelines and the drive for innovation.	Arvan (2018); Butterworth (2018); Calo (2017); Claessens et al. (2018); European Commission (2020); O'Sullivan et al. (2019); Truby et al. (2020).
Public Acceptance, Investment, and Research Challenges	Concerns about misuse, technical limitations, and potential negative consequences of AI technologies can lead to public mistrust, reduced funding, and a decline in research, posing a threat to the advancement of ethical AI.	Giffi et al. (2018); Kassens-Noor et al. (2021); Lokshina et al. (2021); Lokshina & Lanting (2020); Morley (2022); Winfield and Jirotko (2018).
Geopolitical Tensions	The role of AI in geopolitical tensions, including the competition for dominance, challenges the collaborative effort required for global AI regulation and ethical governance, potentially leading to the misuse of AI in power rivalries.	Ali (2019); Bryson and Malikova (2021); Bächle & Bareis (2022); EU (2018a); Roberts et al. (2023); Thomann (2023)

Tableau 13 : Weaknesses of AI declarations (source: author)

Lack of a standard framework for ethical AI

Jobin et al. (2019) comprehensively reviewed 84 ethical AI documents and found significant diversity in their included principles. Notably, no single ethical principle was present in all documents, indicating a lack of universality in the approach to AI ethics. However, they identified common themes presented in over half of these documents, such as transparency (86%), justice (81%), responsibility (71%), non-maleficence (71%), and privacy (56%). In a

related aspect, Fjeld et al. (2020) analyzed 36 documents from diverse organizational backgrounds and identified eight frequently mentioned principles. These were fairness/non-discrimination (cited in every document analyzed), followed by privacy and accountability (each at 97%), transparency/explainability (94%), safety/security (81%), professional responsibility (78%), maintaining human control over technology (69%), and upholding human values (also 69%). A similar analysis conducted by Hagendorff (2020) but on a smaller sample of 21 declarations revealed that four ethical principles — accountability, privacy, justice (each cited in 77% of the declarations), and transparency (68%) — appeared frequently. More recently, Franzke (2022) undertook a comparable analysis of 70 international ethics guidelines documents from academia, NGOs, and the corporate realm. They frequently found transparency, security, and privacy repeated. In a similar vein, Royakkers et al. (2018) performed a systematic review of the literature on ethical technology. They found recurrent themes like privacy, security, autonomy, justice, human dignity, control of technology, and balance of powers. This again underscores the diversity in ethical considerations discussed in the context of AI. Building upon these analyses, Corrêa et al. (2023) expanded the scope of research with a quantitatively more extensive and diverse sample, analyzing 200 documents from 37 countries across six continents and in five different languages, highlighting a total of 17 ethical principles frequently mentioned, each cited in more than 10% of the documents. Their study also revealed that while the top five ethical principles remained consistent in North America and Europe, Asia notably introduced the principle of beneficence/non-maleficence as its fifth most cited (74%) and accountability/liability in sixth place (70%). The study also uncovered regional differences: for instance, in the USA, principles like freedom/autonomy and beneficence/non-maleficence featured prominently, whereas in the UK, the emphasis was more on cooperation/fair competition and diversity/inclusion principles. Furthermore, Corrêa et al. observed variances in principle prioritization based on institution type, with governmental bodies emphasizing transparency (89.5%), private corporations focusing on reliability (87.5%), and CSOs/NGOs advocating fairness (88.2%).

Tomašev et al. (2020) outlined ten principles essential for effective AI collaborations aimed at sustainable development. These principles focus on setting realistic AI expectations, valuing simplicity, ensuring inclusivity and ethical compliance, defining clear goals, and establishing long-term partnerships for solving significant challenges. They also stress the significance of aligning incentives, maintaining trust, exploring cost-effective AI development, improving data readiness, and processing data securely while respecting human rights and privacy. These guidelines serve as a comprehensive framework for organizations using AI to

achieve sustainable goals. According to Furrer (2022), ten main principles have been proven successful in teaching, enforcing, and implementing safe systems. These principles include safety culture, safety standards and policies, safety governance, safety management system, safety principles (which define rules for the construction, evolution, and operation of safe systems), safety implementation, safety assessment and audit, safety runtime monitoring, and secure software. Siala and Wang (2022) reviewed 253 articles on AI ethics in healthcare from 2000 to 2020, culminating in the development of the SHIFT framework—Sustainability, Human centeredness, Inclusiveness, Fairness, and Transparency—to address ethical concerns and guide RAI implementation in healthcare. From the GAI aspect, Oniani et al. (2023) propose nine ethical principles for GAI in healthcare entitled “GREAT PLA”: governability, reliability, equity, accountability, traceability, privacy, lawfulness, empathy, and autonomy. Also, Beswick and Gandhi (2023) introduced the RAFT (Responsible, Accountable, Fair, and Transparent) framework for responsible generative AI.

The variability and absence of a uniform approach can be viewed as a drawback in AI declarations. The diversity of principles and subjects indicates potential difficulties in developing a worldwide consensus on what constitutes ethical AI. This lack of homogeneity may result in inconsistent implementations and interpretations of ethics in AI development and governance. Accordingly, it creates challenges in enforcing global ethical standards for AI.

Divergence in definitions and terminology in AI ethics

One prominent weakness in AI ethics declarations is the disparity in how different declarations define and interpret ethical concepts. According to Theodorou and Dignum (2020), one of the critical challenges in developing ethical AI is defining the ethical principles that should guide the design and use of these systems. The many sources and situations from which these declarations come can explain this contradiction. Each declaration frequently reflects its origin's cultural, legal, and social subtleties, resulting in diverse interpretations and uses of comparable ethical concepts. According to Whittlestone et al. (2019), the ambiguity and varied interpretations of key terms used in ethical principles often mask the conceptual complexities and differing viewpoints across populations, making it challenging to apply these principles universally. For instance, the Montreal Declaration (2018) describes the 'well-being principle' as minimizing user stress, enhancing their mental and physical capacities, and improving living conditions. In contrast, the Australian AI Ethics Framework (2019) emphasizes the concept of 'net benefit,' suggesting that AI systems must create benefits for people that are more significant

than costs. While widespread agreement exists on the significance of 'fairness' in principle, Binns (2018) notes significant political debates over its precise definition. Additionally, cultural variations influence the prioritization of values; for instance, individualistic societies may place higher importance on personal privacy than collectivist ones. In a broader scope, Franzke (2020) argues that most existing guidelines do not adequately define “ethics” and related terms. Ryan and Stahl (2021) offer a critical analysis highlighting AI's comprehensive spectrum of ethical requirements. Their work sheds light on the varied emphases within ethical frameworks—from focusing on algorithms, privacy, and security to safety. Moreover, they note the differences in the tone of these frameworks, from prescriptive "must do" principles to more suggestive "if possible" guidelines. The length and accessibility of these documents also vary significantly, from succinct one-pagers to extensive guides of over 266 pages, catering to a broad audience, including end-users, developers, companies, policymakers, and society. This diversity underscores the challenge of achieving a standard definition for AI principles.

Another weakness captured is the divergence in the use of terminologies. For instance, the Montreal Declaration and Chinese declarations adopted the term “responsible AI,” while the Future of Life Institute's declaration emphasizes “beneficial AI,” while UNESCO focuses on “Trustworthy AI.” This notion variance underscores the lack of a unified approach to conceptualizing AI ethics. Moreover, the principles of these declarations are often interconnected across different facets of ethical AI (Abedin, 2022; Smuha, 2019). A pertinent example illustrates the complex interconnection between the robustness and explainability of deep neural networks. Evidence suggests that robust models tend to be more interpretable (Kim & Canny, 2017; Uddin et al., 2022), and conversely, more interpretable models can enhance robustness (Noack et al., 2021).

Challenges in translating principles into practices

One of the most significant weaknesses identified in the current framework for ethical AI is the difficulty in translating established ethical principles into practical, actionable design protocols and measures within the realms of AI in practice (Turilli, 2007). Prem (2023) underscores this issue by highlighting the significant lack of systematic approaches for monitoring AI ethics, particularly in the crucial stages of deployment and data creation. Despite efforts by companies like Facebook to institutionalize ethics through councils and boards, there remains a profound disconnect in translating these ethical frameworks into actionable practices. Prem's (2003) emphasis on the necessity of coaching, consulting, and labeling as vital

components for effective implementation further illustrates this threat. Another challenge in developing ethical AI is ensuring these systems are transparent and accountable (Cath, 2018). AI systems can be complex and difficult to understand (Yang et al., 2020), making it challenging to determine how they make decisions and whether those decisions are ethical. To address this challenge, researchers are exploring ways to make AI systems more transparent, such as by developing XAI techniques that can provide insight into the decision-making process of AI systems (Meske et al., 2022). Zhou et al. (2020) identify a further challenge involving the integration of fundamental principles—transparency, justice and fairness, responsibility, non-maleficence, and privacy—at every stage of an AI system's lifecycle. This integration ensures the ethical design, implementation, and deployment of AI systems.

Translating high-level ethical principles into actionable design protocols is a significant challenge in practice (Turilli, 2007). If AI declarations do not effectively bridge this gap, there is the threat that these principles may remain theoretical and not influence actual AI design and implementation practices. One of the main reasons is the lack of clarity for practitioners over who should be responsible for ensuring alignment between product design and ethical principles, as illustrated in Figure 21 (Morley et al., 2023).

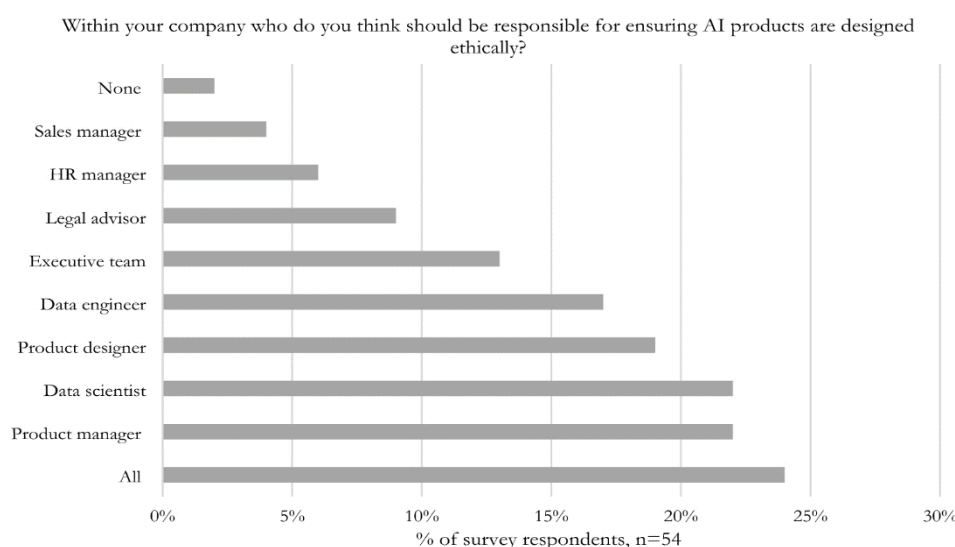


Figure 21 : Responsibility distributed over individuals or ensuring the pro-ethical design of AI products (source: Morley et al., 2023)

Another challenge is the tension between fields of application and among principles. To illustrate, Mittelstadt (2019) emphasizes that while AI ethics has agreed on principles similar to those in medical ethics, there are significant obstacles in applying these principles to AI research and governance. According to the same author, there are substantial contrasts between

medicine and AI, including a lack of common objectives, fiduciary duties, professional history and norms, proven techniques for practical implementation of concepts, and robust legal and professional accountability mechanisms in AI. Similarly, Whittlestone et al. (2019) identify four fundamental tensions arising from AI applications. The first tension is between using data to enhance service quality and efficiency versus upholding individual privacy and autonomy, highlighting privacy protection challenges and the need for meaningful consent. The second tension centers on balancing algorithmic decision-making accuracy and predictions against ensuring fairness and equal treatment. The third involves the benefits of increased personalization in digital spaces against fostering solidarity and citizenship. The fourth tension contrasts the convenience and empowerment of automation with the promotion of self-actualization and dignity. To clarify the last tension, automation can potentially democratize access to arts, science, and languages for previously excluded groups, yet it also risks causing widespread deskilling, practice ossification, homogenization, and cultural diversity loss. Following this line of thought, Thiebes et al. (2021) designate two significant limitations of the existing principles, frameworks, and guidelines related to Trustworthy AI (TAI). The first limitation is the potential conflict of TAI principles. One example is the tension between beneficence and justice values in AI applications such as predictive policing. While predictive policing might assist society by enhancing crime prevention, it may also include discriminatory behaviors, such as utilizing ethnicity in AI model training, which violates the ideal of justice. These issues frequently necessitate trade-offs that are ultimately difficult to settle. The second limitation highlighted by the authors is the generality of TAI principles. Current frameworks and guidelines are too broad and offer little practical guidance for implementation. Song et al. (2019) illustrate the apparent trade-off between robustness and privacy, underlining that while bolstering robustness, adversarial defense mechanisms may inadvertently heighten the model's susceptibility to membership inference attacks, thereby elevating the risk of training data exposure. Proceeding from this point, AI ethics principles focus on how to improve the system but “rarely question the business culture, revenue models, or incentive mechanisms that continuously push these products into the markets. In their frame review of ethical AI principles” (Hickok, 2021, p. 43).

These tensions, articulated by scholars like Thiebes et al. (2021), Hickok (2021), and Whittlestone et al. (2019), underscore a weakness in AI declarations as they often struggle to address and balance these competing priorities and challenges comprehensively and practically.

Also, these declarations fail to provide specific directions on translating these principles into practices, posing a challenge for effectively realizing the goals of ethical AI.

Integration with existing legal frameworks

Aligning AI ethical principles to be more practical and formalizing them with existing laws and regulations can be complex (Winfield & Jirotko, 2018). Declarations may not always consider these legal intersections, leading to potential conflicts or overlaps. According to the European Commission (2018a), any activity and trustworthy AI requires “a regulatory framework that is flexible enough to promote innovation while ensuring high levels of protection and safety” (p. 8). In early 2017, the European Parliament requested that the commission submit a proposal for a directive on Civil Law Rules for Robotics, and the parliament suggested the following:

“creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently” (European Parliament, 2017, p. 250).

However, this recommendation received significant criticism from AI experts and stakeholders. In an open letter to the European Commission, more than 200 AI, robotics, industry, law, and ethics professionals declared that bestowing electronic entities with a legal 'person' status is a misguided and impractical idea¹⁶. They contended that this approach overestimates the current abilities of robots, influenced by misconceptions shaped by science fiction and sensationalized media. Moreover, they argued that assigning legal personhood to robots is not an appropriate course of action from both legal and ethical perspectives. In this vein, representatives from the German industry, including the German Banking Industry Committee (2020), have emphasized the need to foster research and development while avoiding the pitfalls of over-regulation.

The General Data Protection Regulation (GDPR) is one of Europe's most well-known legislation, establishing a precedent for the significance of legal underpinnings in managing technology, including in the field of XAI (Hamon et al., 2022). The insights from Pagallo

¹⁶ Open Letter to the European Commission on Artificial Intelligence and Robotics. Retrieved from: <http://robotics-openletter.eu/>

(2017) underscore a significant threat to the effective integration of AI ethics declarations within existing legal frameworks: the nuanced challenges of applying the GDPR's anticipatory data protection measures in the rapidly evolving landscape of Big Data and AI. The GDPR's emphasis on preemptive risk assessments and its complex secondary rules (the practical implementation of these overarching principles) highlight the broader issue of ensuring that AI ethics declarations align with and enhance current legal standards. Furthermore, the challenge of balancing legal flexibility highlights the need for clear, actionable guidelines to keep pace with technological advancements and competition between regulatory systems.

Due to this importance of law as a basis for AI, in her 2019-2024 Political Guidelines for the European Commission, Ursula von der Leyen, the then-incoming President of the Commission, committed to proposing legislation focused on the human and ethical aspects of AI within her first 100 days in office (Von der Leyen, 2019, p. 13). This commitment led to the release of a white paper on AI in February 2020, which presented various options for investment and regulation in AI and initiated a public consultation process (European Commission, 2020b). In response to the white paper, Ireland, known as a digital frontrunner, raised concerns about potential over-regulation in emerging markets and pointed out the ambiguity in understanding the regulatory acquis. Ireland advocated for a more incremental and learning-based approach to regulation rather than preemptively attempting to predict and address all possible outcomes, as detailed in the Government of Ireland's (2020) report (p.3). Among the scholars advocating against AI regulations, Claessens et al. (2018) argue that overly strict or burdensome regulations could stifle innovation and limit the potential benefits of AI. Also, The White House has released draft guidance for regulating AI applications, which suggests avoiding regulatory or non-regulatory actions that needlessly hamper AI innovation and growth (Calo, 2017). Some argue that measures such as transparency, accountability, and fairness might promote AI growth and innovation (Truby et al., 2020), while others believe that the real task for AI regulators is to create a rules structure that protects the public and promotes industry innovation—not to trade off one against the other (European Commission, 2020b).

Pertaining to this matter, Owczarczuk (2023) highlighted three significant challenges regulators face when creating AI regulations. First, there is a need to develop a precise definition of AI, which is essential for establishing a comprehensive regulatory framework that encompasses the diverse range of AI technologies. Second, addressing ethical threats is critical to ensure AI systems adhere to ethical principles and protect human rights. Finally, the third challenge lies in navigating the competitive landscape of AI regulations across different

regions, such as the European Union, the United States, and China, which may adopt varying approaches to AI governance. Marsoof et al. (2023) address the ethical challenges in using AI for online content moderation, noting that while AI helps regulate internet content, its limitations in accuracy and transparency can lead to removing legitimate content and retaining harmful content. This situation poses threats to human well-being and rights.

Erosion of trust due to limited focus on impact

Recent studies have shed light on critical gaps in ethical AI declarations, revealing a broad spectrum of overlooked issues. Bolte et al. (2022), in their review of the declarations for ethical AI, state that “little has been said, in these guidelines, about the environmental sustainability of AI/ML” (p. 1). Morley et al. (2019) highlight a crucial threat to the trust in AI systems, stemming from the lack of attention paid to the impact of these systems on user autonomy. This issue stems from the assumption in ML that prediction equates to decision-making without adequately researching how predictions translate into actionable decisions. In a similar vein, Hagendorff (2020), in his study about AI declarations, found that lack of attention given to questions related to labor rights, technological unemployment, the militarization of AI, the spread of disinformation, and the misuse/dual-use of AI technology. In parallel, Hagendorff (2020) found that the declarations lack discussion on long-term risks (e.g., artificial general intelligence safety and existential risks). Also, Corrêa et al. (2023) reviewed 200 guidelines and recommendations for AI governance and found only 1.5 % pay attention to long-term AI risks. In the context of AI declarations, this indicates a potential gap in addressing how AI predictions affect end-user autonomy and decision-making, which could undermine trust in AI systems. In addition, Hagendorff (2020) points out a significant omission in discussions about AI technology's sustainability, highlighting that the practices involved in producing and consuming AI can conflict with sustainability goals. This includes concerns like lithium mining, e-waste generation, the singular use of rare earth minerals, high energy consumption, and the employment of low-wage 'clickworkers' for tasks like data labeling and content moderation, as discussed in various sources.

Numerous instances demonstrate a limited focus on the impact of AI technologies. For example, AI systems have been capable of deducing personal attributes such as political preferences or sexual orientation from Facebook likes (Gibney 2018) or assessing mental states by analyzing Facebook posts (Goggin 2019). A notable case in early 2020 involved Clearview, a New York City-based firm, which developed an AI system capable of compiling

comprehensive individual profiles using publicly available internet data. This development sparked widespread concerns and fears about the potential erosion of privacy (Hill 2020). In a more recent example, The Italian Data Protection Authority banned ChatGPT, alleging OpenAI's insufficient legal grounds for data collection and processing and the lack of age verification to protect children from inappropriate responses (Goujard, 2023). While AI declarations, such as the EU Ethics Guidelines for Trustworthy AI, emphasize privacy as a critical principle, there are still instances where privacy violations occur in practice, as seen in the examples of AI systems inferring personal information from social media activities. Cloarec (2022) also explores the role of privacy controls within the Protection Motivation Theory (PMT) framework. PMT, originally developed by Rogers (1975) and later refined by Maddux and Rogers (1983), analyzes how individuals protect themselves from perceived threats by balancing the evaluation of the threat against their coping strategies. Privacy controls, which offer users a sense of control over their personal data on social networking sites, serve as an important cognitive mediator in this framework. The study shows that privacy controls significantly reduce the tendency to falsify information by enhancing users' self-efficacy and reducing their perceived vulnerability. These findings highlight the importance of implementing effective privacy controls to foster trust and ensure data integrity in digital interactions.

In sum, the limited focus on the actual impact of AI systems on users and society represents a significant threat to trust in AI ethics declarations. When these declarations fail to adequately address or emphasize AI technologies' real-world consequences and implications, it can lead to a disconnect between the intended ethical standards and their practical outcomes. This gap can erode public and stakeholder trust in AI systems, as users may perceive a lack of commitment to genuinely understanding and mitigating the potential negative impacts of AI. Moreover, the erosion of trust could have broader implications for AI. It could lead to resistance against AI-driven initiatives, increased regulatory scrutiny, and a general skepticism towards AI advancements.

Usability and practical implementation of ethical tools

Recent advancements in AI ethics have led to the development of various tools intended to guide the ethical design of AI systems. Notable among these are the European Commission's Human Brain Project's responsible research and innovation methodology (Bitsch et al., 2023), the field guide to human-centered (ideo.org, 2015), Involve and DeepMind's guidance on public

engagement in AI ethics (Involve & DeepMind 2019), and Impact Assessment of the Regulation on Artificial intelligence (EC, 2020). These tools represent significant efforts to operationalize ethical considerations in AI development. While these tools are promising, they are not yet "production-ready," requiring considerable effort from ML developers, thereby adding friction rather than easing the application of ethics in AI (Morley et al., 2019). This suggests a challenge for AI declarations in ensuring that ethical guidelines are practically implementable and that tools developed to enforce these ethics are user-friendly and effective in real-world scenarios. The insights from Attard-Frost et al. (2023) identify a notable weakness in the realm of AI ethics guidelines: the predominant focus on algorithmic decision-making aspects such as fairness, accountability, and transparency while mainly neglecting the ethical considerations of AI business practices and the political economies within which AI operates. Ray (2023) thoroughly examined the landscape surrounding evaluating conversational AI systems like ChatGPT, focusing on performance metrics and ethical implications. The study meticulously reviews a range of existing benchmarks, including General Language Understanding Evaluation¹⁷ (GLUE), SuperGLUE¹⁸, Stanford Question Answering Dataset¹⁹ (SQuAD), Conversational Question Answering²⁰ (CoQA), Persona-Chat²¹, Dialogue System Technology Challenges (DSTC), BIG-Bench²², Holistic Evaluation Of Language Models²³ (HELM), and Multilingual Multi-Domain Language Understanding²⁴ (MMLU), highlighting their respective strengths and weaknesses. It also scrutinizes standards set by entities like OpenAI, IEEE's Ethically Aligned Design, the Montreal Declaration, and the Partnership on AI's Tenets to assess their applicability to ChatGPT.

This omission points to a gap in the comprehensive coverage of AI ethics, suggesting the necessity for guidelines to broaden their scope. Addressing this shortfall involves integrating considerations of AI development and deployment's business and political contexts into the ethics guidelines, ensuring a holistic approach to AI governance encompassing the technical and socio-economic dimensions of AI's impact on society.

¹⁷ GLUE consists of nine tasks, including question-answering, sentiment analysis, and textual entailment.

¹⁸ SuperGLUE builds upon GLUE and includes more challenging tasks, pushing the boundaries of NLP models

¹⁹ provides passages and corresponding questions that require understanding of the passage to answer

²⁰ designed to evaluate models on their ability to handle conversational question answering, which requires understanding of the conversation history

²¹ It consists of a dataset of over 131,000 utterances where models are trained to engage in dialogue while adhering to predefined personas

²² It covers a wide range of tasks such as text classification, summarization, translation, question answering, and more.

²³ It focuses on designing challenging tasks that require deep comprehension, including linguistic nuances, common sense, and logical reasoning

²⁴ It focuses on evaluating language models across multiple languages and domain

Ethical AI in a competitive world: challenges and pressures

The view that AI development should be primarily driven by market forces with minimal government intervention, as argued by some (Butterworth, 2018; European Commission, 2020; O'Sullivan et al., 2019), reflects a broader trend in the industry toward competitiveness and innovation. The concern expressed by Claessens et al. (2018) that stringent regulations could hinder AI innovation points to a significant threat: the possibility that in trying to harness and regulate AI, we might inadvertently restrict its potential and competitiveness in the global market. Within this regard, the White House's draft guidance for regulating AI applications, suggesting a cautious approach to regulation not to hamper AI innovation (Calo, 2017), underscores this tension. This guidance reflects the competitive pressure within the industry to continue advancing and innovating. Truby et al. (2020) suggest that measures like transparency, accountability, and fairness might promote AI growth and innovation, further highlighting the competitive aspect. It implies that the industry seeks to self-regulate and innovate responsibly, balancing ethical considerations with competitive advancement.

Arvan's 2018 'moral-semantic trilemma' posits significant challenges in programming ethical AI. It identifies three potential pitfalls: too semantically strict (where ethical guidelines are overly rigid and inflexible, failing to accommodate the nuanced nature of real-world scenarios), too semantically flexible (where guidelines are so broad or vague that they provide little actionable guidance or standards), and overly unpredictable (where the outcomes are not consistent or reliable). Considering this trilemma in the context of AI ethics declarations reveals a significant threat. First, it underscores the difficulty in creating AI ethics declarations that are neither too rigid nor too vague and yield predictable and reliable outcomes in diverse scenarios. Second, The complexity of ethical programming poses a challenge for AI developers in adopting and implementing AI ethics principles, as impractical or vague guidelines may hinder their integration into AI development processes. Third, the difficulty in resolving the trilemma can lead to skepticism or reluctance among developers and stakeholders regarding the feasibility and reliability of ethical AI, potentially affecting the broader adoption and trust in AI systems guided by these declarations. Hence, this represents a threat to AI declarations where pursuing ethical AI might be considered a competitive disadvantage. This short-term view can significantly impede the development of ethically aligned AI systems.

Public acceptance, investment, and research challenges

The public generally has a positive outlook on the advent of intelligent technologies, which offer new opportunities in various aspects of life (Kassens-Noor et al., 2021). Nevertheless, growing concerns exist about the misuse of Intelligent Autonomous Systems (IAS) and their potential negative consequences. A case in light of this is autonomous vehicles, which still face technical challenges, such as their limited ability to predict and react to uncommon traffic situations (Lokshina & Lanting, 2020). These technical limitations have raised concerns about user trust, particularly regarding the safety of riding in fully autonomous vehicles (Giffi et al., 2018). Winfield and Jirotko (2018) suggest that while there's no exact 'formula' for establishing trust in technology, people commonly trust technologies when they are beneficial, safe, and effectively regulated. Advancing this argument, Lokshina and her colleagues (2022) denote five potential indicators of users' trusting beliefs in intelligent machines: calculative cost/benefit perception, perceived societal benefits, organizational situation normality, and technology structural assurance.

Morley (2022) also warns of potential outcomes like a public backlash against algorithmic systems (Cookson 2018), reduced investor and research funder confidence, and an overall decline in interest in AI development. This situation mirrors the fallout from the Lighthill Report 1973, which criticized AI research for not meeting its hyped expectations. This perspective threatens AI ethics declarations where a lack of incentives for ethical AI development could lead to wider repercussions, including public mistrust, reduced funding, and a decline in AI research and adoption.

Geopolitical tensions

Geopolitical tensions, particularly those between major global powers, threaten the collaborative structure required for efficient AI regulation. Bryson and Malikova (2021) explore the advent of a new AI 'cold war' plot, primarily focusing on the United States and China. This narrative implies a binary competition in which other nations are relegated to peripheral roles, either aligning with one of these powers or being eclipsed by their dominant position. Although legitimate security concerns and uncertainties about digital technology underpin some components of this narrative, the extreme binary portrayal lacks a substantial factual basis. In a related aspect, the European Union seeks to set itself apart from the United States and China by focusing on an approach that is ethical, human-centric, and grounded in values, as exemplified

by the European Commission (2018b): “There is strong global competition on AI among the USA, China, and Europe. The USA leads now, but China is catching up fast and aims to lead by 2030. For the EU, it is not so much a question of winning or losing a race but of finding a way to embrace the opportunities offered by AI in a way that is human-centered, ethical, secure, and true to our core values” (pp. 12-13). The digital revolution, as pointed out by Ali (2019), also brings social challenges, such as amplifying prejudice and perpetuating existing power structures. The focus on competition and national superiority exacerbates these issues in a geopolitically charged atmosphere, preceding the need for ethical considerations and inclusive governance in AI development. Thomann (2023) underscores the escalating geopolitical fragmentation and the role of AI in these dynamics. According to the author, the growing use of AI as a geopolitical tool amidst increasing global power disparities and hybrid warfare, including psychological warfare, raises questions about whether current approaches are robust enough to address the challenges of malicious AI use and threats to international psychological security against the backdrop of intense power rivalries. In this interpretation, the potential weaponization of AI in geopolitical strategies highlights the urgency for AI ethics declarations to address the ethical and economic aspects of AI and engage with the complex geopolitical realities where AI might serve purposes that contradict ethical standards.

Bächle and Bareis's (2022) study focuses on the intricate geopolitical processes underlying Autonomous Weapon Systems (AWS), highlighting how these technologies link to state power ambitions. The same authors underline that AWS's vagueness and diverse interpretations extend beyond intellectual or technological difficulties; geopolitical movements profoundly root these issues, as nations and preeminent powers construct varied meanings around AWS in their political discourses. These geopolitical dynamics, which include competing interpretations, national interests, and the strategic use of AI technologies, have significant consequences for AI ethical pronouncements. They complicate the development of internationally agreed standards and laws for AWS and have a broader impact on the effectiveness and apparent neutrality of AI ethical pronouncements in international forums.

The study by Roberts et al. (2023) investigates the influence of two geopolitical trends — digital sovereignty and digital expansionism — on global AI governance. These trends represent different state approaches to controlling digital technologies, with sovereignty focusing on increased state control and expansionism on extending state influence in digital domains. The analysis, which includes case studies of China, the European Union, and the

United States, suggests that the extraterritorial policies and competitive narratives, particularly from the U.S., might hinder comprehensive global cooperation in AI governance. However, the study also identifies potential areas for alignment, such as data governance and technical standards, which might help to build trust and cooperation in international organizations like the G20 or the United Nations. This implies that while geopolitical dynamics present challenges, they also offer opportunities for collaboration in establishing AI ethics and governance frameworks.

3.4 Discussion

This chapter enriches the dialogue surrounding ethical AI by exploring a spectrum of principles, ranging from widely acknowledged to those less consistently highlighted. The quantitative evaluation identifies dominant core principles such as transparency, fairness, accountability, and privacy. Yet, it also exposes notable deficiencies, particularly in the realms of legal compliance and managerial oversight, which are pivotal for the practical implementation and reinforcement of these principles. For example, legal compliance is referenced in merely 15% of the declarations, and managerial aspects are notably absent. To mitigate these gaps, adopting the responsible innovation framework proposed by Stilgoe et al. (2013) — which prioritizes anticipation, reflexivity, inclusion, and responsiveness — stands out as a valuable approach. This framework facilitates the proactive identification of future ethical dilemmas and encourages a critical evaluation of existing practices, steering AI developments toward responsible innovation. It promotes the engagement of a broad range of stakeholders (including users and consumers) in the decision-making process, such as their involvement in board meetings. Moreover, a focus on responsiveness enables the dynamic adjustment of ethical principles in response to emerging challenges, such as addressing the proliferation of disinformation through generative AI technologies and fostering an AI governance model that is both flexible and effective in the face of technological advancements.

The integrative review identifies opportunities and limitations within the existing literature on ethical AI. The limitations adhere to some of the literature, such as Hagendorff (2020), which illuminates six critical areas that current AI ethics tools and frameworks frequently overlook. Firstly, an examination of whether algorithmic decision-making systems outperform or fall short of human decision routines is absent. Secondly, the guidelines do not investigate the essential instances in which AI-driven decision-making could be superior or inferior to human choices. Thirdly, the environmental and social costs concealed by AI systems

remain unaddressed. Fourth, the ethical discussions also scarcely touch upon the ethical dilemmas akin to the trolley problem, as extensively discussed in literature such as Awad et al. (2018). Fifth, the conversation lacks mention of public-private partnerships or the implications of industry-funded research within the AI domain. Lastly, there's minimal consideration of how AI systems could be politically misused, including concerns like automated propaganda, bots, fake news, deep fakes, micro-targeting, election fraud, and related issues. Also, the absence of a unified ethical AI framework, evident in the diversity of principles and their application across specific disciplines, presents a notable challenge. By illuminating the overlooked areas in current AI ethics literature, this review prompts a critical reevaluation of existing guidelines and sparks the development of more comprehensive and robust ethical standards. It acts as a catalyst for ethical innovation, challenging entrenched paradigms and stimulating the creation of novel tools and frameworks adept at addressing the multifaceted challenges of AI. This review not only highlights the necessity of interdisciplinary collaboration between computer science, ethics, sociology, and environmental studies but also enriches the dialogue, fostering innovative solutions sensitive to AI's social, environmental, and political dimensions.

This chapter extends the dialogue on ethical AI by scrutinizing the ethical principles that underpin recent AI declarations and examining the internal and external factors that influence their efficacy. Diverging from prior literature, such as Hagendorff (2020) and Fjeld et al. (2020), this study sets itself apart through its broad scope, adopting a global viewpoint to provide an exhaustive analysis that spans varied geographical contexts. Additionally, this research distinguishes itself by thoroughly examining the opportunities and limitations of ethical AI declarations. These insights hold the potential to profoundly impact policymakers, the business community, and academia, offering nuanced perspectives on the integration and operationalization of ethical principles in AI development and implementation.

3.5 Policy implication

The insights derived from this investigation hold particular value for policymakers for several reasons. Firstly, by advocating for policies that prioritize transparency, they can bolster public confidence in AI technologies. Secondly, the study recommends formulating or enhancing AI-centric policies and regulations. Additionally, it underscores the importance of striving for international consensus on ethical AI standards and promoting governmental participation in global discussions to establish universally recognized ethical guidelines.

Furthermore, by emphasizing the need for diverse stakeholder participation, this study encourages policymakers to use more inclusive strategies in AI policy-making.

3.6 Managerial implication

These findings serve as a roadmap for the business sector to align AI development processes and product offerings with established ethical norms. This alignment fosters consumer trust by prioritizing inclusivity and positions businesses at a competitive advantage by addressing neglected ethical considerations and developing forward-looking strategies that guarantee sustainability in the long run. This review encourages managers to adopt interdisciplinary approaches in their AI projects. This can lead to more innovative and well-rounded solutions that address the multifaceted challenges of AI, integrating ethical considerations into every stage of AI development and deployment.

Also, the findings from have several critical managerial implications that can guide organizations in developing and implementing effective ethical AI practices. First, there is a clear need for managers to advocate for the establishment of standardized ethical frameworks that harmonize global standards and practices. Engaging with international bodies and industry consortia to contribute to the development of universal ethical guidelines is essential. Once these frameworks are developed, managers must implement them within their organizations to ensure consistency and alignment with global best practices.

Second, the divergence in definitions and terminology related to ethical AI principles poses a significant challenge. It is the responsibility of managers to ensure that their data and AI governance teams employ and foster the use of clear and consistent terminology for ethical AI principles within their organizations. This can be achieved by conducting training programs and creating internal lexicons to guarantee that all stakeholders possess a shared comprehension of critical terms and concepts, and guarantee an effective guide AI practices. Additionally, these definitions should be regularly updated to reflect technological advancements, such as the emergence of generative AI, which may modify the interpretation of principles like fairness and privacy.

Third, geopolitical tensions have the potential to influence the progress and use of AI by managers. It is crucial to be aware of the geopolitical consequences and create plans to minimize the potential dangers linked to international AI initiatives. To effectively navigate

these challenges, managers should consider including geopolitical analysts or experts in their consultancy teams or boards of directors. These specialists possess ability to comprehend and analyze geopolitical issues and their consequences for AI, thereby assisting enterprises in designing accurate strategies and reducing risks related to global AI implementations.

Four, the difficulty of converting conceptual principles of ethics into tangible and implementable standards. Managers are required to create comprehensive implementation strategies that translate these concepts into practical actions that can be carried out on a daily basis. This entails developing explicit policies, protocols, and performance indicators to monitor compliance with ethical norms.

Five, the erosion of trust resulting from an excessive focus on the consequences of AI highlights the necessity for increased transparency and accountability. Managers should enact robust transparency measures, such as open reporting of AI system performance and decision-making processes. It is also imperative to establish clear accountability structures that can promptly address and rectify any ethical breaches in order to maintain trust with stakeholders.

Sixth, the usability and practical implementation of ethical tools are critical for embedding ethical considerations effectively in AI systems. Organizations should allocate resources and implement pragmatic ethical instruments, such as fairness evaluation frameworks, bias identification algorithms, and explainability tools. It is imperative to provide employees appropriate training on the efficient utilization of these instruments. Additionally, aligning ethical AI practices with existing legal frameworks is crucial for ensuring compliance and operational integrity.

Finally, continuous improvement and regular training are essential to keep pace with the evolving landscape of AI ethics. Implementing continuous improvement processes for ethical AI practices and developing regular training programs will keep employees informed about the latest ethical standards, practices, and regulatory requirements.

Academics should leverage the identified gaps and opportunities to advance the discourse on ethical AI. This involves moving beyond purely theoretical debates to create and effectively implement practical ethical frameworks in real-world scenarios. Additionally, there's an opportunity to explore the role of AI in enhancing democratic processes and the potential for AI technologies to contribute to more equitable economic development by

reducing biases in decision-making processes and increasing access to information and services.

Lastly, this research suggests a novel avenue for scholars to investigate: integrating ethical AI principles with corporate social responsibility (CSR) and corporate digital responsibility (CDR) initiatives. This integration could offer a comprehensive approach to ethical considerations in AI, combining the technical aspects of AI ethics with broader social and environmental responsibilities. This perspective broadens the scope of ethical AI discussions and invites a multidisciplinary approach to developing technologically sound and socially beneficial solutions.

3.7 Limits and future research

The limitations of this thesis include its focus on a select group of 27 ethical AI declarations, which may restrict the broad applicability of the findings to the wider AI ethics discourse. Furthermore, the analysis concentrates solely on the content of these declarations without examining their real-world execution, limiting our ability to assess their actual impact and compliance with outlined ethical norms. Therefore, it becomes essential for future studies to delve beyond the principles of ethical AI, probing into the multifaceted nature of AI's responsible development and usage. Such inquiries should address the intricacies of ensuring AI's ethical deployment, formulating a more actionable and comprehensive framework for RAI.

Moreover, this chapter encourages future explorations into integrating AI ethics within organizational strategies and the potential for public engagement and discourse in shaping ethical AI frameworks. These avenues promise to enrich the understanding of ethical AI's role in society and its implementation in real-world contexts, offering insights that could significantly influence the development of technologies that are not only innovative but also ethical and socially responsible.

3.8 Conclusion

In the realm of ethical AI, this chapter investigates a comprehensive exploration through a dual-methodological lens, employing quantitative frequency analysis and an ILR. The frequency analysis reveals the landscape of ethical AI principles into three significant levels. Dominant Principles such as transparency, fairness, accountability, and privacy are identified,

boasting substantial mentions of 81%, 67%, 63%, and 56%, respectively. These principles underscore a widespread consensus on the core tenets essential for ethical AI development. Following closely are intermediate principles, including security, safety, and explainability, each receiving moderate attention with mentions around 48%, 44%, and 44%. These principles, while significant, highlight areas where further clarity and emphasis may be needed. On the other end of the spectrum, Areas with Less Emphasis emerge, notably data governance and disinformation, highlighted less frequently in the declarations, at 11% and 4%, suggesting potential gaps in the current ethical AI discourse and areas needing more clarity.

Identifying opportunities for AI declarations to address ethical dilemmas and promote social good contributes to the optimistic vision of AI as a force for positive societal change. It validates the potential of ethical declarations to maintain trust and leverage AI technologies for the benefit of society, providing strategic insights into harnessing these opportunities. Also, global participation and comprehensive coverage are identified as significant benefits of AI ethical declarations, underlining the importance of harmonizing global norms to ensure their efficacy.

Conversely, the investigation into weaknesses, such as the lack of a common framework for ethical AI, divergence in defining ethical principles, challenges in translating principles into practices, and issues integrating with existing legal frameworks, highlights critical areas for improvement. These findings point to the necessity for a more unified approach to ethical AI governance, clarifying ethical principles and their practical implementation. Other factors related to trust, practical implementation, competitive pressures, and geopolitical tensions highlight the external challenges facing the effective implementation of AI ethics declarations. This aspect of the study contributes to understanding the external landscape's complexity and the need for adaptive strategies to mitigate these threats.

Addressing the gaps and shortcomings uncovered in this research presents a crucial chance for policymakers and industry leaders to refine and implement SMART ethical AI principles—Specific, Measurable, Achievable, Relevant, and Time-Bound. Establishing and harmonizing global standards emerge as fundamental steps toward creating a cohesive ethical framework that encourages trust and international cooperation.

Chapter 4: An Integrative Theoretical Framework for Responsible Artificial Intelligence

Chapter 1: What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI? What are the risks of GAI and their impact on individual society organizational and environment?



Chapter 2: What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?



Chapter 3: What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?



Chapter 4: What are the essential dimensions of Responsible AI (RAI) and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?



Chapter 5: What are the specific generative AI risks? How do these GAI risks affect different aspects of society, business, and the environment?



Chapter 6: What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?

About this chapter

This chapter has been published in the **journal**:

Haidar, A. (2024). An integrative theoretical framework for responsible artificial intelligence. *International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)* (ABS 1) 13(1), 1-23. <http://dx.doi.org/10.4018/IJDSGBT.334844>

Presented at the **conference**:

Haidar, A. (2023). A framework for responsible AI. In *R&D Management Conference*. 19-21 June, Seville, Spain

4.1 Theoretical background

4.1.1 The need for AI governance within the company

In the age of digital technology, the rapid rise of AI across business landscapes is not just a monument to human skills and creativity but also a significant turning point requiring a rethinking of governance paradigms. As organizations navigate the complexities of integrating AI into their operations, there is a need for strong AI governance not only as a regulatory compliance checkbox but also as a strategic cornerstone critical to sustaining innovation, ensuring ethical behavior, and safeguarding stakeholder trust.

Incorporating AI holds promises like unprecedented efficiency, improved decision-making capabilities, and the ability to set new value chains and business models. However, this potential carries some risks. Several vital drivers power the leap into AI-driven business models: the velocity variety and volume of big data, the technological infrastructure, and the security challenges like its costs (Pencheva et al., 2020; Halim et al., 2023). Schneider (2022) outlines four primary obstacles in managing AI: the complexity and lack of transparency in AI's results, the potential for unpredictable outcomes that may defy management, biases in AI decision-making based on data, and the swift pace of technological progress. In their extensive review of the literature, Wirtz et al. (2022) identify six categories of AI-related risks: (1) technological, including risks related to the AI systems' foundational technology, data, and analysis; (2) informational and communicational, focusing on risks like the manipulation of information and communication pathways; (3) economic, concerning the possibility of AI-

driven firms abusing their market position; (4) social, which includes the risk of social biases being perpetuated by AI; (5) ethical, addressing concerns over AI's failure to emulate human virtues such as equity and responsibility; and (6) legal and regulatory, pointing out that intensified regulatory scrutiny could stifle AI innovation and progress.

A series of significant incidents involving renowned corporations and their brands have occurred as a direct consequence of the rising use of AI. These incidents accentuate the critical need for establishing comprehensive AI governance. Here are a few examples that have been recently examined:

1. Apple's Siri gender bias allegations: according to a 2021 UNESCO study, Siri and other voice assistants perpetuated gender bias by automatically switching to female voices (West et al., 2021). The study contended that by supporting the idea that women are weak and need assistance, these AI systems promoted harmful stereotypes. As a result, Apple and other businesses have started addressing gender prejudice by giving customers several voice options for their virtual assistants.
2. Incidents involving ChatGPT: The Italian Data Protection Authority accused OpenAI in March 2023 of not having a legitimate legal basis for gathering and using personal data to train ChatGPT. It further asserted that OpenAI does not adhere to the General Data Protection Regulation and does not have an age-verification mechanism to shield children from exposure to improper content. As a result, Italy prohibited the chatbot (Goujard, 2023).
3. Incidents involving Microsoft: users of Bing Chat noted that, during prolonged interactions, the chatbot not only tended to provide erroneous information but also unexpectedly demonstrated emotional responses, implying the establishment of an unintended personality-like feature in its responses (De Vynck et al., 2023).

4.1.2 Frameworks for responsible AI

The term 'governance,' as the political sciences have traditionally developed it, usually points to alternatives to formal government at the societal or state level (Stahl et al., 2021). Business research often describes the structures and processes within organizations, while the term 'regulation' applies to discussions at a higher level, according to Braithwaite & Drahos (2000). As AI evolves and integrates into various sectors of businesses, such as supply chains

and society, such as social cohesion, the demand for RAI has never been more robust. Responsible AI refers to

human-centered and ensures users' trust through ethical ways of decision-making. The decision-making must be fair, accountable, not biased, with good intentions, non-discriminating, and consistent with societal laws and norms. Responsible AI ensures, that automated decisions are explainable to users while always preserving users privacy through a secure implementation (Göllner et al., 2024, p. 46).

This shift towards responsibility aims to leverage AI's potential for positive societal effects and reduce its associated risks. Numerous frameworks have been put ahead and accepted by governments, scholars, and organizations worldwide to accomplish these goals. These frameworks share common objectives but approach them through different lenses and methodologies. Dignum (2019) categorizes the methods for integrating ethical reasoning into AI systems into three main groups. Top-down strategies apply universal moral principles to make specific decisions, using established ethical theories as a foundation within a computational setting to particular scenarios. On the other hand, bottom-up strategies develop broad guidelines based on specific instances, equipping the AI with enough examples of behavior in similar contexts to generalize and compile these insights. Hybrid strategies merge elements from both top-down and bottom-up methods, promoting moral contemplation crucial for making ethical decisions. Most current models adopt a principled stance, positioning sustainable AI within the cross-disciplinary realm of AI ethics, as highlighted by Larsson (2020). Schiff et al. (2020) have introduced criteria for an effective sustainable AI framework. This framework should be comprehensive, acknowledging AI's broad spectrum of ethical, social, and economic implications. It needs to be actionable, enabling the transformation of theoretical principles and objectives into tangible action plans. Moreover, it should be flexible to adjust to different AI technologies and organizational environments and be iterative, with its principles applicable across an AI's entire lifecycle. The framework aims to be instructive, offering straightforward guidance that allows users of average expertise to implement, personalize, and resolve issues in diverse situations. Crucially, it should be inclusive, engaging insights from a wide range of stakeholders, especially those affected by AI deployment, including members of the general public.

Jobin et al. (2019) conducted a study on 84 ethical frameworks, pinpointing 11 crucial principles for the development of RAI: transparency, justice and fairness, non-maleficence

(avoiding harm), responsibility, privacy, beneficence (doing good), freedom and autonomy, trust, dignity, sustainability, and solidarity. Similarly, Eitel-Porter (2021) advocates for ethical AI, highlighting five fundamental principles: fairness, accountability, transparency, explainability, and privacy. Following this approach, various organizations have articulated high-level principles for pursuing sustainable AI. For example, OpenAI declares its dedication to reducing harm, fostering trust, leading trust and safety initiatives, and continuously learning and refining its data modeling processes.

Recognizing that AI impacts a broad range of stakeholders, RAI frameworks increasingly emphasize participatory design and stakeholder engagement. This approach involves including diverse voices in developing and deploying AI systems, particularly those of marginalized and affected communities. By doing so, developers can better understand the social context of their technologies, anticipate potential harms, and design AI systems that serve the broadest possible spectrum of users and use cases.

4.1.3 Exploring the limitations of the principles-approach for AI

While the principles-based approach to AI governance provides a solid ethical framework, its implementation reveals numerous shortcomings that require further examination.

Varying interpretations and implementational challenges

The flexibility of ethical principles also introduces a significant challenge: their interpretation can significantly vary, complicating the translation into specific, actionable guidelines or standards. Floridi et al. (2018) highlight the difficulty in implementing these abstract principles in a consistent manner, which in turn complicates the process for businesses trying to determine the most appropriate framework to adopt (Qiang et al., 2023). This variation in interpretation muddles decision-making and hampers these principles' universal applicability and effectiveness.

Legal enforceability

Another critical limitation is the non-binding nature of these principles. As noted by Brundage et al. (2018), the absence of legal enforceability can significantly dilute the impact of these ethical guidelines. Without the backing of the law, adherence to these principles relies

largely on AI developers' goodwill and ethical commitment and deploying entities, which might not always align with societal expectations or the greater good.

Cultural and social contextual relevance

The assumption that ethical principles have universal applicability overlooks the diverse cultural and social contexts in which AI operates. Gunkel (2018) emphasizes the challenge of applying a one-size-fits-all set of principles across varied contexts, underscoring the importance of adapting AI governance to reflect cultural sensitivities and social norms.

Stakeholder representation in principle development

The development of ethical AI principles often involves a limited set of stakeholders, potentially omitting crucial perspectives from those most affected by AI's deployment. Diakopoulos et al. (2017) argue that the lack of diverse stakeholder input could lead to a governance framework that does not fully consider the breadth of societal impact, thereby limiting its effectiveness and fairness.

Insufficiency for ensuring ethical AI

Mittelstadt (2019) contends that principles alone are insufficient for guaranteeing ethical AI, particularly highlighting the challenges in achieving transparency and explainability within AI systems. The author further notes the absence of established methodologies for translating principles into practice outside academic settings, stressing the need for proven methods to bridge the gap between ethical aspiration and practical application.

Keeping pace with technological evolution

As AI technology evolves, so do its ethical and social implications. Principles established in one era may quickly become outdated as new challenges emerge. Jobin et al. (2019) point out that the principles that may have once been adequate can fall short in addressing the contemporary ethical and social issues accompanying AI technology advancement.

4.2 Proposed integrative theoretical framework

Establishing core principles (e.g., fairness, safety, etc.) captures the technical aspects essential for the effective implementation of ethical AI (Eitel-Porter, 2021). Building upon the insights from examining limitations in current frameworks, we argue that while principles provide a valuable foundation for crafting guidelines for RAI, we should supplement them with other dimensions, as shown in Figure 22.

The second dimension pertains to implementing responsible innovation management (RIM) principles. The third dimension concerns legislative aspects, concentrating on AI-related laws to guarantee safe outcomes. Lastly, the fourth dimension delineates sustainable AI impact pathways, ensuring that AI deployment aligns with the sustainable development goals (SDGs) and avoids harm to the environment, economy, or society.

The rationale behind selecting these four dimensions—technical, legal, responsible innovation management, and impact on SDGs—derives from their alignment with the prevailing discourse and research on RAI. A thorough examination of pertinent literature and existing theoretical frameworks demonstrates that these dimensions encapsulate the most critical aspects of RAI while remaining inclusive and manageable. Moreover, the framework mirrors the interdisciplinary nature of AI, as each dimension originates from distinct research domains across various disciplines, rendering the framework inherently multidisciplinary. This interdisciplinary approach facilitates a more holistic understanding of RAI, integrating diverse viewpoints and insights to address the intricate challenges associated with AI development and deployment. Furthermore, the integration of these dimensions within the framework prevents the introduction of unnecessary complexity or redundancy.

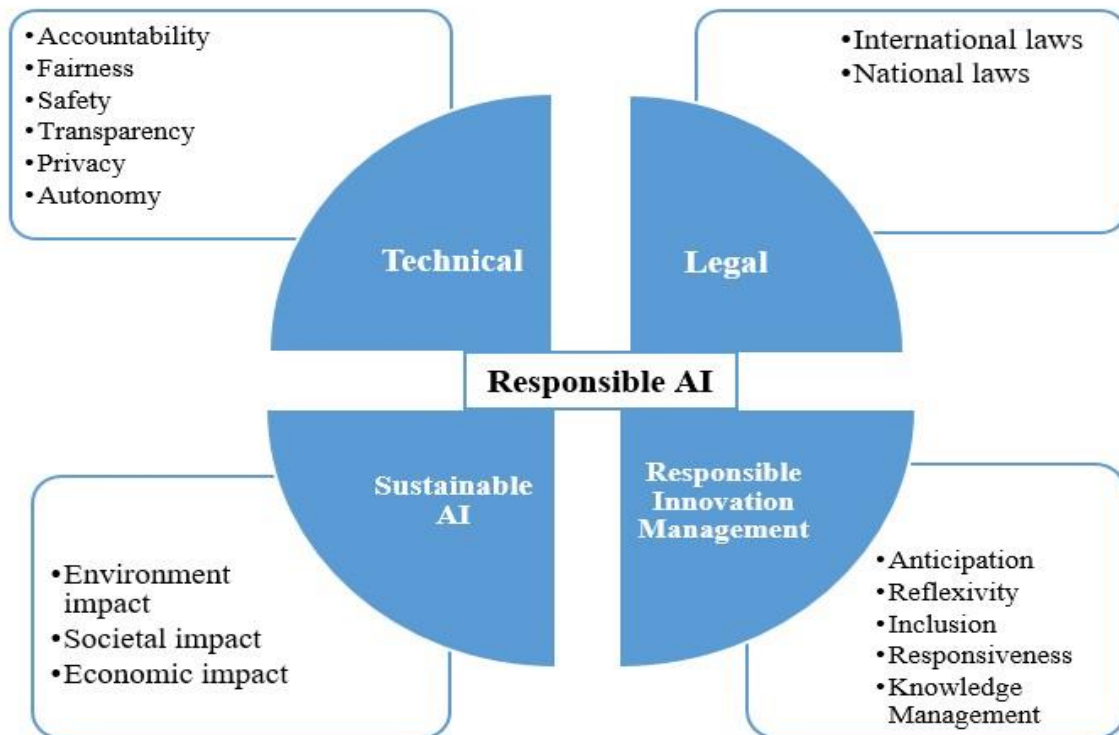


Figure 22 : Theoretical framework for responsible AI (source: author)

4.2.1 Technical dimension

To fully address the complex technical challenges of machine learning, it is crucial to consider the process at the core of all AI research: data collection. Recently, there has been much excitement over applying machine learning techniques to a broad range of data types, including web data, images, and bioinformatics. While there is much potential in these newer data types, the vast majority of machine learning still revolves around the analysis of tabular data and data for which a measure of feature relevance is assumed (for example, in time series prediction). The dominant technique at present for learning from tabular data is supervised learning. In supervised learning, an agent is presented with example inputs and their desired outputs, given a measure of success in predicting the output from the input via a mapping function (Li et al., 2023). In such a scenario, the goal is to learn a predictive rule that can then be applied to unseen input instances. This process proves quite potent, often requiring minimal assumptions about the underlying data generation process. However, this technique's effectiveness depends greatly on the quality of the provided data and the ease of manipulating it.

An obvious solution to the labeled data shortage is to develop algorithms that can learn from unlabelled data or label data effectively (Gomes et al., 2022). Despite notable advancements, especially in semi-supervised learning, data labeling remains the most resource-efficient approach for a given learning task. Self-learned labels will always be noisy, and a human with a clear understanding of the task will always be able to label the data more effectively. This is a consequence of the “No Free Lunch” theorem applied to machine learning: learning without a priori assumptions about the data is impossible (Wolpert, 2021). Obtaining labeled data is a finite and expensive endeavor (Man & Cahl, 2022). Often, human labeling is necessary, demanding significant investments in both time and finances. For instance, accumulating a labeled protein dataset took researchers many years (Aggarwal et al., 2021). Maximizing the utility of existing labeled data also entails avoiding redundant efforts. If a research group has already labeled a specific dataset, sharing it rather than duplicating the labeling process is preferable. Yet, collaboration is not commonplace in today’s fiercely competitive landscape, where publication metrics define success.

In summary, data, often described as the new oil (The Economist, 2017), are essential raw materials for generating information (Zins, 2007). As Boar (1997) defined, IT encompasses essential technologies for handling these invaluable resources—particularly in collecting, storing, transforming, and disseminating information. AI, a pivotal component of IT, leverages these data to automate complex tasks and generate actionable insights. Nevertheless, the intricate relationship between AI and data also presents significant technical challenges, such as overfitting and data leakage, which are critical for assessing the performance of AI models, as evidenced in Table 14.

Issue/Challenge	Description	References
Data Quality Issues	Generation of noisy, incorrect, or inconsistent data; calibration issues; poor sensor nodes quality; environmental effects.	Alwan et al. (2022); Bosu & MacDonnell (2019)
Sampling Bias	Sampling bias refers to the systematic error that arises when the sample collected is not representative of the entire population, potentially leading to inaccurate conclusions and misleading results.	Wang & Cheng, (2020); Vetter & Mascha (2017)
Data Drift	Data drift refers to the alteration or change in the input data distribution over time, which may affect the performance of predictive models.	Alsuwatet al. (2023); Sethi & Kantardzic (2017)
Quantization Error	A systematic error resulting from the difference between the continuous input value and its quantized output, and it is like round-off and truncation errors.	Colagrossi et al. (2023); Zhang et al. (2023)
Feature Redundancy	The inclusion of redundant or irrelevant features in the dataset reducing model generalization.	Lin et al. (2015); Tabakhi & Moradi (2015)
Data Leakage	Unauthorized access and exposure of sensitive information.	Alneyadi et al. (2016); Wang et al. (2019)
Memory Management Issues	Challenges related to efficiently managing memory usage during model training and inference, particularly critical when dealing with large models or datasets.	Cui et al. (2016); Zhang et al. (2015)
Overfitting	Overfitting is primarily a technical problem, arising when a model learns the training data too well, capturing its noise as if it were a real pattern, which leads to poor generalization to unseen data.	Burnham & Anderson (2002); Mutasa et al. (2020)
Underfitting	Underfitting occurs when the model is too simplistic to capture the underlying structure of the data.	Jabbar & Khan (2014); Wagner et al. (2021)

Tableau 14 : Key Technical Issues Faced in AI Implementation (source: author)

In recent years, researchers like Jobin and colleagues (2019), Floridi and Cowls (2022), and Royakkers and his team (2018) have studied principles-based frameworks to promote AI that benefits society. Jobin's team found that no single principle was present in all AI guidelines, but transparency, justice, fairness, non-maleficence, responsibility, and privacy were common in more than half. Similarly, a thorough review of the ethics in technology literature showed recurring themes like privacy, security, autonomy, justice, human dignity, control over technology, and power balance (Royakkers et al., 2018). A key reason for stressing principles like transparency and accountability in AI, similar to those in other IT areas, is their demonstrated advantages in various digital applications, such as e-governance. Yimer (2021) noted, "Once transparency is established, accountability will control government officials."

This chapter analyzed ten AI-related statements from 2017 to 2021 using a frequency test to determine AI's most common ethical considerations. Principles mentioned in at least half of these declarations were chosen for focus, reflecting their prevalence and significance. Setting this 50% threshold helped strike a balance between highlighting the most critical principles and excluding those less agreed upon or discussed. This method enabled a more streamlined and focused review of the core principles of AI. These principles fall into two primary categories: (1) their added value or impact on sustainability and (2) the technical requirements necessary for sustainable AI development. The latter category covers the technical aspects, focusing on accountability, privacy, transparency, safety, fairness, and autonomy.

4.2.2 Responsible innovation dimension

The concept of responsible innovation (RI) calls for firms to transform societal members' values and behaviors towards socio-ethical issues (Ceicyte & Petraite, 2018). While responsible innovation is not new, “the major novelty and practical relevance of responsible innovation is in integrating existing approaches and in making an explicit link between innovation and responsibility” (Genus & Iskandarova, 2018). the concept is interpretively flexible, and there are competing narratives. Burget et al. (2017) reviewed the literature on responsible innovation. They observed that policymakers and scientists defined and conceptualized RI differently, concluding that: “Responsible Innovation is essentially an attempt to govern research and innovation to include all the stakeholders and the public in the early stages of research and development. The inclusion of different actors and the public is, in turn, meant to increase the possibilities to anticipate and discern how research and innovation can or may benefit society as well as prevent any negative consequences from happening” (p. 15).

RI is introduced top-down by policymakers and scientists (Zwart et al., 2014). However, the problem with the current concept of responsible innovation is that researchers and policymakers focus primarily on responsible science and technological development (Lettice et al., 2013) without differentiating between research, development, and commercialization (Pellé and Reber, 2017). This poses critical challenges for implementing RI in the business context. First, focusing on science and technological development indicates a narrow view of innovation, as other types of innovation are not considered, such as social innovations (Blok et al., 2015). Second, commercialization is an essential stage in the innovation process (Choi and

Majumdar, 2014). RI is an emerging concept, and documentation of its implementation in the business context is still scarce. However, documentation about related social and sustainable innovation investigations in business is more common (Caulier-Grice et al., 2012). Like RI, social and sustainable innovation also addresses overarching societal challenges. Furthermore, they require the involvement of multiple stakeholders, and they consider social and environmental impact in addition to economic impact as desirable innovation outcomes. The scientific documentation of social and sustainable innovation in the business context can be informative for conceptualizing and understanding responsible innovation in the business context.

While various models for responsible innovation are discussed in the literature (for example, Burget et al., 2017), the framework developed by Stilgoe et al. (2013) stands out as one of the most influential. This framework includes four key dimensions: anticipation, reflexivity, inclusion and deliberation, and responsiveness. Additionally, “knowledge management,” a concept introduced by Lubberink et al. (2017, p. 721), is incorporated into this model, with further details provided. Table 15 explains the five dimensions.

It’s crucial to explore practical applications of these RI dimensions to move beyond merely theoretical discussions. Papagiannidis et al. (2023) offer practical insights, recommending that businesses adopt new AI strategies to stay competitive and enhance efficiency. These strategies align with the RI dimensions. For example, in the anticipation phase, enterprises assess the impact of AI on employees’ roles and responsibilities, especially as AI tends to automate tasks that are typically mundane and disliked by staff. During the reflexivity phase, organizations should establish clear policies that include continuous critical evaluations of AI systems’ effectiveness and ethical considerations. In the inclusion phase, management needs to develop transparent policies and ensure all employees communicate and understand these. Responsiveness involves using dashboards to interact with people and AI systems, facilitating swift organizational adjustments to new ideas and challenges. Lastly, in knowledge management, dashboards are tools for managing information, monitoring key performance indicators and metrics, and integrating new knowledge into the innovation process.

The following summarizes anticipation:

- Firms are required to include all stakeholders and the public to anticipate the potential problems and assess the available alternatives (Wickson et al., 2014; Burget et al., 2013)

- Firms must fully engage in multiple activities to enhance their understanding (e.g., Social trends, market trends, technological development, etc.) (Gazilusoy et al., 2013). An instance argues they can use a green data classification strategy (AGDCS) so the company knows in what data center location they should store their data in cold or hot regions.
- Usually, there are two ways a company can generate innovative ideas: first, the traditional way, using a pen and pencil, brainstorming activities, and idea boxes (Bocken et al., 2014). More modern and innovative way (Berker, 2010), collaborative business modeling, bayesian formalism for digital image steganalysis (Ziou & Jafari, 2014), and Ashby's Law of Requisite Variety and complexity theory (Espejo, 2017).

The second element is reflexivity, and it can be described as:

- A company's value system and beliefs can significantly impact innovation (Lubberkin et al., 2017; Larson, 2000; Dana et al., 2021).
- Company amount of knowledge can be a barrier to innovation, so as much as it has a specific team and leadership with higher expertise and experience, it will positively contribute to innovation (Lubberkin et al., 2017; Lampikosoki et al., 2014).
- Early inclusion of stakeholders can also enhance reflexivity (Wickson, 2014).

The third element is inclusion and deliberation, illustrated as:

- Academia-industry research and development collaboration and government funding influence innovation efficiency.
- The government's role encompasses three key areas: (1) supporting private sector research and development; (2) funding federal laboratory research activities and the efficient transfer of knowledge to the private sector; and (3) promoting industry-university research and development collaboration (Leyden & Link, 2022).
- The involvement of a diversity of stakeholders during different stages of the innovation process

The fourth element is Responsiveness, summarized by:

- Responding to new knowledge as this emerges and to emerging perspectives, views, and norms
- It implies that socially or ethically desirable values have to shape the evolution and use of technology and research (Pellé and Reber, 2013).
- safeguard achieving the desired innovation outcomes that address grand challenges and/or prevent detrimental effects (Lubberkin, 2017).

The final element for RI is Knowledge Management

It covers actions to overcome practical knowledge gaps that can arise with innovation: for instance, creating knowledge within the firm, creating knowledge with other external actors, or obtaining knowledge from external sources and subsequently integrating it into the innovation process (Lubberink et al.2017).

RI Dimensions	Key elements	References	Tools	References
Anticipation	Identifying potential impacts, risks, and uncertainties	Stilgoe et al. (2013)	Scenario thinking, value mapping, ideation for business modelling, translating vision into mission, game theory/perspective, and crowdsourcing are all valuable tools and approaches for businesses to strategize and make informed decisions.	Long et al. (2020)
	Engaging with a wide range of stakeholders	Stilgoe et al. (2013)		
	Understanding the dynamics influencing the formation of the innovation is essential for innovators.	Burget et al. (2017)		
Reflexivity	Reflecting on the values and assumptions that underpin the innovation is crucial, as it helps anticipate unintended consequences that may arise.	Stilgoe et al. (2013)	Formal evaluations, third party critical appraisal, informal (self) assessment culture, knowledge-concept-process mechanisms, and empowered and open communication are all important components of effective organizational learning and development.	Long et al. (2020)
	Considering the social, cultural, and political contexts in which the innovation is developed and used	Stilgoe et al. (2013)		
	Involving stakeholders early in the process is crucial for successful outcomes.	Wickson & Carew (2014)		
Inclusion	Ensuring that the innovation is developed and used in a way that is inclusive and participatory	Stilgoe et al. (2013)	Crowdsourcing via software, living lab structures, collaborative business modelling, partnerships, consultancy of experts, focus groups, and integrating views and opinions are all valuable methods for gathering diverse perspectives and generating innovative ideas in the context of business innovation and development.	Long et al. (2020)
	Engaging with diverse stakeholders and considering their perspectives and needs throughout the innovation process	Stilgoe et al. (2013)		
	Compelling of stakeholders for the purpose of substantively better decision-making and mutual learning	Wickson & Carew (2014)		
Responsiveness	Developing flexible and adaptable governance arrangements that can respond to changing circumstances and emerging ethical and social concerns	Owen et al. (2013); Lubberink et al. (2017).	Customizing or mainstreaming, preventing organizational inertia, adjusting/withdrawing innovation, and monitoring the external environment post-introduction of innovation are all important strategies for managing the implementation of innovative ideas and ensuring their success in the marketplace.	Long et al. (2020)
	Engaging with a wide range of stakeholders to ensure that potential issues are identified and addressed in a timely and effective manner	Stilgoe et al. (2013)		
Knowledge Management	Managing the various types of knowledge that are relevant to the innovation process, including both explicit and implicit knowledge	Owen et al. (2013); von Schomberg (2012)	Learning management systems, document management systems, web content management systems, decision support systems, knowledge repositories	Maican & Lixandriou (2016); Gupta et al. (2022); Gottschalk et al. (2023)
	Ensuring that the best available evidence informs the innovation process, like engaging in different activities to obtain the necessary knowledge, and that stakeholders are able to make informed decisions about the development and use of innovations	Lubberink et al. (2017)		

Tableau 15 : Responsible innovation dimensions

4.2.3 Legal dimension

Recent developments in AI, cloud computing, 5G, and Web3 have increased the need for either enhanced regulatory measures, including laws, regulations, and antitrust actions, or improved international data standards, data architecture, and data interoperability through better interfaces (Dąbrowska et al., 2022). Halder et al. (2022) highlighted how legal policies influence organizational privacy practices, pointing to legal frameworks' crucial role in guiding RAI governance. The specific legal frameworks applied to AI will vary depending on the context of AI applications and whether they adhere to national or international laws. For example, growing data privacy and protection concerns have spurred global laws and regulations. AI legislation and regulations fall into three major categories: governance of autonomous intelligence systems, responsibility and accountability for the systems, and privacy and safety concerns (Wirtz et al., 2018). The development of public sector strategies for AI management and regulation has become increasingly important at the local, national and international levels (Library of Congress, 2019; European Commission, 2020) and in a wide range of fields, including public service management (Ojo et al. 2019), law enforcement, the financial sector (Bredt, 2019), robotics (Iphofen and Kritikos, 2019), the military (Chelioudakis, 2017), and international law (Maas, 2019). Given these risks, there is growing pressure for algorithms to become transparent. From a legal perspective, transparency is relevant in several ways: clarifying how self-learning systems work can clarify liability issues more quickly and allow people to pursue their claims more straightforwardly. This will become an essential issue in the case of accidents involving autonomous vehicles. From the data protection law perspective, transparency is critical in putting "informational self-determination" into practice, particularly algorithmic decisions. Since early 2016, many national, regional, and international authorities have begun adopting AI strategies, action plans, and policy papers (Berryhill et al., 2019).

Table 3 showcases fundamental laws and highlights differences in AI regulations between the European Union (EU) and the United States (US). Our study used descriptive and analytical methods to compare AI legal frameworks between the EU and the US. We detailed specific laws and regulations in each region (descriptive comparison) and analyzed their effects on AI practices (analytical comparison). Moreover, we contrasted the EU's centralized legislative approach with the US's decentralized system, discussing each approach's unique challenges and advantages.

The EU has taken a proactive approach to AI regulation by establishing a thorough regulatory framework prioritizing data protection and RAI development. The General Data Protection Regulation (GDPR) is a cornerstone legislation that imposes strict guidelines on data collection, processing, and usage, ensuring that individuals have control over their data. Building on the GDPR, the EU further enhanced its regulatory stance by introducing the Data Governance Act (2021), aimed at improving data governance, facilitating data sharing, and fostering the growth of RAI. In June 2023, the EU updated this act to include a broader range of AI systems and adopted a more rigorous risk-based regulatory approach. This revision also expanded the requirements for transparency and accountability and strengthened the anti-discrimination measures, establishing a more stringent standard for AI practices within the EU. Although these regulations do not solely focus on AI, they are crucial for the development and deployment of AI technologies. The Digital Services Act regulates online platforms, requiring them to conduct risk assessments, enhance transparency, and manage illegal content. This will significantly affect AI-driven platforms used for tasks like content moderation or targeted advertising. The Digital Markets Act (DMA) ensures fair competition in the digital marketplace by setting rules for major online platforms, significantly affecting large AI corporations such as Google and Meta, classified as gatekeepers under this act.

In contrast, AI regulation in the United States consists of diverse state laws and regulations without a unified federal framework. This decentralized approach allows states the flexibility to experiment with various regulatory strategies tailored to local needs. However, this can also lead to uncertainties for businesses operating across multiple states and complicate the understanding of consumer rights. Table 16 includes examples of differing laws across various US states.

EU Regulation (law/description)		US Regulation (states/law/description)		
General Data Protection Regulation (GDPR) (2016)	Comprehensive data protection law designed to give individuals control over their personal data and simplify the regulatory environment for international business.	California	California Consumer Privacy Act (2018)	Grants consumers more control over their personal data, requires businesses to be transparent about how they collect and use data, and establishes a private right of action for consumers.
Regulation on a European approach for Artificial Intelligence (AI Act) (2021)	The AI Act categorizes applications of AI into four levels of risk: unacceptable risk, high risk, limited risk and minimal or no risk.		The Bolstering Online Transparency Act (BOT) (2022)	Requires online businesses to disclose to users whether they use automated decision-making tools (ADTs) and to provide users with the opportunity to opt out of ADT-based decisions.
Digital Services Act (DSA) (2022)	Regulates online services, including social media platform, online marketplaces, and app stores.	Illinois	Illinois AI Video Interview Act (2021)	Regulates the use of AI-powered video interviews in the hiring process, requiring employers to notify applicants of AI use, obtain their consent, and provide them with access to their interview data.
Digital Markets Act (DMA) (2022)	Aims to ensure fair competition in the digital market by imposing obligations on large online platform gatekeepers.	Indiana	Bill 5 Consumer data protection (2023)	Prohibits businesses from collecting, using, or disclosing personal data without the consent of the individual, and establishes a privacy regulator to enforce the law.
Data Governance Act (2023)	Aims: (1) To increase the availability of data for reuse; (2) To promote the fair and responsible use of data; (3) To increase trust in data	Maine	Data Privacy and Protection Act, HP 1270 (2023)	Establishes a comprehensive framework for data privacy protection in the United States, including requirements for data minimization, data security, and data breach notification.
		Pennsylvania	HB49	The proposed registry would require businesses using artificial intelligence systems to provide their contact details, the AI system's purpose and type, as well as a consent statement for the information to be stored by the Department of State.

Tableau 16 : EU and US AI legislation

4.2.4 AI and sustainability

4.2.4.1 Definition of sustainable AI

Van Wynsberghe (2021) describes sustainable AI as a movement to enhance ecological integrity and social justice throughout the entire lifecycle of AI products—from idea generation and training to retuning, implementation, and governance (p. 213). In this context, Bolte et al. (2022) categorize sustainable AI into two dimensions: AI for sustainability, which entails using AI to support the United Nations’ sustainable development goals (SDGs), and the sustainability of AI, which considers the environmental impacts of developing and using AI technologies. Nishant et al. (2020) conducted a comprehensive literature review identifying challenges in the research on AI for sustainability. These include an overreliance on historical data in machine learning models, unpredictable human responses to AI-driven initiatives, heightened cybersecurity threats, adverse effects of AI technologies, and challenges in assessing the impact of these technologies. Sustainability is often seen through the triple bottom line lens (Elkington, 1998), which balances economic, environmental, and societal elements in line with the SDGs. Scholars like Yong et al. (2020) emphasize that these three components are equally crucial to the success of a business. This section explores how AI interacts with sustainability concepts.

4.2.4.2 Opportunities and challenges for AI for sustainability

Regarding sustainability, AI presents a unique set of opportunities and challenges. While AI has the potential to drive significant improvements in how we manage and conserve resources, it also brings forth complex challenges that require careful consideration, particularly in terms of environmental impact, social acceptance, and economic viability.

3.2.4.2.1 Environmental consequences of AI advancements and concerns

Environmental sustainability is “meeting the resource and services needs of current and future generations without compromising the health of the ecosystems” (Morelli, 2011). Related concepts and themes include low-impact transportation, sustainable agriculture, the conservation of environmental assets (such as biodiversity, water resources, energy use, renewable energy, raw materials such as food and minerals, and land use sustainability), and the management of waste and pollution (such as waste reduction, recycling, reuse, repair, using environmentally responsible material, pollution monitoring, and pollutant reduction) (Samiha,

2013; Pillemer et al., 2011). Many of these issues have been investigated in research on the use of AI for environmental sustainability (as shown in Table 17). For example, there has been a particular emphasis on using AI to increase biodiversity. Many ecologists use ARIES(Artificial Intelligence for Ecosystem Services) systems to simulate ecosystem services (Willock et al., 2018). Numerous ML models for data analysis are built into the program to help explain complicated, nonlinear relationships.

Given the direct effect of carbon footprints, several tools have been proposed to make the impact of training models more visible. These tools can be schematically divided into the following categories:

- Integrated tools, such as Experiment Impact Tracker²⁵ (Henderson et al., 2020), Carbon Tracker²⁶ (Anthony et al., 2020), and CodeCarbon²⁷, are Python packages that report measured energy consumption and the associated carbon footprint.
- Online tools, such as Green Algorithms²⁸ (Lannelongue et al., 2020) and ML CO2 Impact (Lacoste et al., 2019), require only a few parameters, such as the training duration, the material, and the location, but are less accurate.

The complex relationship between AI and its environmental impact is summarized in Table 17, which outlines AI's positive and negative consequences on the environment and SDGs.

Negative consequences

Sustainability research is intensifying its focus on creating and managing future scenarios, mainly driven by rapid environmental changes and societal shifts (Vervoort & Gupta, 2018), with the ambitious target of the Paris Climate Agreement to limit warming to 1.5° Celsius motivating further exploration into realistic and favorable climate futures (Granjou et al., 2017). Additionally, the concept that all data-reliant emerging technologies contribute to environmental impacts, such as increased carbon emissions and electricity consumption, is gaining attention. This impact encompasses the entire life cycle data (LCD) of technologies, including phases such as pre-construction, usage, and demolition, each potentially further broken down depending on the specific research focus (Khasreen et al., 2009). Some scholars, including Chen & Zhao (2012), have expanded LCD to include seven stages—data generation, transmission, use, sharing, storage, archiving, and destruction. The volume of data is exploding, and according to Masanet et al. (2020), data center workloads have increased six-fold in the

²⁵ <https://github.com/Breakend/experiment-impact-tracker>

²⁶ <https://github.com/lfwa/carbontracker>

²⁷ <https://codecarbon.io/>

²⁸ <http://www.green-algorithms.org/>

past decade. This is primarily attributable to expanding mobile and cloud-computing traffic and developing and adopting computer-intensive applications and digital technologies, such as AI and the Internet of Things (IoT). According to Statista data (Figure 23), the total volume of data created, captured, copied, and consumed worldwide by 2024 will surpass 149 zettabytes yearly, and much of it will be unstructured.

Most companies' current data center infrastructure will probably not be adequate to adapt to their future data center workloads, implying the need for more storage capacity (Shehabi et al., 2018). Gartner's 2022²⁹ Strategic Roadmap for Storage predicts a 23-fold increase in enterprise storage capacity by 2030, necessitating that Infrastructure and Operations (I&O) leaders re-evaluate data center technologies and implement energy conservation initiatives to reduce costs. In the training data phase, NLP algorithms have a substantial negative impact; for example, Strubell et al. (2019) found greenhouse gas emissions (GHG) equivalent to 300 flights between New York and San Francisco. The development of digital technologies such as blockchain and machine learning is expected to drive an increased demand for data center services. Bitcoin, the most notable proof-of-work blockchain example and most valuable cryptocurrency based on market capitalization consumed 105 terawatt hours (TWh) in 2021, 20 times more than in 2016 (IEA, 2022). In the data storage phase, more data constitutes more energy consumption by data centers. According to Jones (2018), data centers consume an estimated 200 TWh annually, more than the national energy consumption of some nations, such as Iran. Many governments and academic institutions consider this problem a matter of significant interest. The New Climate Economy report by The Global Commission on the Economy and Climate (includes seven participating countries: Colombia, Ethiopia, Indonesia, Norway, South Korea, Sweden, and the United Kingdom) stated that investments made in the 15 years after 2014 will determine the future of the planet's climate system. One of the most remarkable predictions regarding the future energy demands of IT is attributed to Andrae and Adler (2015), who have stated that data center electricity consumption will increase approximately 15-fold by 2030. The authors show that communication technology (CT) could use as much as 51% of global electricity, contributing to 23% of the globally released GHG emissions in 2030. Levitch (2024) cites a Gartner report stating that the annual energy cost for powering a single server rack in a U.S. data center can reach up to \$30,000, with storage accounting for approximately \$3,300 or 11% of this cost. Avom et al. (2020) have found a positive correlation between ICT penetration and carbon emissions in 21 sub-Saharan African

²⁹ <https://www.gartner.com/en/documents/4012543>

nations. Other studies show the same results in developed countries, such as China. Zhong et al. (2021) have found that the interaction effect of urbanization and Internet penetration positively and significantly impacts carbon emissions. The global acceleration of industrialization and urbanization is increasing the energy demand, leading to long-term global pressure to reduce carbon emissions (Franco et al., 2017). Khattak et al. (2020) analyzed the impact of technological innovation, economic growth, and renewable energy use on carbon emissions in the five major emerging countries of Brazil, Russia, India, China, and South Africa (BRICS countries) from 1980–2016. The findings showed that innovation activities have failed to reduce carbon emissions for all these countries except Brazil. Nerini et al. (2019) have shown how climate change threatens achieving 16 sustainable development goals (SDGs) but note that taking action to combat climate change may undermine 12 other SDGs.

Water consumption is another sustainability concern. Data centers are the largest and fastest-growing electricity and water consumers in the United States (Sverdlik, 2016). Water is used to cool servers in most data centers, as it is cheaper than electricity. Servers, which are essentially stacked computer hard drives, generate a significant amount of heat. These servers are housed in large, windowless buildings approximately the size of several football fields. According to various industry estimates, a midsize center uses between 80 million and 130 million gallons of water annually for cooling. This is equivalent to the amount of water used in three hospitals, cultivating 100 acres of almond trees, or maintaining two 18-hole golf courses annually, and this is only for computer chips (FitzGerald, 2015). Scientific research has identified considerable inefficiencies in our use of natural resources. According to a study by Marín-Beltrán et al. (2022), approximately 78% of the total water withdrawn is wasted or mismanaged, along with 31% of the energy produced, and 85% of ores and 26% of non-metallic minerals extracted. These figures highlight the significant challenges related to efficient resource use.

The digital industry, which includes the production of electronic equipment, produces 50% more greenhouse gas emissions than the aviation industry, with at least 50% of these emissions directly attributed to electronic equipment production (Ferreboeuf, 2019), which raises the issue of electronic waste (e-waste). According to Kastanaki and Giannis (2022), a total of 24.6 million feature phones and 6.6 million smartphones were generated as waste between 2010 and 2020, and it is anticipated that 2.3 million feature phones and 40.7 million smartphones will be generated as waste between 2021 and 2035. If not correctly handled, this will lead to severe waste of natural resources, human and animal health hazards, and increased harm to the ecological environment (Hussain & Mumtaz, 2014; Andrade et al., 2023).

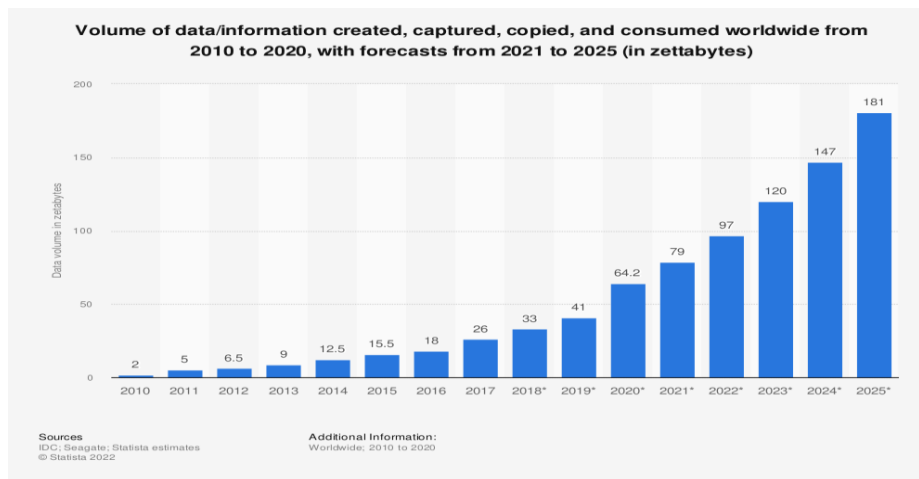


Figure 23 : Volume of data created, captured, copied, and consumed worldwide from 2010 to 2020 with forecasts for 2021 to 2025 (source: Statista)

Positive consequences

AI has the potential to contribute significantly to environmental protection, offering innovative solutions that can enhance resource management, conservation efforts, and climate change mitigation. The findings of Ahmad et al. (2021) indicate that AI is becoming a key enabler of a complex, new, and data-related energy industry, serving as a key “magic tool” to boost operational performance and efficiency in an increasingly competitive environment. Vinuesa et al. (2020) have investigated the relationship between AI and SDGs and concluded that AI can accomplish 134 targets across all the goals. According to Xu et al. (2018), neural networks and objective-oriented techniques can improve the classification of *vegetation* cover types based on satellite images, with the possibility of processing large amounts of images in a relatively short time. A major recent *agricultural* application of AI has involved data capture through devices such as movement sensors on, or even in, farm animals to enable reporting of their behaviors, health, and condition (Rutten et al., 2013). Combining aerial imagery with soil moisture-meter data from AI can significantly reduce the number of soil-moisture sensors needed to produce a high-resolution map, as Vasisht et al. (2017) demonstrated.

Furthermore, AI will support low-carbon energy systems. Bonire and Gbenga-Ilori (2021) proposed a model combining the source testing method, artificial neural network (ANN) technique, and the IoT system to efficiently predict and reduce pollutant gas emissions. AI can also help improve the health of ecosystems (Vinuesa et al., 2016). High energy efficiency from alternative sources is a top priority. One way to achieve this is by integrating renewable energy resources and AI, which can aid in developing a novel energy management system (Kumar et al., 2021). According to Hannan et al. (2021), AI can help meet 42 out of 169 targets of the SDGs. This is because of the ability of AI to improve the operation and efficiency of renewable

energy sources, reduce the operating costs of producing energy, and effectively reduce environmental impacts (Jha et al., 2017). In addition, the role of AI applications in renewable energy systems can include forecasting; emission reduction; cost minimization; robust and smooth control; high power quality without fluctuations, even when input is intermittent; expansion of novel technologies for optimal production using available natural resources; environmental awareness; enhanced energy management; distribution of energy; and energy delivery (Al-Shetwi et al., 2020, Ahmad et al., 2021).

Climate change has significantly impacted the quality of our planet's natural resources and has led to the deterioration of its ecological variety. Over a million species are threatened with extinction, highlighting the critical need for conservation strategies emphasizing biodiversity protection to retain its numerous benefits to human life. Silvestro et al. (2022) have created a methodology, conservation area prioritization through artificial intelligence (CAPTAIN), that quantifies the trade-off between the costs and benefits of area and biodiversity protection, allowing the exploration of multiple biodiversity metrics. Panigrahi et al. (2022) have developed a deep learning algorithm similar to YOLOv3 to create a real-time biodiversity map in response to the biodiversity issue. This algorithm could help to transform wildlife classification, poaching detection, species identification, and geo-localization of species where the entire mapped region can be transfigured to an activity heat map updated in real-time and easily scalable to a larger region. Different companies are following this trajectory; for instance, among 54 of the most eco-innovative Polish companies, Pichlak and Szromek (2021) found that most are developing technologies for biodiversity protection.

Integrating AI with satellite imagery can help identify land-use changes. For example, drones have been used for community-based forest monitoring (Paneque-Gálvez et al., 2014). In addition, mapping and quantifying ecosystem services with drones is an efficient way to inform site design and zoning, mainly when information is scarce, outdated, and based on coarse-resolution remote-sensing images (Jiménez López and Mulero-Pázmány, 2019). Research into AI's impact on land use has broadened to explore its role in urban planning, leading to the concept of "sustainable cities." Attaran et al. (2022) note that there is no widely accepted definition of this term despite numerous attempts to establish one. D'Aniello et al. (2020) describe a smart city as one that uses ICT to enhance the quality of life for its residents by providing better services and using resources efficiently. Šurdonja et al. (2020) add that a city becomes smart by combining social capital with traditional and modern ICT to promote sustainable economic growth and improve living standards.

Sustainable AI and the Environment with the corresponding SDG

Positive Impact	Negative Impact
Resource management (SDG 12) - Improved Energy Efficiency (Ahmad et al., 2021) - Optimized water usage (Alam et al., 2022; Tariq et al., 2021) - Enhanced waste management (Andeobu et al., 2022)	Energy consumption (SDG 7, SDG 12, SDG 13) - Energy-intensive model training - Power-consumption data centers (Andrae and Adler, 2015; Gartner, 2022) Environmental hazards from e-waste
Climate change mitigation (SDG 13) - Reduced greenhouse gas emissions (Bonire & Gbenga-Ilori, 2021) - Improved climate prediction - Enhanced disaster response	Resource depletion (SDG 7, SDG 12, SDG 15) - Mining of rare-earth element - Land degradation and habitat destruction (Kankara et al., 2013; Zhang et al., 2022)
Biodiversity conservation (SDG 14 & 15) - Monitoring species (Silvestro et al., 2022; Panigrahi et al., 2022) - Habitat restoration - Reduce marine Pollution (Keramitsoglou et al., 2006). - Improve the classification of vegetation (Xu et al., 2018) - Enhance Agriculture (Rutten et al., 2013; Vasisht et al., 2017) - Forest monitoring (Paneque-Gálvez et al., 2014; Jiménez López and Mulero-Pázmány, 2019)	E-waste generation (SDG 11, SDG 12) - Rapid AI hardware obsolescence (Contreras-Koterbay, 2020; Maas, 2019) - Disposal and recycling challenges (Yu et al., 2022) - Environmental hazards from e-waste (Pinto, 2008)
Pollution reduction (SDG 3) - Air quality monitoring (Kumar et al., 2021) - Improve renewable energy (Jha et al., 2017; Srivastava, 2020). - Water pollution control - Noise Pollution reduction	Carbon emissions (SDG 13) - Fossil fuel-based energy sources for AI (Chen et al., 2022; BP, 2021) - Coal fuel-based energy sources for AI (Cheng et al., 2017; - Increase in GHG (Strubell et al., 2019; Zhong et al., 2021; Nerini et al. (2019) - Due to the acceleration of urbanization (Sun and Dong, 2022; Zhang et al., 2017). - Indirect carbon emissions from AI
Sustainable Cities (SDG 11) - Urban sustainability (Yigitcanlar & Cugurullo, 2020). - Improve residents' quality of life (D'Aniello et al., 2020).	

Tableau 17 : Relationship between AI and environment with corresponding SDGs (Source: author based on literature)

3.2.4.2.2 Societal implications of AI: benefits and challenges

The section on AI's impact on society explores both its positive and negative effects, comprehensively discussing various aspects and their corresponding implications (illustrated in Table 18).

Negative impact

Philosophers like Bostrom (2014) have pointed out that in the utopian world of abundance, where machines and robots do all the work, humans may be reduced to having a second-rate status. Stephen Hawking famously warned that “[t]he development of full artificial intelligence could spell the end of the human race” (Cellan-Jones, 2014). Robots are becoming increasingly affordable and accessible, which raises the possibility that humans will become more dependent on them than on themselves on a larger scale (Wirtz et al., 2018). In a case study on Uber drivers, Rosenblat and Stark (2016) found that relying on algorithm-based decision-making decreases drivers' experiences of being in control and makes them feel negatively toward the company in question. The following debate will introduce the negative impacts of AI, such as fake news, privacy violation, human autonomy, discrimination, and E-justice, in different fields of society.

Recent evidence has indicated that the propagation of false information can mislead and potentially trigger severe societal complications (Fedeli, 2020; Zhang & Ghorbani, 2020). Fake information involves three main aspects: first, *political*, to spread misinformation about politicians and public policies; second, *anti-scientific*, which aims to disseminate false information to subjects, such as regarding aliens and vaccines; and third, *economical*, to manipulate stock markets and sell fraudulent products (Lazer et al., 2018). Societal acceptance of the information and news on social media is highly dependent on the verification and fact-checking features available on these platforms (Olan et al., 2022). This motivates scholars to study the impact of awareness and knowledge about fake news (Pundir et al., 2021). In the business context, the potential impact of fake news on businesses and consumers can be substantial (Berthon & Pitt, 2018). Di Domenico and colleagues (2021) investigated the effect of fake news from a consumer perspective, identifying five key themes: the creation and dissemination process using fake social media profiles, the characteristics of the spreading channels, the outcomes, fabricated legitimacy, and public attitudes towards fake news. Additionally, fake news represents a significant financial risk for businesses. For example, a false story in 2016 claimed Pepsi's Chief Executive Officer (CEO) Indra Nooyi telling Trump supporters to “take their business elsewhere” went viral, causing Pepsi's stock to drop by nearly 4% just before the 2016 US presidential election (Berthon & Pitt, 2018). In spite of this,

according to Allcott and Gentzkow (2017), the term “fake news” gained relevance following the 2016 US presidential elections. According to the same authors, 156 misleading news stories got just under 38 million shares on Facebook 3 months before the 2016 US presidency. Social media has, therefore, given consumers the power to build or kill a brand (Kohli et al., 2015). One of the tools used to create misinformation is deepfake technology powered by AI and ML, which offers automated procedures to create fake content that is becoming harder and harder for human observers to detect (Kietzmann et al., 2020). It can “merge, combine, replace and superimpose” (p. 1) various forms of media to quickly produce synthetic material that obscures the distinction between real and fake information (Maras & Alexandrou, 2019).

Privacy violation is another aspect that needs to be discussed, as modern civilization is highly dependent on computing systems, which touch all aspects of business, government, and individual lives. The negative impact of AI on privacy and security has become a growing concern in recent years. AI-driven technologies, such as facial recognition and data analytics, have raised concerns about the erosion of individual privacy (Brundage et al., 2018). The widespread use of these technologies has led to mass surveillance and the potential for misuse by both public and private entities (Zuboff, 2019). AI’s negative impact on privacy and security is further exacerbated by the potential misuse of AI for malicious purposes. Bad actors can exploit AI to create new and more potent cyberattacks, espionage, and disinformation campaigns (Brundage et al., 2018). For instance, deepfake technology, which uses AI to manipulate images and videos, has spread disinformation and caused reputational harm to individuals and organizations (Chesney & Citron, 2019). Fortes et al. (2017) provides empirical evidence showing that privacy concerns significantly impact consumers' intention to make online purchases. The relationship between privacy concerns and purchasing intention is partially mediated by trust, perceived risk, and attitude. This underscores the importance of addressing privacy concerns to maintain consumer trust and mitigate perceived risks associated with AI technologies.

Additionally, AI-driven autonomous weapons have raised ethical concerns about the potential loss of human control in life-and-death decisions, leading to calls for international regulation (Blauth et al., 2022). From the healthcare aspect, Electronic medical records (EMRs) in the health industry can provide many benefits for physicians, patients, and healthcare services if they are adopted by healthcare organizations (Bohr & Memarzadeh, 2020). Yet, concerns about privacy and security related to patient information can cause relatively low EMR adoption by several health institutions (Keshta & Odeh, 2021). In the context of implementing AI in clinical practice, Kelly et al. (2019) advised AI algorithm developers to be cautious about

possible risks. These risks include shifts in datasets, unintentional fitting of confounders, unintentional discriminatory bias, difficulties in generalizing to new populations, and unforeseen adverse effects of new algorithms on health outcomes.

The use of AI in education has also brought about some negative impacts. One significant concern is the potential for exacerbating existing educational inequalities, as access to AI-driven educational resources and tools may be limited to students from higher socioeconomic backgrounds, leaving disadvantaged students further behind (Williamson & Eynon, 2022). Additionally, privacy issues arise from the vast collection of student data by AI-based systems, which pose threats to students' personal information and increase the risk of data breaches (Zeide, 2017). Furthermore, an overreliance on AI in education might lead to decreased human interaction, affecting students' development of essential social skills and emotional intelligence (Chiang & Yin, 2022).

Moreover, cybersecurity attacks represent a central dimension in terms of privacy challenges. Bubukayr and Almaiah (2021) indicated how major cybersecurity attacks (malware attacks, phishing attacks, software failures, etc.) compromise systems based on four smartphone assets (devices, data, applications, and network connectivity). Moreover, the security challenges associated with cloud computing have been widely studied and examined (e.g., Ahmad et al., 2021). For example, at the University of Illinois Urbana-Champaign, Proctorio's remote-testing software was criticized for privacy invasions, accessibility issues, anxiety-inducing monitoring, and biased recognition favoring white faces over those of students of color (Khan, 2021).

AI biases exist in various applications ranging from face recognition, voice recognition, and search engines to self-driving cars and medical robots (Howard & Borenstein, 2018). They often reflect deep imbalances in institutional infrastructures and reinforce societal unfairness (Zou & Schiebinger, 2018). Algorithm-based HR decision-making can harm employees' integrity as these tools threaten self-determination in three ways: (1) by decreasing opportunities for human sense-making, (2) through a tendency to rely on technology in instances where reflexivity is required, and (3) through a lack of moral imagination (Leicht-Deobald, 2019). An experiment conducted by Lee (2018) revealed that when an algorithm makes recruitment decisions and performance evaluations, they are less likely to be interpreted as fair and trustworthy and are more likely to evoke more negative emotions. Yet, AI systems will continue to be biased as long as the data they receive are biased, according to the "bias in, bias out" principle (Mayson, 2018). According to Martin (2018), if an algorithm is designed to

preclude individuals from taking responsibility for a decision, then the algorithm designer should be held accountable for the ethical implications of the algorithm in use.

Moreover, De Cremer and Kasparov (2022) argue that AI itself has no ethics and that good or bad decisions made by algorithms are caused by human choices made at an earlier stage. Another factor that could amplify AI gender bias is the existence of this type of bias in society (Nadeem et al., 2020). Therefore, bias in an algorithm might push corporations to tackle specific actions. Amazon, for example, terminated its AI recruitment system, as it was considered biased against women (The Guardian, 2018). On a state scale, the commercial AI software COMPAS, used by judges to help decide whether a person charged with a crime should be released from jail before their trial, demonstrates significant bias against black defendants (Courtland, 2018). According to Holder (2014), the COMPAS algorithm violates ethical principles in its design because it includes factors that a person cannot control, such as a parent's criminal record or the first time the police stopped someone. In marketing, van Giffen et al. (2022) identified eight distinct ML biases: social, measurement, representation, label, algorithmic, evaluation, deployment, and feedback bias.

Positive impact

Shifting the focus to the numerous positive impacts of AI on society reveals its potential to revolutionize various aspects of life. AI has demonstrated a positive effect on privacy and security in multiple ways. AI-powered cybersecurity tools have been developed to enhance threat detection and mitigation, making it possible to identify potential vulnerabilities and respond to cyberattacks more effectively (Bonfanti, 2022). Furthermore, AI has been instrumental in developing secure authentication methods, such as biometric identification and behavioral analysis, providing a more reliable and personalized means of verifying users' identities (Murugavalli et al., 2016). AI-driven fraud detection and prevention systems are increasingly being used by businesses and financial institutions to protect against fraudulent activities (Villar & Khan, 2021; Dhieb et al., 2020)

AI has the potential to significantly impact healthcare by improving monitoring and diagnostic capabilities. For instance, AI can diagnose skin cancer more accurately than a professional dermatologist (Esteva et al., 2017). One of the main reasons is that using AI in such diagnosis can be processed more quickly and efficiently because it is analyzed based on a large body of knowledge and data (Rigby, 2019). In this manner, AI improves patient engagement and participation, service quality, and reduces medical errors and medical costs (Lee & Yoon, 2021). Wang et al. (2019) stated that colonoscopy tests assisted by artificial intelligence in China revealed 20% more polyps (growth of tissue from a surface in the body)

than those without. AI-driven algorithms can rapidly analyze large volumes of complex data, leading to accelerated drug discovery and development processes (Tripathiet al., 2023). By identifying potential drug candidates and predicting their efficacy, AI has the potential to reduce the time and costs associated with traditional drug development methods (Vamathevan et al., 2019). Furthermore, AI-powered remote patient monitoring systems enable healthcare providers to track patients' health status in real-time, improving disease management and early detection of health issues (Vyas et al., 2020). Personalized medicine, which tailors treatments to individual patient needs, is also being transformed by AI, as algorithms can predict patients' responses to specific therapies based on genetic and clinical data (Shah, 2018)). Safavi and Kalis (2019) estimate that "AI applications could create up to \$150 billion in annual savings for U.S. healthcare by 2026." These advances demonstrate the significant positive impact of AI in enhancing healthcare outcomes and contributing to a healthier society.

In communication and information, AI has facilitated improved communication and information dissemination, enabling better access to knowledge and more informed decision-making. For instance, AI-powered language translation services have broken language barriers and facilitated cross-cultural communication (Ogie et al., 2018). Furthermore, AI-driven recommendation systems have personalized content delivery and enhanced user experiences across various platforms, such as streaming services, social media, and e-commerce websites (Bawack et al., 2022). AI has significantly enhanced accessibility for disabled individuals, opening up new opportunities for improved quality of life and greater social inclusion. For instance, AI-powered speech recognition technology has enabled the development of applications that convert spoken language into text, providing valuable assistance to deaf and hard-of-hearing individuals (Saon et al., 2017; Shezi et al., 2020). Similarly, AI-driven computer vision systems have been employed in various applications, such as object recognition and scene description, to assist visually impaired individuals in navigating their surroundings (Abdolrahmani et al., 2020). These systems provide tailored content suggestions by analyzing user preferences and behavior, promoting more engaging and relevant user experiences.

In education, AI-driven adaptive learning systems have revolutionized how students learn by providing personalized learning experiences. These systems tailor educational content to individual students, catering to diverse learning styles and leading to improved learning outcomes (Chen et al., 2016; Xie et al., 2019). AI has also enhanced educational resources by supporting the development of more interactive and engaging content, such as intelligent tutoring systems, which provide real-time feedback and support to students (Schmid et al.,

2022). Moreover, AI applications have assisted educators in better understanding student needs by analyzing student performance data and providing insights into areas requiring additional support. This data-driven approach enables educators to create more targeted and effective teaching strategies (Datnow & Hubbard, 2016; Gupta et al., 2021).

AI has the potential to contribute to social equity by addressing social inequalities and enhancing access to services and resources for marginalized groups. According to Haluza & Jungwirth (2023), “the most direct way that AI could contribute to sustainability is through its ability to analyze data and discover patterns that humans may not be able to detect” (p. 13). The process of AI analyzing data can address the problem of social inequality (Zajko, 2020). Furthermore, AI can help enhance accessibility to services and resources for marginalized groups by providing tailored solutions and bridging gaps in service provision (Guo & Li, 2018; Zeleznikow, 2017). For instance, AI-powered chatbots can be employed to deliver mental health support in areas with limited access to mental health professionals (Abd-alrazaq et al., 2020). AI can also be used to develop personalized learning experiences that cater to the unique needs of learners from diverse backgrounds, thus promoting equal educational opportunities for all (Morimoto, 2022).

AI is vital for achieving zero hunger by 2030, a key goal of the SDGs. Addressing the digital divide between developed and underdeveloped countries is crucial. Despite economic resource limitations, these regions have untapped potential to meet global population demands. AI is used in agriculture for various tasks, including crop surveillance, irrigation management, disease identification, yield forecasting, and supply chain optimization. It enhances food safety and security by ensuring higher crop yields through multi-temporal remote sensing (RS) techniques, aiding in crop phenotyping, land cover monitoring, soil moisture prediction, and biomass modeling (Ahmad et al., 2024). Tamasiga et al. (2023) provide strategies for developing countries to pursue SDG 2, focusing on eradicating hunger and malnutrition through sustainable agricultural practices. These strategies include fostering partnerships between developed and developing nations, investing in AI and big data analytics, providing data access and training for farmers, and developing supportive regulations. Additionally, measures like cash transfers and social security baskets are recommended to protect vulnerable populations from price spikes during supply chain disruptions. By integrating these approaches, AI can significantly contribute to achieving zero hunger by enhancing agricultural efficiency and resilience in developing countries.

Aspect of Society	Positive Impact of AI	References	SDGs	Negative Impact of AI	References	SDGs
Privacy & Security	AI-powered cybersecurity tools for threat detection and mitigation	Bonfanti, 2022; Kaloudi & Li, 2020; Meier, 2021	SDG 9, SDG 16	Increased surveillance and erosion of privacy	Zuboff, 2019; Chesney & Citron, 2019	SDG 16
	Secure authentication methods using AI	Murugavalli et al., 2016; Wazid et al., 2021	SDG 9, SDG 16	Potential misuse of AI for malicious purposes	Altmann & Sauer, 2017; Blauth et al., 2022	SDG 16
	AI-driven fraud detection and prevention	Villar & Khan, 2021; Dhieb et al., 2020	SDG 16			
Healthcare	Improved diagnostics and treatment through AI-powered tools (e.g. cancer)	Bohr & Memarzadeh, 2020; Esteva et al., 2017; Rigby, 2019; Lee and Yoon, 2021.	SDG 3	Privacy concerns with patient data	Keshta & Odeh, 2021;	SDG 3, SDG 10
	Remote patient monitoring and personalized medicine	Vyas et al., 2020; Zhang et al. 2020.	SDG 3	Potential biases in AI algorithms affecting treatment recommendations	Kelly et al., 2019	SDG 3, SDG 10
	Enhanced drug discovery and development	Paul et al., 2021; Vamathevan et al., 2019.	SDG 3	Unequal access to AI-powered healthcare	Shaheen, 2021.	SDG 3, SDG 10
	Create savings for healthcare	Safavi and Kalis, 2019.				
Education	Personalized learning experiences through adaptive AI systems	Schmid et al., 2022; Chen et al., 2016; Xie et al., 2019	SDG 4	Widening digital divide	Rowe & Perry, 2016; Gabriel et al., 2022	SDG 4, SDG 10
	Enhanced accessibility to education resources for remote and rural areas	Schmid et al., 2022; Gopal et al., 2020	SDG 4	Data privacy concerns for students and educators	Zeide, 2017	SDG 4, SDG 16
	AI-powered tools for assessing learning outcomes and progress	Datnow & Hubbard, 2016; Gupta et al., 2021	SDG 4	Overreliance on AI, potentially reducing human interaction in education	Chiang & Yin, 2022	SDG 4
Social Equity	AI-powered tools for identifying and addressing social inequalities	Zajko, 2021; Lutz, 2019	SDG 10	Reinforcement of existing biases in AI algorithms	Howard & Borenstein, 2018	SDG 10
	Enhanced accessibility to services and resources for marginalized groups	Guo & Li, 2018; Zeleznikow, 2017	SDG 10	Exacerbation of inequalities due to unequal access to AI technologies	Zou & Schiebinger, 2018	SDG 10
	AI can achieve zero hunger	Tamasiga et al., 2023; Ahmad et al., 2024	SDG 2	Discrimination (E-justice issues)	Lee, 2018; Holder, 2014; Babuta & Oswald, 2018; Hoekstra & Sloan, 2020; Fryer, 2020	SDG 5, SDG 10, SDG 16
Communication & Information	Enhanced accessibility for disabled individuals	Saon et al., 2017; Shezi et al., 2020; Abdolrahmani et al., 2020	SDG 10	Fake news	Alcott & Gentzkow, 2017; Lazer et al., 2018; Berthon & Pitt, 2018; Di Domenico et al., 2021	SDG 16
	Enhanced information retrieval	Bawack et al., 2022	SDG 4, SDG 9	Human autonomy	Krafft et al., 2017; Mousavi et al., 2020; Martin, 2018	SDG 16
	Eliminate cultural barriers (e.g., AI-powered language translation)	Noja, 2021; Ogie et al., 2018		Cybersecurity	Bubukayr & Almaiah, 2021; Bachura et al., 2022	SDG 16

Tableau 18 : Overview of positive and negative impacts of AI on society and their corresponding SDGs (source: Ahmad based on Literature)

3.2.4.2.3 Economic implications of AI: prospects and pitfalls

AI has become a popular area of research in science and technology. The relationship between AI's positive and negative impacts on various aspects of the economy and their corresponding SDGs is illustrated in Table 19.

AI negative consequences

Focusing on the negative consequences of AI in the economy, it becomes essential to explore how these advanced technologies can have unintended effects. A regression that uses internet connectivity and per capita gross domestic product (GDP) as independent variables to clarify GDP growth shows that the effect of AI is a significant determinant of negative GDP growth prospects from 2000 until 2017 (Puaschunder, 2019). Korinek and Stiglitz (2017) showed that the diffusion of AI and other worker-replacing technological progress could negatively influence income distribution and employment. The rapid increase in income inequality supports this argument and has been going along with the rapid technological changes that started in the 1980s (Atkinson, 2013). At the country level, income inequality strongly correlates with innovation activities. Norton (2017) can summarise this dimension and indicates that AI technology may enlarge national inequality gaps: (1) boosting the advantage of capital over labor; (2) relative declines in medium- and low-skilled employment shares; (3) weakening of labor institutions; and (4) reduction of tax bases, thus weakening governments' capacity for redistribution.

There is a continuous discussion regarding the impact of new and developing technology on employment and skills. Pessimism is increasing as intelligent computers learn to perform more sophisticated activities previously performed by people, putting millions of jobs worldwide at risk. As many questions remain unanswered, one can refer to these discussions as "the great AI fear" (Clark, 2020). This fear can be captured in the following studies. Frey and Osborne (2017) have attracted the attention of media and policy practitioners. According to them, roughly 47% of US employment is at risk of computerization. For the same country, Acemoğlu and Restrepo (2017) found that "one more robot per thousand workers reduces the employment-to-population ratio by about 0.18–0.34 percentage points and wages by 0.25%–0.5%." Aghion and his colleagues (2019) discussed robotization's impact on employment in France between 1994 and 2014. The authors found that robotization reduces employment for educated and non-educated workers and stresses the importance of the labor

market and educational policies. In Denmark, the findings of Holm and Lorenz (2021) suggest that the proliferation of AI can increase inequality in the labor market by enhancing the abilities required for high-skill occupations while having proportionally more significant adverse effects on other jobs. In a separate report, McKinsey forecasted that 75 million and 375 million workers may need to transfer occupations by 2030 due to the high rate at which regular work is automated (Manyika et al., 2017). Robots can provide two types of services – tangible and intangible (Wirtz et al., 2018). Intangible services, robots can be humanoid (e.g., reception at a school) or non-humanoid (e.g., autonomous taxi). Van Doorn et al. (2017) classify intangible services into video-based (e.g., ATM service), voice-based (e.g., Siri), and text-based (e.g., online casino service). Lu et al. (2019) anticipate that humans will deliver tasks with high emotional needs while service robots will handle analytical functions with fewer emotional needs. In the field of labor economics, the substitution of human labor with emerging technologies and robots has been hotly discussed. AI will change the jobs of research economists by automating some aspects of econometrics (Athey, 2019). AI is most likely to displace radiologists in the healthcare industry as a regular radiologist can read 50–100 X-rays a day, while an AI-supported system can read 10–100 times more images (Pesapane et al., 2018). The effect of AI and employment can be summarised in task substitution, task complementarity, and the creation of new jobs. According to Georgieff and Hye (2021), task substitution dominates only for workers with low digital skills, but productivity effects dominate for workers with high digital skills. Utilizing AI technology to deliver consumer value necessitates the adoption of new practices, skills, operational procedures, and business models. Many businesses have discovered, for instance, that AI algorithms do not produce definitive answers but rather offer tentative solutions (e.g., probability-based predictions), which require human interpretation, justification, and action to make concrete, valuable results (Tarafdar et al., 2019)

This may not always be positive for businesses. Cost issues will accompany the adoption of AI. In addition to the expenditures associated with the implementation (i.e., system costs, project and consulting fees, system integration costs, and system upgrading costs), various other charges will likely be incurred when organizations transit to AI (Bughin et al., 2018). The transition costs may include those incurred while restructuring the organization, severance for workers who the new technology may displace, and the costs of upgrading the skills of existing employees (Bughin et al., 2018). Based on 119 announcements of 62 listed firms invested in AI, Liu and his colleagues (2022) also found that AI investment negatively influenced the firms' market value. The firms' stock prices decrease by 1.77% on the announcement day.

Nonmanufacturing businesses and companies with limited information technology capabilities or low credit ratings are impacted more negatively than other businesses. The rise of AI has led to market monopolization (Rasskazov, 2020). AI has been so powerful that it has enabled companies to dominate and absorb others, thereby creating superstar businesses. These multinational corporations are headquartered in a few countries but operate on a global scale. This amplifies the stark economic differences between developed and developing nations, as developing nations have a significant accessibility problem due to a lack of internet and technology. However, the concept of market monopoly by “superstar firms” is not exclusive to corporations. There is a risk that countries that lead to AI advancement will reap all the benefits, becoming “superstar countries” and reaping all the rents associated with the development of AI (Korinek and Stiglitz, 2021).

Implementing AI-driven farming techniques and technologies can exacerbate existing inequalities in the agricultural sector by disproportionately benefiting larger, well-funded farms while leaving smaller, resource-poor farms disadvantaged (Jarosz, 2012; Gwagwa et al., 2021).

AI also poses potential negative impacts on the financial services and fintech sectors. For instance, AI may inadvertently increase systemic risk in financial markets, as the widespread use of similar AI models could lead to herd-like behavior and exacerbate financial crises (Danielsson et al., 2022; Bookstaber, 2017). Additionally, AI may contribute to the financial exclusion of vulnerable populations, as automated decision-making processes could disproportionately affect those with limited access to financial services or a lack of credit history, for example, racial bias in credit scoring of loan applications (Bhutta et al., 2022; Berg et al., 2019).

AI positive impact

AI’s positive impact on the economy is closely linked to ethical values, as it has the potential to generate increased wealth and well-being, ultimately contributing to improved living standards and human flourishing. The EU’s High-Level Expert Group on AI makes this very clear when it states: “AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation” (p.4). Corporate digital transformation leads to increased innovation, ultimately improving corporate performance and promoting the rapid development of enterprises (Guo et al., 2022). AI has the potential to accelerate research and development processes significantly (Santus et al., 2021) across various industries, like healthcare (e.g., the use of deep learning in drug discovery, Chen et al., 2018) as it can analyze large datasets, optimize workflows, and make predictions with remarkable speed and accuracy

(Sun et al., 2020). AI has the potential to enhance collaboration and information sharing by facilitating communication, automating tasks, and enabling efficient data management across various domains (Khan et al., 2020)

AI offers various benefits in the financial services and fintech sector that enhance its overall efficiency. First, AI can significantly improve risk assessment, fraud detection, and trading by using advanced analytics and pattern recognition to identify unusual activities and make more informed decisions (Bhatore et al., 2020). The majority of data generated from Internet banking and Automated Teller Machine (ATM) transactions, which account for approximately 2.5 quintillion bytes, is unstructured and valuable for fraud detection, risk management, and customer satisfaction (Shakya & Smys, 2021). Leveraging the latest Big Data Analytics methodologies can help address the challenges and competition among banks in managing this vast amount of data (Shakya & Smys, 2021). Second, AI can revolutionize customer service and personal finance management by providing personalized recommendations, automating routine tasks, and enabling chatbots for immediate support, thereby delivering a more seamless and tailored customer experience (Boustani, 2022)

Basri (2020) found that AI-assisted social media marketing increased the number of clients and customer bases and increased business profitability. Domino's Pizza, for instance, has decreased delivery time and increased the accuracy of delivery time prediction from 75% to 95%, and mining companies in Australia have reduced mining costs, improved worker safety, and increased productivity by 20% by utilizing autonomous trucks and drilling technology. The rise of AI enhances customer engagement in an evolving technological environment, allowing marketers to use storytelling techniques to transport and persuade their customers narratively (Van Laer et al., 2018). Viewed as a key growth factor, AI can allow any organization to achieve the following: increase the efficiency of operations, maintenance, and supply chain operations, optimize and improve the customer experience, improve products and services (with new features), as well as item recommendation processes (retail and other industries) (Kuzey et al., 2014; Chinchanchokchai et al., 2021).

Automation has significantly increased productivity across various industries, streamlining processes and reducing human error (Barbut, 2014). As a result of the rapid development and adoption of AI, PricewaterhouseCoopers (PwC, 2018) believes that the global GDP might increase by up to 14% (the equivalent of \$15.7 trillion) by 2030. In the same year, McKinsey (2018) estimated that AI may deliver an additional economic output of around US\$13 trillion by 2030, increasing the global GDP by about 1.2% annually. The case for productivity growth from AI can look to empirical research on robotics for support. The

research by Graetz and Michaels (2015) highlights that robotics contributed approximately 0.4 percentage points to annual GDP growth in 17 countries from 1993 to 2007, accounting for about one-tenth of total GDP growth during this period. Studies, including one conducted by Li et al. (2022), show that the digital economy initially boosts growth but later restricts it through its effects on carbon emissions. Furthermore, Cockburn et al. (2018) discuss how AI has reshaped the innovation process and enhanced the capabilities of R&D firms, leading to significant economic spillovers. Acemoğlu (2018) suggests that AI can mitigate the economic impacts of aging populations and enhance growth by increasing total factor productivity. Shackleton (2018) stated that the current level of AI development has not changed the TFP of some high-end manufacturing industries and has only helped improve the TFP of traditional manufacturing. According to Bughin and Hazan (2017), investments in AI are increasing. For instance, Anthropic, an AI startup, is being backed by Google with an investment of nearly \$400 million. The startup is currently developing a competitor to OpenAI's ChatGPT (Alba & Bass, 2023). Revenue collected from the robot tax can be redistributed as universal basic income or as transfers to workers displaced by robotic systems and AI and unable to be relocated to new jobs (Merola, 2022).

According to Chen (2023), AI plays a significant role in each stage of recruitment, such as recruitment promotion, job search, application, screening, assessment, and coordination. Wamba-Taguimdje et al. (2020) emphasized AI's organizational benefits, particularly its capacity to enhance organizational (financial, marketing, and administrative) and process-level performance. By leveraging these AI characteristics, businesses can increase their business value. Also, integrating AI into an organization or workplace can boost productivity significantly and enhance the quality of decision-making, enabling individuals to make better and faster choices (Knight, 2015). A human resource (HR) department's sourcing process takes time. The workload may be shared using AI and machine intelligence, making it simpler to find eligible candidates. A study by Vrontis et al. (2020) highlights that AI can improve the accuracy of human resource management (HRM) decisions, particularly for non-experts, and reduce the time required for such decisions.

AI has the potential to significantly enhance agriculture by optimizing farming practices, increasing crop yields, and allowing for rapid diagnosis of plant diseases. ML and AI-powered tools can analyze large volumes of data to identify the optimal conditions for crop growth, enabling farmers to make data-driven decisions that maximize productivity and minimize waste (Ghosal et al., 2020; Bachmann et al., 2022). AI can also play a crucial role in

the early detection and management of plant diseases, helping to prevent severe damage to crops and reducing the need for excessive pesticide use (Kamilaris & Prenafeta-Boldú, 2018).

Aspect of Economy	Positive Impact of AI on Economy	References	SDGs	Negative Impact of AI on Economy	References	SDGs
Employment & Workforce	Automation leading to increased productivity	Barbut, 2014; Brynjolfsson & McAfee, 2014	SDG 8, SDG 9	Job loss due to automation	Felten et al., 2019; Frey & Osborne, 2017; Aghion et al., 2019; McKinsey, 2017	SDG 8
	Creation of new job opportunities in AI-related fields	Bessen, 2017; Hunt et al., 2022	SDG 8, SDG 9	Widening skills gap	Norton, 2017; Holm & Lorenz, 2021; Athey, 2019; Pesapane et al., 2018; Georgieff & Hyee, 2021	SDG 4, SDG 8
Economic Growth & Productivity	AI can improve efficiency and productivity across industries	PwC, 2018; Iain et al., 2019; Kromann et al., 2020	SDG 9	AI can hamper economic growth	Puaschunder, 2020	SDG 8, SDG 9
				AI may exacerbate income inequality and wealth concentration	Atkinson, 2013; Korinek & Stiglitz, 2017; Norton, 2017	SDG 10
Businesses	Increased Investment in AI by Businesses	Bughin et al., 2017; Lunden, 2021; Fickling, 2019	SDG 8, SDG 9, SDG 17	AI can increase cost for businesses	Bughin et al., 2018; Intalar & Pharm, 2022	SDG 8, SDG 9, SDG 12
	AI optimize Human Resource Management (e.g., optimize employment, enhance accuracy in decision-making, better understand	Garge et al., 2021; Wamba-Taguimdje et al., 2020; Vrontis et al., 2020; Nawaz, 2020	SDG 5, SDG 8, SDG 10	AI has the potential to decrease company's market value	Liu et al., 2022	SDG 8, SDG 9
	AI can enhance Marketing (boosts client base and profitability; enhancing delivery efficiency; leverages storytelling techniques)	Basri, 2020; Minevich, 2020; Van Laer et al., 2018	SDG 8, SDG 9, SDG 12	AI led to market monopolisation	Rasskazov, 2020	SDG 8, SDG 9
Innovation & Research	AI can spur innovation and the development of new products/services	Guo et al., 2022	SDG 8, SDG 9, SDG 3, SDG 12			
	AI can accelerate research and development processes	Santus et al., 2021; Sun et al., 2020; Wang & Srinivasan, 2017	SDG 3, SDG 4, SDG 9,			
	AI can improve collaboration and information sharing	Khan et al., 2020; Androutsopoulou et al., 2019	SDG 4, SDG 9, SDG 17			
Financial Services & Fintech	AI can improve risk assessment, fraud detection, and trading	Bhatore et al., 2020; Eltweri et al., 2021; Shakya & Smys, 2021	SDG 8, SDG 9	AI may increase systemic risk in financial markets	Danielsson et al., 2022; Bookstaber, 2017	SDG 8, SDG 9
	AI can enhance customer service and personal finance management	Boustani, 2022; Gigante & Zago, 2023	SDG 1, SDG 8, SDG 9, SDG 10	AI may contribute to financial exclusion for vulnerable populations	Bhutta et al., 2022; Berg et al., 2019	SDG 1, SDG 8, SDG 10, SDG 16
Agriculture & Food Industry	AI can optimize farming practices and increase crop yields, rapid diagnosis of plant disease	Bachmann et al., 2022; Doshi & Varghese, 2022; Ghosal et al., 2020; Tzachor et al., 2022	SDG 2, SDG 8, SDG 12	AI-driven farming may exacerbate existing inequalities in agriculture	Jarosz, 2012; Foster et al., 2023; Gwagwa et al., 2021	SDG 2, SDG 10
	AI can reduce food waste and improve supply chain efficiency	Kamilaris & Prenafeta-Boldú, 2018	SDG 2, SDG 8, SDG 12	Overreliance on AI may reduce resilience to unforeseen challenges	Dong, 2021; Gulyas & Edmondson, 2021	SDG 9, SDG 11, SDG 12
	Optimization of land and water use	Liu et al., 2013; Xiang et al., 2021	SDG 2, SDG 6, SDG 15			

Tableau 19 : Overview of the relationship between AI and the economy (source: author based on literature)

4.3 Discussion

The swift progress in AI technology is profoundly transforming the business landscape, presenting opportunities and challenges. Plekhanov et al. (2022) discuss how digital transformation brings complex strategic challenges, emphasizing the importance of organizational understanding of AI principles, especially in sustainability (Van Wynsberghe, 2021). Demdoun et al. (2021) highlight this by pointing out the scant attention to environmental factors in southeastern Algeria, which mirrors a worldwide pattern of insufficient management of AI's environmental effects. Consequently, there is an increasing need for more robust governmental and consumer initiatives to foster sustainable AI practices.

Considering these complexities, analyzing existing frameworks can provide deeper insights. Researchers like Mikalef et al. (2022), Dignum (2019), and Trocin et al. (2023) have explored the opportunities, challenges, and principles for RAI, and this study contributes to the field by offering a framework for RAI implementation. From the perspective of international organizations, the OECD Principles on AI (2019) set forth high-level, values-based guidelines for RAI that highlight the benefits for people and the planet and promote transparency and accountability. The European Commission's Ethics Guidelines for Trustworthy AI (2019) define seven key trustworthiness criteria, including human oversight and societal well-being (HLEGAI, 2019). From an academic viewpoint, Nylén (2021) presents a comprehensive framework centered on digital innovation strategy, focusing on user experience, value proposition, digital evolution scanning, skills, and improvisation. Likewise, Correani et al. (2020) stress the need to define the scope of digital transformation, underscoring the essential role of data in enhancing transformed activities, tasks, and services that deliver value to customers.

In this chapter, we find similar findings to Vinuesa et al. (2020) regarding AI's emergence as a powerful enabler in alleviating poverty (SDG 1) through enhanced economic growth and social protection and in combating hunger (SDG 2) via improved agricultural productivity and food security. In health (SDG 3) and education (SDG 4), AI's contributions to better healthcare services, disease prevention, and personalized learning are noteworthy. However, the impact on gender equality (SDG 5) remains complex, showing both positive and negative effects. AI also supports clean water and sanitation (SDG 6) by optimizing water management and facilitates the transition to sustainable energy (SDG 7) through energy

optimization. In urban development (SDG 11), AI drives smart cities and resource efficiency, while its influence on peace, justice, and strong institutions (SDG 16) is mixed, highlighting the need for careful management of surveillance and privacy issues. Furthermore, our chapter goes beyond Vinuesa et al. (2020) by exploring AI's nuanced impact on [specific SDG], providing a deeper understanding of its potential and challenges in this area.

Regarding the SDGs' economic pillar, our findings align with Vinuesa et al. (2020) in recognizing AI's potential to reduce inequalities (SDG 10) by creating new job opportunities in AI-related fields and optimizing human resource management. However, both studies also highlight the potential for AI to exacerbate income inequality and wealth concentration, necessitating careful policy interventions. Also, both studies identify AI as a critical enabler of decent work and economic growth (SDG 8) through automation and increased productivity. However, it also poses risks, such as job displacement due to automation. Furthermore, AI facilitates industry innovation and infrastructure (SDG 9) by improving efficiency and fostering innovation across various sectors, consistent with the positive impacts observed in Vinuesa et al. (2020). Regarding responsible consumption and production (SDG 12), both studies acknowledge AI's role in optimizing resource use and improving supply chain efficiency while cautioning against overreliance on AI, which can reduce resilience to unforeseen challenges. Additionally, our research extends the discussion by exploring AI's nuanced impact on financial services and fintech, demonstrating its ability to enhance risk assessment, fraud detection, and customer service, thereby contributing to the stability and inclusivity of financial systems. Moreover, while Vinuesa et al. (2020) emphasize AI's role in fostering partnerships for the goals (SDG 17), our findings provide a deeper analysis of AI's impact on innovation and research, highlighting its ability to accelerate research processes and improve collaboration and information sharing. This comprehensive analysis underscores AI's pivotal role in economic sustainability, corroborating the broad positive impacts that Vinuesa et al. (2020) identified while offering new insights into its potential and challenges.

Regarding AI's significant influence on the environmental pillar of the SDGs. Our findings align with Vinuesa et al. (2020) in recognizing AI's role in enhancing resource management (SDG 12) through improved energy efficiency, optimized water usage, and enhanced waste management. However, they also highlight the negative impact of increased energy consumption due to energy-intensive model training, power-hungry data centers, and the environmental hazards of e-waste. In terms of climate action (SDG 13), both research

acknowledges AI's potential to reduce greenhouse gas emissions, improve climate prediction, and enhance disaster response. Nonetheless, both studies point out the adverse effects of resource depletion, such as mining rare-earth elements and land degradation, which pose significant environmental challenges. Biodiversity conservation (SDGs 14 and 15) and pollution reduction (SDG 3) are another areas where AI shows considerable promise. Additionally, while Vinuesa et al. (2020) emphasize AI's contribution to sustainable cities (SDG 11), our findings provide a deeper analysis of AI's role in urban sustainability and improving residents' quality of life. This comprehensive analysis corroborates the broad positive impacts that Vinuesa et al. (2020) identified while offering new insights into AI's potential and challenges in fostering environmental sustainability.

Unlike the frameworks mentioned in the literature, the proposed integrative framework moves beyond a singular focus, such as user-centric approaches. It provides a broad and in-depth look at AI's technical, legal, and sustainability aspects, making it flexible and applicable across different industries. This comprehensive analysis makes the study relevant to a wide range of readers, including data analysts who will appreciate its technical detail and legal and managerial policymakers who can find valuable insights for governance and strategic planning.

The strength of the proposed framework lies in its potential for widespread applicability across different countries or regions. Although the legal aspects may differ due to the variety of AI-related laws and regulations unique to each jurisdiction, the core principles of RAI remain unchanged. In this regard, organizations must align with local and international laws to ensure effective implementation. Therefore, incorporating a comprehensive and integrative framework into an organization's digital strategy is vital for addressing AI challenges and maintaining a commitment to fairness, accountability, transparency, and security during innovation and expansion. This study enhances understanding of RAI by underscoring ethical and legal issues, promoting stakeholder involvement, offering a framework for responsible practices, and guiding diverse groups towards ethical AI deployment.

Having detailed the strengths of the RAI framework, it is pertinent to delve into the policy implications of this study, which are aimed at benefiting a wide array of stakeholders within the AI sphere, including businesses, developers, and policymakers. Here's how the RAI framework could specifically impact these groups:

- **Businesses:** The RAI framework provides a robust guide for companies looking to integrate AI into their operations responsibly. By adhering to the principles of responsible innovation outlined by Stilgoe et al. (2013), businesses can gain a competitive edge, bolster customer trust, and reduce risks associated with AI deployment, such as biases or privacy breaches. For practical application, businesses should consider implementing mechanisms to address specific technical challenges, such as quantization error, which can affect the accuracy and reliability of AI models. This involves providing adequate training and resources to technical teams and incorporating advanced verification methods like QEBVerif, a quantization error-bound verification technique proposed by Zhang et al. (2023). Such methods ensure that the AI systems operate within expected parameters, reducing the likelihood of errors that could impact business operations or customer experiences.

Further extending this framework, businesses could integrate models like the one developed by Zhu et al. (2023), which utilizes a multi-strategy approach to relation extraction to reduce noise in unstructured teaching resources for online education. This model exemplifies how businesses can apply the RAI framework to ensure that technical challenges in AI deployment are responsibly addressed, particularly in how AI interacts with educational content. Moreover, companies are encouraged to develop internal policies that reinforce the adherence to RAI principles across all AI project development and deployment levels. This might include regular ethical audits, the establishment of AI ethics committees, and ongoing training programs for staff to remain abreast of best practices and emerging challenges in AI ethics and governance.

- **Developers:** AI developers have the opportunity to leverage the RAI framework as a comprehensive guide for crafting AI systems that are intrinsically responsible and adhere to high ethical standards. While the technical aspect of the framework offers detailed guidelines on implementing principles such as fairness, transparency, and safety, developers need to understand that these technical specifications are just part of the solution. To truly align with RAI, developers must also consider their work's wider ethical, social, and organizational consequences. This broader approach encourages a deeper integration of ethical considerations into the development process, moving beyond just technical compliance to address how AI systems impact society.

To extend the practical application of this framework, developers could establish cross-functional teams that include ethicists, user experience designers, and domain experts. These teams can collaborate to ensure that AI systems are technically proficient, socially

attuned, and ethically grounded. For example, developers could implement user-centered design practices to gather feedback from diverse user groups, ensuring that AI systems are accessible and beneficial to a broad spectrum of the population. Additionally, simulation environments could be used to test AI behaviors in various social scenarios, helping to predict and mitigate potential negative impacts before deployment. By adopting these comprehensive development practices, developers can create AI systems that are more trustworthy, socially responsible, and closely aligned with human values, ultimately leading to innovations that are both technologically advanced and ethically sound.

- **Policymakers:** Policymakers can gain essential perspectives from the RAI framework to shape regulatory environments that promote RAI innovation while protecting public interests. This is especially pertinent given the diverse regulatory approaches to AI, such as the EU's recent revisions to the AI Act and the varying regulations at the state level in the US. The principles outlined in the RAI framework offer a solid foundation for policymakers to manage these different legal contexts, steering AI development toward innovative, ethical, and advantageous solutions for society in various regions.

4.4 Conclusion

In a time when AI rapidly transforms industries and sectors, implementing RAI dimensions has become more essential than ever. This study significantly contributes to the field by introducing an integrated and comprehensive RAI framework. This framework covers technical principles such as accountability, privacy, transparency, safety, fairness, and autonomy, and it spans legal compliance, responsible innovation management, and alignment with SDGs. RIM is vital as it translates ethical AI practices into tangible elements within the corporate culture. Organizations can proactively identify and respond to potential AI impacts by incorporating aspects like anticipation and reflexivity, promoting ethical decision-making and inclusive stakeholder involvement. Concurrently, the legal dimension is critical for navigating AI's complex legal landscape and preventing ethical breaches. While the RIM and legal aspects are crucial in integrating responsible practices in AI, it is essential to recognize that all dimensions of the framework, including technical and sustainability aspects, are interconnected and equally vital. The primary scientific contribution of this research lies in its synthesis of these diverse dimensions into a unified framework, offering a comprehensive

approach to RAI that is innovative and globally applicable. This study also provides valuable insights for stakeholders in AI and businesses aiming to understand and implement RAI effectively, thereby fostering the responsible use of AI technologies.

While the framework establishes a solid foundation for RAI, it also recognizes the difficulties associated with its implementation. There is a highlighted need for strict enforcement mechanisms and clear accountability measures to ensure the framework's practical application in real-world scenarios. Furthermore, the framework's global applicability, coupled with the need for adaptation to diverse legal environments, underscores its flexibility and adaptability. However, it's essential to acknowledge that this study has its limitations, primarily focusing on the business uses of AI and not covering the entire range of AI technologies, such as those used in military contexts.

This study sets the stage for further academic inquiry into AI and significantly impacts AI technologies' practical development and regulation. A comparative analysis could reveal differences in AI adoption across various regions and their respective outcomes. Developing specific indicators for each framework dimension would enhance assessment and evaluation tools. Longitudinal research, tracking changes over time, could deepen our understanding of the dynamic nature of AI and its broader effects. Additionally, a critical avenue for future research is the empirical application of our RAI framework in targeted sectors, like the technology industry, to evaluate its principles and flexibility in practical settings. This targeted application could yield more profound insights into the utility and effectiveness of our framework. Regarding regulatory implications, the comprehensive design of our RAI framework provides a template for informing future legislative and policy decisions concerning AI. It highlights the necessity for flexible regulatory approaches that can adapt to the rapid progression of AI, ensuring that ethical, transparent, and inclusive practices continue as the technology evolves. This proactive strategy is vital for managing the evolving AI landscape and promoting its responsible development and regulation.

Chapter 5: Generative AI: A Quantitative Study on Emerging Risks and Impacts

Chapter 1: What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI? What are the risks of GAI and their impact on individual society organizational and environment?



Chapter 2: What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?



Chapter 3: What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?



Chapter 4: What are the essential dimensions of RAI and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?



Chapter 5: What are the specific generative AI risks? How do these GAI risks affect different aspects of society, business, and the environment?



Chapter 6: What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?

About this chapter

This chapter has been submitted for the **journal**:

Haidar, A. and Balague, C. (2024). Generative AI: A Quantitative Study on Emerging Risks and Impacts. *Business & Society* (HCERES A; ABS 3; FNEGE 2)

This chapter has been selected and presented at the following **conferences**:

- Haidar, A., & Balagué, C. (2024). A Generative AI: A Quantitative Study on Emerging Risks and Impacts In *Digital Transformation Society Conference (DTS 2024)*. 20-23 May, Naples, Italy.
- Haidar, A., & Balagué, C. (2024). A Generative AI: A Quantitative Study on Emerging Risks and Impacts In *International Society for Professional Innovation Management (ISPIM) Conference*. 9-12 June, Tallinn, Estonia.
- Haidar, A., & Balagué, C. (2024). A Generative AI: A Quantitative Study on Emerging Risks and Impacts In *R&D Management Conference*. 17-19 June Stockholm, Sweden

5.1 Theoretical framework of generative AI risks

5.1.2 Theoretical foundation of this study

In our framework for classifying GAI risks, we integrate and extend existing theoretical models by identifying two primary dimensions: the nature of risk and the level of impact, as illustrated in Figure 24. This approach draws upon Hilgartner's (1992) observation that definitions of particular risks encompass three conceptual elements: an object deemed to 'pose' the risk, a putative harm, and a linkage alleging some form of causation between the object and the harm. Building on this foundation, our dual-dimensional approach is further inspired by Boholm and Corvellec's (2011) relational theory of risk, which posits that risk is inherently relational, emerging from the cognitive processes that link a 'risk object' to an 'object at risk' in specific contexts.

Building on this foundation, we also draw on insights from Scheer et al. (2014), who differentiates between "risk" and "hazard." Hazards refer to the potential to cause harm to targets such as human health or the environment, whereas risks involve the likelihood of exposure and the extent of the resulting damage. This distinction clarifies that our framework

focuses on both the harms associated with GAI and the risks stemming from these harms. In line with this, the Center for Security and Emerging Technology in Washington defines AI harm as “when an entity experiences harm (or potential for harm) that is directly linked to the behavior of an AI system” (Hoffmann & Frase, 2023, p. 2). Thus, in the context of our study, our framework examines both the inherent potential harms (risk object) posed by GAI and the associated risks (object at risk) by assessing the probability and impact of these harms on the number of incidents.

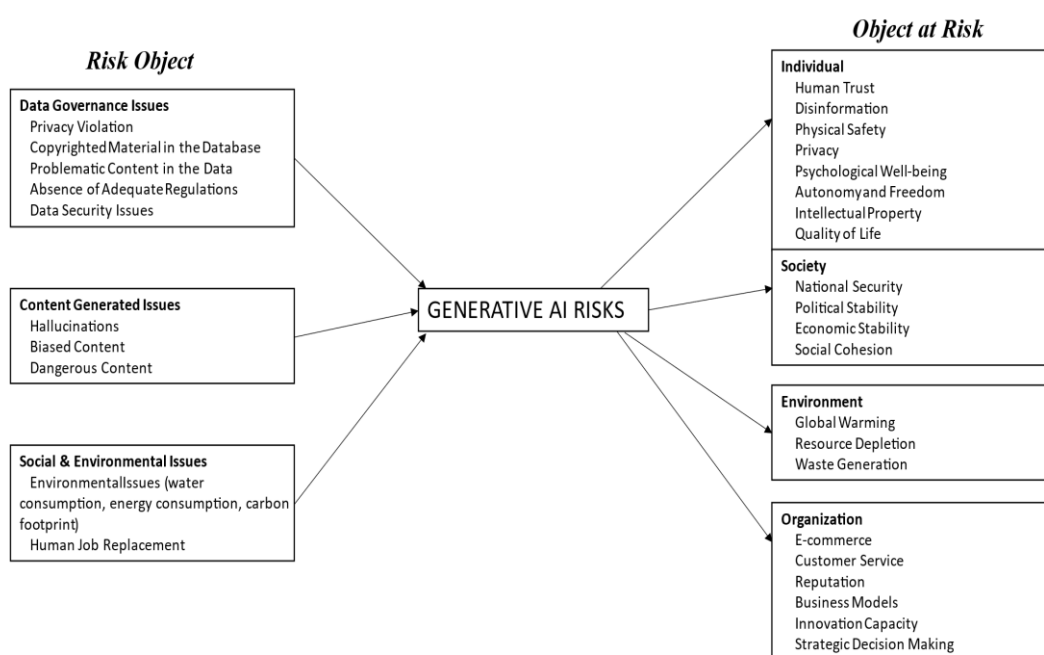


Figure 24 : Theoretical Framework of the study

5.1.2.1 Risk object

5.1.2.1. Definition

Boholm and Corvellec's (2011) relational theory of risk defines 'risk objects' as entities recognized as dangerous, not through inherent qualities but through societal and cognitive processes that construct them as such. Unlike 'hazards,' which are typically seen as innately risky, 'risk objects' emerge from a dual process of selection and conceptualization, where specific characteristics are emphasized to shape the object's identity as dangerous, while other

traits are minimized or ignored. This emphasis on certain characteristics effectively establishes the 'risk object' as a significant component in a given setting.

For example, classifying certain dog breeds as 'dangerous' illustrates this concept well. The act of labeling a breed like pit bulls as risky does not just recognize an existing danger but actively constructs the breed as a risk object. This is done by delineating a boundary that separates 'dangerous' dogs from 'non-dangerous' ones, creating a category based on perceived genetic and behavioral traits.

Thus, 'risk objects' represent a nuanced concept in risk theory, highlighting the socially constructed nature of what is deemed risky, in contrast to 'hazards,' which are generally accepted as inherently dangerous without the need for social or cultural confirmation

4.1.2.1.2 Risk object in GAI context

This dimension is influenced by Steimers (2022), who investigates the existential risks linked to AI, particularly highlighting how risks originate in systems that use contemporary machine learning techniques. Our research delves deeper and broadens these risks within the context of GAI, as detailed in Table 20, which targets specific areas.

Dimension: Risk object			
Sub-dimensions	Categories	Description	References
Data Governance	Privacy violation	. Unauthorized access and data breaches of sensitive information . Misuse of surveillance and collected data . Operation without individual knowledge or consent . Need for data protection principles in AI system design	Kunz & Wirtz (2023); Shum (2023); Humphreys et al. (2023)
	Copyright material in the database	. Replication of copyrighted content by AI models like GPT-4 . Risk of violating intellectual property laws . Blurring of lines between innovation and imitation . Need to reevaluate concepts of "authorship" and "work" in creative fields	Henderson et al. (2023); Smits and Borghuis (2022).
	Problematic content in the data (fake news, toxic speech)	. Significant risk of fake news and toxic speech in AI training datasets . Potential for AI systems to learn and disseminate misinformation . Adoption of harmful language patterns by AI . Escalation of harmful activities by AI-powered social bots . Production of toxic speech by models like GPT-3, mimicking extremist style	Hajli et al. (2022); McGuffie & Newhouse (2020).
	Absence of adequate regulations	. Lack of comprehensive regulatory frameworks for General AI (GAI) . Presence of critical issues in GAI, including ethical dilemmas, biases, and accountability challenges . Economic disparities and potential misuse due to inadequate regulations	Shouyang et al. (2023); Oduro et al. (2022); Amariles & Baquero (2023).
	Data security issues	. Societal and ethical concerns from potentially misleading or offensive AI-generated content . Influence on user's preferences and misinformation risks . Ethical challenges arising from AI content that blurs lines between reality and production . Risk of AI integration into search engines leading consumers to illegal websites	Sullivan and Fosso Wamba (2022). Ashraf et al. (2024); Subhash (2023); Rozado (2023)
Content Generated	Hallucinations	. Potential hallucinations for ChatGPT, such as creating fictitious sources . Undermining of integrity and accuracy in structured texts and academic writing . Impact on business contexts: efficiency and creativity tempered by risks of hallucinations and misinformation . Lack of cautious use and clear guidelines from publishers regarding ChatGPT's authorship and application.	Alkaissi & McFarlane (2023); Stokel-Walker (2023); Fui-Hoon Nah et al. (2023).
	Biased content (political bias, gender biased, discrimination...)	. Potential political biases in ChatGPT, with indications of a left-leaning response tendency . Biases attributed to training rules imposed on human trainers . Evident stereotypical behavior toward various social categories like religion and politics	Hartmann et al. (2023); Rozado (2023); McGee (2023); Kocoń et al. (2023); Busker et al. (2023).
	Dangerous content	. ChatGPT's potential to prompt harmful behaviors, such as misuse of medications, violence, and self-harm. . Inconsistencies in OpenAI's content moderation, with uneven enforcement across demographics. . Higher flag rates for negative comments towards women than men.	Subhash (2023); Rozado (2023).
Social & Environmental Issues	Environmental issues (water consumption, energy, carbon footprint)	. Annual consumption of 200 TWh by data centers. . Expected 15-fold increase in data center electricity consumption by 2030. . Data centers as major consumers of electricity and water in the U.S (a midsize data center uses as much water annually for cooling—80 to 130 million gallons—as three hospitals or two golf courses combined)	Çelebi (2023); Andrae and Adler (2015), Sverdlík (2016); FitzGerald (2015);
	Replacement of human jobs by machine	. AI-induced job replacement causing employee anxiety, impacting stress levels, organizational commitment, and productivity. . In healthcare, for example, a regular radiologist reads 50–100 X-rays a day, while an AI system can process 500–10,000, significantly more efficiently.	Nazareno & Schiff (2021); Khan (2021); Pesapane et al. (2018)

Tableau 20 : Risk object dimensions (source: authors)

Data governance issues

Privacy violations in GAI constitute a significant concern due to unauthorized access and data breaches exposing sensitive personal information from large datasets. These violations raise the possibility of misusing surveillance data, frequently without the consent or knowledge of the participants. Sharma and Barua (2023) emphasize the complexity of this issue in the age

of expansive data collection by AI systems. Shum (2023) suggests that safeguarding privacy rights necessitates embedding data protection principles at every stage of AI system development, from data collection to storage and processing. Humphreys et al. (2023) also highlight the need for robust privacy frameworks that keep pace with the rapid evolution of generative AI technologies. To tackle privacy and security concerns in GAI, developers employ privacy-preserving techniques (Majeed, 2023), implement adversarial defense mechanisms (Deldjoo et al., 2021), and enact regulatory measures and policies.

The replication of *copyrighted content* by GAI models such as ChatGPT and GPT-4 raises a challenge to existing intellectual property regulations, as noted by Henderson et al. (2023). These models frequently operate in a gray area of imitation and innovation, blurring the limits of authorship and work in creative professions. Similarly, Feuerriegel et al. (2023) highlight a twofold concern with GAI concerning data governance and intellectual property rights. The initial problem is the AI's ability to unintentionally replicate copyrighted works, even without exposure to the copyrighted content. An example of this would be an AI producing a picture that mimics the Adidas logo without any prior training. The second problem is the AI's ability to create derivative works, posing the original creators' transformative rights risks and prompting intricate legal debates over originality and creative attribution in these systems. Smits and Borghuis (2022) argue that there is an urgent need to reevaluate traditional concepts of intellectual property in light of these advancements to ensure that creators receive protection in the digital age.

AI systems are susceptible to learning and propagating fake news and toxic speech in their training datasets, as Hajli et al. (2022) outlined. These risks worsen when AI-powered social bots disseminate misinformation or adopt language patterns that can amplify harmful activities. The risk that models such as GPT-3 will produce toxic speech that mimics the language of radicals emphasizes how urgently strong data handling governance is needed. McGuffie & Newhouse (2020) underscore that without stringent controls and checks, the propensity for GAI to spread problematic content is significant, highlighting the need for more proactive governance mechanisms.

The current regulatory landscape for GAI is lacking, particularly in financially and economically sensitive contexts. Shouyang et al. (2023) draw attention to the pressing ethical dilemmas, biases, and accountability issues that arise from such a regulatory gap. Oduro et al. (2022) argue that this regulatory gap contributes to economic disparities and creates the

possibility of misuse AI. Amariles & Baquero (2023) contend that ensuring GAI applications' transparency, safety, and ethical adherence remains challenging without comprehensive regulatory frameworks.

Societal and ethical concerns emerge from the potential of AI-generated content to mislead or offend, which is a growing concern (Sullivan and Fosso Wamba, 2022). The ability of AI models to influence user preferences and generate misinformation risks highlights the lack of interpretability (Subhash, 2023). Also, according to Rozado (2023), the results from various tests are uniform, with 14 out of 15 politically oriented tests indicating that ChatGPT's responses lean towards left-wing perspectives, holding political biases. Ashraf et al. (2024) provide examples such as integrating GAI into search engines, which might inadvertently lead consumers to illegal websites, underscoring the need for enhanced data security measures.

The technological aspects of AI systems give rise to technical dangers. These could consist of, but are not restricted to:

1. **Data-related issues:** These risks may be caused by biases in training data, inadequate or poor-quality data, or concerns about data privacy (Higgins et al., 2011). Sensitive information breaches, biased algorithmic decision-making, and privacy abuses are a few examples. In GAI, the training data effectively acts as a selection and conceptualization process similar to identifying 'risk objects.' Just as specific characteristics are emphasized to shape an entity's identity as dangerous, training datasets in AI may inadvertently emphasize social biases. These biases are not inherent to the data but arise from the societal contexts and cognitive frameworks of those who compile and curate the datasets. Consequently, GAI models trained on such data may generate outputs that perpetuate these biases, creating a significant risk of reinforcing harmful stereotypes and social prejudices in various settings. This scenario underscores the urgent need for more diverse and carefully vetted training datasets to mitigate the risk of bias amplification in AI-generated content.
2. **Algorithmic issues:** These risks could result from unanticipated outcomes of the optimization process or biases in the algorithms themselves, weaknesses in AI models, or biases in the algorithms themselves (Fletcher et al., 2021; Dieber & Kirrane, 2020). Examples include unfair treatment of certain groups, over- or under-representation of particular features, or reinforcing of preexisting biases in society.

3. Human factors: Risks associated with AI may stem from human decision-making or behavior, including the misinterpretation of AI-generated results, excessive reliance on AI systems, or the malicious use of AI technologies (Jones, 2015; Karkar et al., 2017; Brundage et al., 2018). Examples include disinformation campaigns driven by AI, AI-enhanced surveillance, and the development of autonomous weapons.
4. Lack of accountability: The issue of accountability in AI involves assigning responsibility for decisions made or actions taken by AI systems (Dignum, 2017). Given AI systems' complexity and frequently obscure nature, determining the reasons behind specific outcomes or decisions is often difficult. This challenge in pinpointing responsibility can result in a lack of accountability. This issue becomes especially critical when decisions have substantial consequences, such as healthcare, autonomous driving, and legal judgments.
5. Errors in algorithms: Algorithms are the foundation of AI systems, and errors in these algorithms can lead to unpredictable, often adverse effects (Wang et al., 2019). Coding mistakes, algorithmic biases, or problems with the data used to train the AI—such as skewed or inadequate data—can all cause this. These errors result in inaccurate forecasts or choices, which can have negative real-world effects like monetary losses or bodily injury.
6. Privacy violation: AI systems, especially those that employ machine learning, typically need vast datasets for training and functionality. This extensive data usage can raise privacy issues, especially when it involves sensitive or personal information (Sarathy & Robertson, 2020). In the absence of adequate protective measures, there is a risk of data breaches, misuse of personal data, or infringements on privacy, which could harm individuals or groups whose data are utilized.
7. Autonomy: The influence of AI on human autonomy poses a major issue (Bryson, 2018). With AI becoming increasingly prevalent in everyday life, there are worries that these systems may start to dictate or replace decisions that humans traditionally made. This situation could undermine personal autonomy, leaving individuals with less control over their choices and actions. In severe instances, there could be a risk of becoming too dependent on AI, leading individuals or organizations to rely on AI systems when human judgment is crucial.
8. Safety: The safety of AI systems is a crucial issue, particularly in contexts where AI has direct physical impacts, such as in autonomous vehicles or robotics (Amodei et al., 2016). Safety concerns also extend to AI systems' potential for harm through errors or

malicious use. For example, an AI system could cause damage if it makes a mistake in a healthcare context or is used maliciously for cyberattacks. Ensuring the safe use and operation of AI systems is a significant challenge in the field.

Content Generated

GAI models, such as ChatGPT, are known to sometimes produce 'hallucinations,' a term used to describe the generation of false sources or facts (McIntosh et al., 2023). This issue can significantly affect the trustworthiness and accuracy of their outputs, especially in structured documents and scholarly writing. In business environments where ChatGPT is valued for its efficiency and creativity, these inaccuracies require careful handling. Alkaissi & McFarlane (2023), Stokel-Walker (2023), and Fui-Hoon Nah et al. (2023) emphasize the need for well-defined guidelines from publishers and users to minimize the risk of misinformation caused by such hallucinations, thereby ensuring the authenticity of authorship and the appropriate use of generated content.

There has been significant scrutiny over the content produced by GAI systems like ChatGPT for potential biases. Hartmann et al. (2023) have identified a left-leaning bias in the portrayal of political subjects and responses related to political figures. The nature of these biases is multifaceted, often stemming from the training protocols used by human trainers that guide the AI, leading to a propensity for stereotypical behavior toward different social groups. The manifestation of bias varies across categories, with substantial disparities noted in content relating to religion and politics, as further discussed by Rozado (2023), McGee (2023), Kocoń et al. (2023), and Busker et al. (2023). This issue highlights the complex challenge of ensuring that GAI operates fairly and equitably, respecting diversity without perpetuating discrimination.

GAI's capability to produce content has also been linked to the propagation of harmful behaviors, as evidenced by studies highlighted by Subhash (2023) and Rozado (2023). These studies suggest that ChatGPT can produce statements that may inadvertently encourage actions such as unnecessary medication use, mass shootings, and suicide. Furthermore, disparities in content moderation systems, as implemented by developers like OpenAI, reveal biases in automated moderation across demographic lines. Such biases become evident in the treatment of negative remarks, which are reportedly flagged more frequently as hateful when directed at women, suggesting a gender bias in moderation practices. This underscores the need for more

sophisticated and balanced moderation mechanisms that can identify and mitigate the generation of dangerous content while ensuring fairness and impartiality.

Social and environmental issues

The operation of data centers, which are critical to the functioning of generative AI technologies, poses significant environmental challenges. As noted by Çelebi (2023), data centers are currently estimated to consume around 200 TWh of electricity annually, which is expected to surge by approximately 15 times by 2030. Andrae and Adler (2015) have highlighted that these centers represent some of the largest and most rapidly expanding electricity and water consumers in the United States. This consumption is not trivial; for instance, Sverdlik (2016) points out that a midsize data center can use between 80 to 130 million gallons of water each year for cooling purposes alone, comparable to the water needs of three hospitals or two golf courses.

The advent of AI technologies has been met with increasing concern over the potential displacement of human jobs, contributing to employee anxiety regarding job security. As Nazareno & Schiff (2021) describe, this anxiety can significantly affect stress levels, organizational commitment, and overall productivity within the workforce. In the healthcare industry, for example, the efficiency of AI is both a boon and a bane; Pesapane et al. (2018) cite that while a typical radiologist might read 50–100 X-rays in a day, an AI-supported system could interpret 10–100 times more images within the same timeframe. While such advancements in AI capabilities promise unparalleled improvements in operational efficiency, as Khan (2021) suggests, they also raise concerns about the future of employment and the need for strategies to manage the transition for affected workers.

5.1.2.2 Object at risk

4.1.2.2.1 Definition

Unlike risk objects, which we identify due to their potential for danger, we recognize objects at risk for their susceptibility to loss, vulnerability, and subsequent need for protection. These objects are valued for their present worth and potential enduring significance, necessitating careful management and safeguarding.

The process of identifying something as an object at risk involves not only recognizing its value but also actively choosing to preserve and protect it. For instance, in the example of managing dangerous dogs, children are protected from harm (objects at risk) just as dogs are shielded from negligent owners. This dynamic reflects a broader societal or group consensus on what is valuable and how such values are prioritized and protected, often leading to policies that reflect these priorities.

However, what is considered valuable—and thereby what becomes an object at risk—is not static but subject to change based on cultural, social, and temporal contexts. This makes evaluating and protecting objects at risk a dynamic and continually negotiated process within society. For example, regulations restricting dogs in parks to protect park users can clash with the rights of dog owners, illustrating the complex negotiation of values and the protections afforded to objects at risk.

4.1.2.2.1 Object at risk in GAI context

Table 21 illustrates the four dimensions of GAI: individual, organizational, societal, and environmental.

Dimension: Object at risk			
Sub-dimensions	Categories	Description	References
Individual	Human trust	. GAI erodes trust by automating decisions without transparency or fairness. . Deepfakes and personalized content challenge digital trust and autonomy. . Human involvement in AI development helps maintain a human-centered approach.	Ray (2023); Limna et al. (2023). Dunn et al. (2023); Fui-Hoon Nah et al. (2023); Shneiderman (2020);
	Disinformation	. AI spreads disinformation, impacting public opinion and democracy. (Example: Election interference).	Floridi (1996, 2011); Grinberg et al. (2019).
	Physical safety	. AI-related accidents, injuries, and deaths highlight safety risks. . The danger is prevalent in environments with operational AI machinery. (Example: A man ended his life after an AI chatbot encouraged him to sacrifice himself to stop climate change).	Oravec (2022); Huggler (2015); El Atillah (2023)
	Privacy	. AI raises privacy issues with potential misuse of collected personal data. . Concerns include unauthorized surveillance and data security breaches.	Sivan and Zukarnain (2021); Wu et al. (2023); Ray (2023); Limna et al. (2023).
	Psychological well-being	. AI impacts psychological well-being via social media algorithms and surveillance. . AI contributes to addiction and anxiety, notably among adolescents.	Twenge et al. (2018).
	Autonomy and freedom	. AI erodes autonomy by influencing choices and behaviors. . Concerns heightened in online advertising and content personalization	Susser et al. (2020); De Vries (2020).
	Intellectual property	. AI models like ChatGPT replicate copyrighted content, prompting a reevaluation of intellectual property laws. . Challenges traditional authorship concepts and complicates content ownership determination.	Ballardini et al., 2019; Chaka, 2023
	Quality of life	. AI impacts quality of life through job displacement, economic instability, privacy issues, and mental health concerns. . AI contributes to global warming and resource depletion, threatening sustainable living conditions.	Denecke et al. (2021); Frank et al. (2021)
Organisation	E-commerce	. AI in e-commerce incurs high initial costs for training, maintenance, and integration. . These costs pose challenges in the competitive online consumer engagement landscape.	Arman and Lamiya (2023); Wael Al-Khatib (2023).
	Customer service	. ChatGPT's impact on customer service raises privacy and personalization concerns. . Effectiveness of AI chatbots in emotional support roles is questioned.	Arman and Lamiya (2023).
	Reputation	. Incorrect AI implementation can damage an organization's reputation due to biased algorithms causing discrimination. . These biases often result in negative publicity.	Akter et al. (2020); McGuire & De Cremer (2022).
	Business models	. AI disrupts existing business models and enables new data-driven models. . Influences product and service development, market expansion, and competitive advantage.	Fichman et al. (2014); Dana et al. (2022).
	Organization's innovative capacity	. Enables new product and service development . Opens new markets . Drives competitive advantages through data-driven insights and decision-making	Dana et al. (2022).
	Strategic decision-making	. Improves efficiency and accuracy . Provides data-driven insights for informed decisions . Over-reliance might diminished human critical thinking . Difficulty in understanding decision-making processes	Mayer et al. (2020); Grewal (2021).
Society	National security (e.g., cyberattack)	. Raises cyberattack concerns . Increases potential for military escalation . Shifts global power dynamics	Buchanan (2020); Sayler (2019); (NSCAI, 2021).
	Political stability	. GAI has the capacity manipulate public opinion and democratic processes . Raises potential for social unrest . Use in surveillance by authoritarian regimes undermines democratic norms	Howard et al. (2018); Murphy (2017); Kosta (2020); Feldstein (2021).
	Economic stability	. Increases financial market volatility through automated trading algorithms . Exacerbates wealth disparities . Concentrates AI-generated wealth among the affluent and educated	Groth (2011); Mollemans (2022); Arumugam et al. (2023); Vinuesa et al. (2020).
	Social cohesion	. AI-driven social media algorithms shape cultural norms and practices . Cross-cultural management in global business benefits from AI technologies . Profound and unpredictable impact	Won et al. (2017); Arif (2023).
Environment	Global warming	. The energy-intensive nature of AI contributes significantly to global climate change, . Impacts ecosystems, wildlife habitats, and human societies . Results in increased temperatures, rising sea levels, and more severe weather events	Babu & Mohan (2018); Fuso Nerini et al. (2019).
	Resource depletion	. Extensive energy requirements strain energy resources . Contributes to broader challenges in sustainable development	Khattak et al. (2020).
	Waste generation	. Particularly evident in critical sectors like nuclear power . Leads to increased waste generation and environmental concerns . Vulnerabilities to cyber attacks exacerbate environmental risks	Poornima (2022).

Tableau 21 : Object at risk dimension (source: authors)

Individual level

The trust of individuals in GAI is being critically undermined as these systems often make decisions and automate processes without clear adherence to essential ethical principles such as transparency and fairness. This erosion of trust is significant in the context of GAI's capability to fabricate deepfakes and personalized content, which poses severe challenges to the authenticity of digital information. Ray (2023) emphasizes that such capabilities threaten individual autonomy and the decision-making capacity in the digital age, where differentiation between real and synthetic is increasingly blurred. Dunn et al. (2023) advocate for the inclusion of humans in the AI development process to ensure ethical considerations are not overlooked, supporting Shneiderman's (2020) call for human-centered design in AI as a way to maintain a necessary check against the unbridled autonomy of AI systems and to ensure that they serve to augment rather than replace human capabilities.

AI systems' ability to propagate disinformation represents a direct threat to public opinion and democracy. Floridi (1996, 2011) has long discussed the ethical implications of information quality and how misinformation can erode the very fabric of a knowledgeable society. Grinberg et al. (2019) highlight the heightened risk during political events like elections, where AI-facilitated disinformation campaigns can potentially influence outcomes and destabilize democratic processes. This level of impact calls for an immediate and strategic response regarding technology design and education, media literacy, and robust legal frameworks to safeguard the public against malicious information campaigns.

The integration of AI into physical spaces introduces significant safety risks. Although designed to increase efficiency and reduce human labor in hazardous environments, AI-operated machinery and robots can inadvertently cause accidents, injuries, or even fatalities. Oravec (2022) examines the consequences of replacing human oversight with AI, advocating for stringent safety protocols. Huggler (2015) provides a practical perspective, exploring the implications of these risks in various industries, suggesting that without comprehensive safety measures, the operational environment could become increasingly perilous for individuals in proximity to AI systems.

AI's pervasive data collection practices raise substantial privacy concerns, as noted by Sivan and Zukarnain (2021). These systems can amass vast quantities of personal data, leading to potential abuses such as unauthorized surveillance and security breaches. Wu et al. (2023)

argue that privacy is not merely an individual concern but also a societal one, as the collective understanding of privacy shapes democratic values and norms. Ray (2023) and Limna et al. (2023) highlight that maintaining privacy necessitates solid legal protections and ethical AI design, prioritizing data security and individual rights.

The psychological implications of AI on individuals, especially within the context of social media and surveillance, are profound. Twenge et al. (2018) document the correlation between increased screen time, fueled by AI-curated content, and rising levels of addiction and anxiety, particularly among adolescents. This suggests that while AI has the potential to connect and inform, it also has the capacity to harm mental health, necessitating a balanced approach to the design and regulation of AI-driven platforms to safeguard against such unintended consequences.

The rise of GAI systems, especially in the context of decision-making and recommender systems, introduces subtle yet profound challenges to personal autonomy and freedom. These systems, through their design to predict and influence user choices, could potentially lead to a covert form of control over individual behaviors and preferences. Susser et al. (2019) raise concerns over the potential of these systems to manipulate choices in nuanced ways that may not be immediately apparent to users, particularly in sensitive areas like online political advertising and personalized content generation. The concern is that these AI systems may reinforce echo chambers or nudge users toward certain behaviors, thus undermining the autonomy of choice. De Vries (2020) elaborates on this issue, suggesting that personalization algorithms could limit exposure to diverse perspectives, effectively shaping an individual's worldview. This erosion of autonomy is not merely a theoretical concern but has practical implications, as it affects how individuals make informed decisions about their health, finances, and social interactions. In the pursuit of enhancing user experience, it's critical to strike a balance that respects individual agency, ensuring that users are not unwittingly subjected to an AI's agenda. Discussions surrounding this challenge revolve around the ethical implementation of AI, where the user's autonomy is paramount, and algorithms are designed to facilitate, rather than dictate, user choice.

The advancement of GAI models like ChatGPT highlights emerging complexities in intellectual property (IP) law. Ballardini et al. (2019) note that the capabilities of AI to generate original content, such as literature, music, and art, necessitate a thorough reevaluation of IP laws. The current legal frameworks, established in a pre-AI era, are premised on human

authorship and creation, leading to various legal and ethical complexities when determining the ownership of AI-generated content. Chaka (2023) points out the need for IP laws to evolve, accommodating the unique nature of AI-generated works that challenge traditional concepts of authorship and creativity. This includes considering whether AI-generated content should be copyrightable, who holds the copyright if it is, and how IP rights are enforced in a digital environment where AI-created content can increase rapidly. The questions AI raises in the realm of IP are profound, touching on the nature of creation, innovation, and the economic value attributed to original works. As AI continues to produce content that rivals human creativity, policymakers, and legal experts are tasked with redefining the parameters of

Organization level

- **E-commerce**

In the realm of e-commerce, significant transformations influenced this sector by emerging internet technologies. The size of the GAI in the e-commerce market, valued at USD 624.51 million in 2022, is forecasted to grow to approximately USD 2,530.89 million by 2032, with a compound annual growth rate (CAGR) of 15.02%³⁰. The rising need for product recommendations fuels this expansion, which is vital in driving e-commerce sales (Shivkumar, 2024).

GAI is one of these emergent technologies and presents a double-edged sword. While they offer innovative approaches to customer engagement and personalized shopping experiences, implementing such technologies carries substantial initial costs. Ghaffari et al. (2024) underline that the dynamic and continuously evolving nature of the e-commerce marketplace complicates the adaptation of relevant product taxonomies. Effective product categorization, crucial for the seamless functioning of these platforms, requires well-labeled initial data for training LLMs. Moreover, the need for continuous human oversight to ensure new products are accurately tagged underscores the dependency on expert intervention, adding to operational costs. The inconsistent categorization by some LLMs—either too specific or excessively generic—can lead to a convoluted taxonomy that obscures product distinctions. As Arman and Lamiya (2023) note, expenses, including training, maintenance, and integration, must be carefully weighed against the potential for revenue growth and enhanced customer interaction. Wael Al-Khatib

³⁰ <https://www.precedenceresearch.com/generative-ai-in-e-commerce-market>

(2023) discusses the challenges these costs pose in a fiercely competitive online landscape, where consumer loyalty is often contingent on the quality and personalization of the shopping experience.

Adding to this complexity, generative AI has implications on digital platforms (including e-commerce platforms like Amazon), where all stakeholders and their interactions can be profoundly transformed. While generative AI facilitates unprecedented levels of personalization—hyper-personalization—allowing complementors (third-party vendors or service providers who offer complementary products or services) to instantly tailor their offerings to individual user preferences, it also poses significant disruption. For example, the introduction of generative AI systems on crowdsourcing platforms has been shown to push complementors to either abandon the platform or shift their efforts to more complex contests to avoid direct competition with these new systems (Lysyakov & Viswanathan, 2023).

- **Customer Service**

The adoption of GAI in customer service, mainly through chatbots like ChatGPT, has stirred debate. According to Arman and Lamiya (2023), while AI chatbots can respond quickly to customer inquiries, they lack the nuanced emotional support that human interaction offers. This deficiency could lead to a degradation in service quality as privacy concerns and the impersonal nature of chatbots might alienate customers who seek a personal touch. The effectiveness of Gen AI chatbots can vary significantly. Even the most advanced Gen AI chatbots may struggle with understanding user input, conveying responses accurately, or providing high-quality and precise information, leading to service failures (Sands et al., 2021). If the chatbot cannot rectify these failures, their negative impact is magnified, posing a substantial risk to a brand's financial performance and reputation (Huang & Dootson, 2022).

According to Ferraro et al. (2024), while GAI-enabled customer service fosters a sense of connection with a brand, it can also lead to consumers feeling more isolated due to the absence of human interaction, exacerbating the risk of service failures. For example, the shift towards self-checkouts in supermarkets has highlighted the importance of human staff in providing meaningful interactions with customers. Dutch supermarket Jumbo has responded by introducing 'slow' and 'chat' checkouts specifically aimed at offering human interaction to the elderly, addressing the growing issue of loneliness (Boyes, 2021).

Ferraro et al. (2024) present six paradoxes regarding integrating Gen AI chatbots with customer experience. Firstly, while Gen AI chatbots can foster a sense of connection between customers and brands, they also have the potential to induce feelings of isolation. Secondly, while Gen AI chatbots may reduce operational costs through more efficient solutions and resource allocation, these savings may come at the expense of societal costs. Thirdly, although Gen AI chatbots can improve the quality of customer service, they lack the empathetic capabilities of human agents. Fourthly, while Gen AI chatbots may enhance satisfaction with customer service outcomes, they can also lead to value degradation as customers become increasingly frustrated. Fifthly, while Gen AI chatbots enable brands to personalize service based on prior interactions, the extensive data collection and analysis may raise privacy concerns. Lastly, while Gen AI chatbots offer tailored service based on prior interactions, the extensive data collection and analysis may also raise privacy concerns.

- **Reputation**

Concerns regarding reputational risks are longstanding for organizations, as evidenced by the *Reputational Risk Report 2015: Reputational@Risk*. This report, based on a survey of over 300 executives, found that 87% considered reputational risk either important or very important. Additionally, it highlighted that 25% of a company's market value is directly associated with its reputation, which can be influenced by various factors, including corporate social responsibility, safety policies, and innovation (Ristuccia & Rossen, 2015, p. 7).

Organizational reputation is another facet vulnerable to the pitfalls of AI. Akter et al. (2020) caution that biases in AI algorithms can lead to discriminatory outcomes, potentially casting a shadow over an organization's public image. McGuire & De Cremer (2022) underscore the reputational damage from incorrect AI implementation, emphasizing that the fallout from such biases can have lasting adverse effects on customer trust and corporate identity. Morris & Brubaker (2024) introduced the concept of a "generative ghost," an AI-powered, agentic representation of a deceased individual that goes beyond static or interactive memorialization by generating novel content and interactions based on the deceased's data and additional metadata and possesses capabilities for autonomous actions in digital or physical realms. The same authors argue that reputational risks of generative ghosts include scenarios where the AI's interactions might negatively impact the memory of the deceased. Such risks can alter perceptions in ways the deceased would not have approved, harm the reputations of relatives,

or negatively affect mental health—for instance, a child learning from the AI that they were less favored. The main categories of these risks are privacy, hallucination, and fidelity risks.

Holweg and his colleague (2022) analyzed 106 cases related to AI controversies, identifying the primary causes of stakeholder concerns and resulting reputational issues. The study revealed that the most prevalent reputational impact resulting from AI failure is the violation of privacy, constituting 50% of the analyzed cases. Additionally, algorithmic bias emerged as the second most common reputational issue, accounting for 30% of the cases. Finally, a significant reputational impact arose from the challenge of explainability, accounting for 14% of the cases. Similarly, Bowen (2020) stresses the role of cybersecurity and AI in influencing an organization's reputation. Yet, the ramifications of AI and cybersecurity breaches on a company's market standing transcend the firm's boundaries and extend into society.

- **Business Models**

AI is a pivotal element for business model innovation (BMI), offering essential momentum for firms to revise their existing strategies (Lee et al., 2019). The recent advent of GAI has notably intensified these discussions. Kanbach et al. (2023) propose a framework comprising six assertions to articulate the potential effects of GAI on innovation activities, work environments, and information infrastructures. The initial assertion posits that GAI democratizes innovation, enabling innovators across various regions and social strata to access necessary expertise, technology, and resources. The second assertion focuses on the sweet spot of GAI, particularly in business models that integrate factual knowledge with creative processes, which are prime for rapid innovation and significant value capture. The third suggests that GAI typically impacts businesses by extending current products and services, emphasizing innovations that are more immediately accessible. The fourth assertion highlights the profound influence of GAI on the roles of white-collar professionals, whose jobs may be completely transformed, replaced, or enhanced. The fifth discusses the evolving skill sets necessary in the GAI-influenced workplace, where roles shift from creators to editors, suggesting a model of human augmentation rather than outright replacement. The final assertion considers the integration of GAI, stating that the focus for companies is not whether to utilize GAI but how it will be incorporated or built into their existing data infrastructures.

Lecocq et al. (2024) explore the transformative potential of generative technologies on business models (BM), identifying six significant ways these technologies can alter both the

process and the outcomes of BM. Firstly, generative technologies facilitate the production and analysis of data crucial for BM design's inspiration, definition, and testing stages. Secondly, they can rapidly generate a vast array of ideas, making them accessible even to those with minimal management or strategy knowledge. Thirdly, these technologies aid entrepreneurs and managers in consistently navigating between the broad strategic concepts and detailed elements of a BM. Fourthly, they enhance analogical reasoning, applying structured knowledge from familiar to novel domains. Fifthly, they bolster reflexive analysis during BM design. Lastly, they blur the distinctions between BM design and implementation, integrating the testing and execution phases. The same authors identify four key risks of using generative technologies in BM design. Firstly, these technologies often rely on existing data and models, which can limit their ability to engage in highly abstract and speculative BM development and may introduce biases due to a restricted data set, resulting in a limited range of proposed BMs. Secondly, the generative technologies may become uncreative and imitative, with users potentially losing track of the origins of ideas and BMs due to the technologies' operation in a legal and creative no man's land. Thirdly, there is a risk that using generative technologies to test and enhance BMs could inadvertently aid competitors if the same technologies are used industry-wide, leading entrepreneurs to withhold or falsify data input to protect their interests. Lastly, the ease and speed of creating detailed BMs might encourage a cycle of constant redesign without implementation, a phenomenon likened to Penelope syndrome, where continuous creation supersedes actual execution.

Fichman et al. (2014) and Dana et al. (2022) elucidate how AI can catalyze the development of innovative, data-driven business models, altering the very fabric of product development, market expansion, and competitive dynamics. This transformation is not simply evolutionary but often revolutionary, requiring companies to adapt rapidly or risk obsolescence.

- **Organization's innovative capacity**

In exploring the intersection of GAI and organizational innovation, it is essential to recognize both the potential benefits and the accompanying risks or negative impacts that such technologies can introduce. Scholars like Dana et al. (2022) and Mikalef and Gupta (2021) illustrate how AI-based recommendations have facilitated the design and creation of new products, thereby advancing business model innovation. Similarly, Lee et al. (2019) support this view by showing the instrumental role of AI in enhancing innovative business models. Yet, as AI technology evolves, it introduces fresh management challenges and workplace

implications. AI differs in how it limits, augments, or replaces human roles when integrated into organizational settings, displaying a range of complex behaviors that can lead to unforeseen outcomes. Furthermore, AI fundamentally questions traditional distinctions between human and machine capabilities, presenting significant shifts in both areas (Benbya et al., 2021).

A particularly poignant example of potential pitfalls is Nadia, a chatbot developed by the Department of Human Services in Australia. Intended to aid the disabled population, it was discontinued due to public concerns over its efficacy, underscoring AI's limitations and adverse impacts on organizational innovation (Benbunan-Fich et al., 2020). This case exemplifies how AI applications can fail to meet user needs and expectations despite the potential benefits, leading to discontinuation and reputational damage.

Moreover, von Wolff et al. (2022) highlight a gap in design knowledge for chatbots that support internal business processes. The focus in existing literature and applications has predominantly been customer-facing, neglecting the chatbots' potential to impact digital transformation efforts internally.

Holmström & Carroll (2024) provide a different perspective to this discussion and offer a typology of innovation strategies impacted by ChatGPT-like technology. Four strategies result from this framework, defined along two key dimensions: automation and augmentation. These strategies are Automated Assistance (high automation, low augmentation), Traditional Tool (low automation, low augmentation), and Assisted Augmentation (high automation, high augmentation). This typology helps with developing flexible strategies to keep up with the quick changes in the digital ecosystem, in addition to helping to understand the various ways that AI may be leveraged for innovation. Furthermore, Holmström and Carroll (2024) suggest that incorporating prompt engineering into organizational workflows can maximize the potential of tools like ChatGPT for innovation. By meticulously designing prompts, AI becomes a formidable ally in problem-solving, research, content generation, decision support, and design, thus enhancing creativity and efficiency across various business operations.

This complex interaction between GAI and organizational innovation, characterized by both transformative potential and significant challenges, underscores the need for ongoing adaptation and careful management of technological implementations in business settings.

- **Strategic decision making**

The integration into organizational strategic decision-making of GAI can significantly transform possibilities and raise some substantial challenges. It is right that GAI would make decisions simpler and broad-based on data analyzed, although simultaneously exposing them to risks.

First, decision-makers who rely solely on GAIs are likely to exhibit automation bias, where they place too much trust in automation results without questioning them (Bracci, 2023). This scenario may occur if managers fail to check the recommendations made by AI against long-term goals set by the company or even moral standards.

Secondly, AI systems often use ‘trained’ data sets but can be inherently prejudiced or incomplete (Budhwar et al., 2023). Therefore, these constraints may lead an AI to distort its outputs, resulting in bad, discriminatory, or ethically disputable decisions. The implications of such consequences are alarming in strategic settings that would impact corporate directionality and employee motivation.

Thirdly, AI systems may lack transparent decision-making processes. Thus, managers sometimes do not understand these decisions as being formulated indistinctly. Consequently, strategic planning might have many problems, such as loss of sense in the absence of means of evaluation and subsequent modification. Furthermore, GAI can sometimes create a transparency issue in decision-making processes. As decision logic becomes more complex and self-evolved, it can be challenging for human managers to understand how certain decisions are derived (Chen & Esmaeilzadeh, 2024). This lack of clarity can be detrimental to strategic planning, as it impedes the ability to evaluate and adjust strategies effectively.

Society level

- **National security**

The recent report released by the Centre for Emerging Technology and Security (CETaS) at The Alan Turing Institute shows how significant the unintended threats that GAI causes to the UK's national security are. In the past, there has been more emphasis on deliberate misuse of GAI, such as in disinformation campaigns, cyberattacks, and other malicious activities. This kind of misuse is projected to accelerate in pace and scope, thereby exposing

many more members of society to these dangers. Nonetheless, the report also calls attention to the non-intentional risks stemming from the misuse and experimentation with GAI tools, particularly in critical national infrastructure, public services, and private individuals excited by AI advancements. This includes the potential for excessive risk-taking driven by an overreliance on AI outputs and a fear of missing out on the latest technological developments. The report stresses policymaking's importance in managing intentional and unintentional threats to maintain national security (Janjeva et al., 2023).

The negative impact of AI on *national security* has been a growing concern in recent years. As AI technologies become more advanced, they may be used to develop more sophisticated weapons and autonomous systems, raising the risk of accidental or intentional harm to civilians and military personnel. Several reports and studies have highlighted AI's potential risks to national security. For instance, a report from the US National Security Commission on Artificial Intelligence (NSCAI) warned that China's use of AI in its military could pose a significant threat to US national security (NSCAI, 2021). Another report from the RAND Corporation noted that AI could be used to develop more advanced autonomous weapons that could pose a significant threat to civilians (Lair, 2020).

Furthermore, AI-enabled cyberattacks could have severe consequences for national security, as they may be more challenging to detect and defend against. Eling and Schnell (2016) define Cyber risk as:

Any risk emerging from the use of information and communication technology (ICT) that compromises the confidentiality, availability, or integrity of data or services. The impairment of operational technology (OT) eventually leads to business disruption, (critical) infrastructure breakdown, and physical damage to humans and properties. Cyber risk is either caused by natural disasters or is man-made, where the latter may emerge from human failure, cyber criminality (e.g., extortion, fraud), cyber war, or cyber terrorism. It is characterized by interdependencies, potential extreme events, high uncertainty concerning data and modeling approach, and the risk of change. (p. 12)

There have been several high-profile cyberattacks on national security in recent years, including:

4. SolarWinds hack: In 2020, a cyber espionage campaign targeted several US government agencies and other organizations using malware implanted in the SolarWinds Orion software (McGuire & Surrey, 2021). The cyberattack on SolarWinds compromised over 15,000 networks, including those of several Fortune 500 companies in the IT sector, such as Cisco, FireEye, Intel, and Microsoft (Krebs, 2021).
5. Colonial Pipeline attack: In May 2021, the Colonial Pipeline, which supplies gasoline and other fuel to much of the US East Coast, was hit by a ransomware attack that forced the pipeline to shut down for several days. The attack was attributed to a criminal hacking group called DarkSide (Bing & Kelly, 2021; Hobbs, 2021).
6. Microsoft Exchange Server hack: In early 2021, a cyber espionage campaign targeted thousands of Microsoft Exchange Server installations worldwide, exploiting software vulnerabilities to gain access to sensitive information. The attack was attributed to a Chinese government-sponsored hacking group (Novet, 2021).

These cyberattacks demonstrate the significant and ongoing threat that cyber threats pose to national security and the need for robust cybersecurity measures and proactive response strategies. On the other hand, Müller and Bostrom (2016) estimated that AI systems will attain human-level intelligence by 2075 and that some scientists are concerned that continuing the development of AI toward superintelligence could be detrimental to humanity.

The last point to mention in the national security context is the rising use of AI in the armed forces and for military purposes. Rashid and his colleague (2023) illustrate seven patterns for using AI in the military and their application, described in Table 22.

Pattern name	Definition	Application in militart
Hyperpersonalization	The hyperpersonalization pattern uses machine learning in forming a unique profile which learns and adjusts for diverse roles over time, for example, showing relevant content, suggesting suitable products or offering individualized advice.	GANs for personalized soldier training
Pattern and anomalies recognition	A powerful computational technique to objectively evaluate visual data	Object detection in military surveillance (e.g., Raven drone)
Conversational Pattern	This involves talking and sharing data via images, text and sound across different channels done through various mediums such as graphics, voice and audio between man and artificial agent.	Chatbots for military communication (e.g., US Army's Sgt. Star)
Predictive Analytics	The use of data to anticipate future trends and events is called AI.	Predictive maintenance for military equipment (e.g., F-35)
Autonomous Systems Pattern	The most sophisticated kind in AI where the processes are automated that have the intelligence for making computers, bots, and systems operate without humans interacting with them.	Self-driving military vehicles (e.g., KF51 panther)
Goal-driven systems	The agent autonomously operates within its defined goals and purpose, actively identifying issues rather than randomly detecting anomalies.	Autonomous drones (e.g., MQ-9 reaper)

Tableau 22 : Patters of AI in military

Modern-day war and defense have been transformed significantly by introducing AI into national security frameworks. Considering surveillance and autonomous guns, Buchanan (202) examines the possibility of AI applications enhancing cyberattacks, causing sharp military escalations. In addition to concerns about traditional warfare methods being wiped out, Sayler (2019) emphasizes that new forms of cyber warfare through AI technology could disrupt major infrastructures, crippling nations even when left untouched physically. Moreover, the possibility of a realignment of global power dynamics has been alluded to by the US National Security Commission on Artificial Intelligence, NSCAI (2021), a situation where many nations compete using AI for military supremacy; thus, this has plunged ethical problems about using AI in military operations with an urgent request to put international regulation mechanisms into existence.

- **Political stability**

Among the most crucial risks associated with AI is *political stability*. AI can be used to manipulate public opinion and undermine democratic processes, which can lead to social unrest and political instability (Howard et al., 2018). Moreover, the application of AI in surveillance brings in a new age of surveillance, so-called “algorithmic surveillance” (Murphy 2017, p. 225), which can lead to violations of individual rights and freedoms, creating further tensions within societies (Kosta, 2020). Several reports and studies have highlighted AI's potential risks to political stability. For instance, a Carnegie Endowment for International Peace report noted that authoritarian regimes use AI to enhance their surveillance and censorship capabilities, leading

to greater political repression and a deterioration of democratic norms (Feldstein, 2021). Another report from the Oxford Internet Institute³¹ found that AI-enabled social media manipulation is increasingly being used to spread disinformation and propaganda, which can significantly impact public opinion and political stability (Howard & Kollanyi, 2016). Furthermore, organizations might leverage AI to shape political and public opinions, as demonstrated in prior studies where AI, particularly social bots, influenced public opinion during political events, such as the use of automated Twitter accounts to manipulate discourse during the 2016 US presidential election campaign (Bessi & Ferrara, 2016).

- **Social cohesion**

As AI technologies become more pervasive, they can significantly impact social cohesion. According to Friedkin (2004), “Social psychological definitions of social cohesion have emphasized individuals’ membership attitudes and behaviors” (p. 411). For instance, AI can bridge or deepen social divisions depending on its design and application. Automated systems that rely on biased data can reinforce existing inequalities and create a more divided society (Chouldechova, 2017). However, AI can foster greater understanding and unity when used to connect individuals across social, cultural, and geographical barriers.

GAI systems can also have profound influences on *culture*. They can often shape cultural norms and practices in unpredictable ways (Arif, 2023). For instance, AI-driven social media algorithms can influence what cultural content gets seen and shared, potentially skewing cultural norms and trends over time (Oladoyinbo et al., 2024). Furthermore, AI technologies that automate creative tasks, like music or art creation, could fundamentally reshape cultural industries.

AI has the potential to either enhance or exacerbate *inequities* within society. For instance, it could either perpetuate bias and discrimination if these biases are present in the data it's trained on or be used to promote inclusivity and diversity, depending on its design and application (D’Cruz et al., 2022).

³¹ <https://www.ox.ac.uk/news/2021-01-13-social-media-manipulation-political-actors-industrial-scale-problem-oxford-report>

Economic stability level

AI's role in economic spheres is multifaceted, having profound implications for financial markets and wealth distribution. Oluwaseyi (2024) highlights four primary risks of implementing AI in financial systems. Data privacy and security risks include increased exposure to unauthorized access and cyberattacks due to the extensive financial data these systems handle, potentially leading to significant financial and reputational damage. Model risk arises from biases in training data that can produce discriminatory results and the inherent lack of transparency in AI algorithms, complicating the understanding and oversight of decision-making processes. Systemic risks involve the growing dependence on AI-driven systems, which might amplify systemic vulnerabilities in financial markets, alongside the potential for AI to be exploited for market manipulation. Lastly, regulatory compliance risks concern the challenge of aligning AI applications with a complex array of financial regulations, posing significant compliance challenges.

Chowdhury et al. (2024) discuss the potential for AI-driven automated trading algorithms to increase market volatility, creating rapid fluctuations that can challenge economic stability. According to a recent report, after interviewing algorithmic trading solution vendors and buy-side traders, it is estimated that the worldwide equity trading commission wallet share for algorithmic trading services increased from roughly 30% in 2018 to 66% in 2022 (Mollemans, 2022). For instance, AI and ML systems that utilize identical data or employ similar tactics might cause collective movement, potentially intensifying market instability (Ait-Sahalia & Saglam, 2024). Additionally, the risk of AI-powered trading systems reacting to incorrect or deceptive signals, triggering severe market fluctuations, represents a substantial systemic threat (Jacob Leal et al. 2016). algorithmic trading services increased from roughly 30% in 2018 to 66% in 2022 (Mollemans, 2022). Arumugam and colleagues (2023) found that buy-side algorithmic traders and high-frequency traders who maintain large client portfolios and hold securities for short intraday periods with low inventory, respectively, can increase volatility through excessive directional and non-directional trading. Within this regard, Algorithmic stablecoins, which rely on financial engineering, algorithms, and market incentives to maintain a peg to a reference asset, are not genuinely stable and are constantly exposed to risk (Clements, 2021). Furthermore, the concentration of wealth among those with access to AI technology is a growing concern. Large tech companies with the resources to invest in AI

already benefit from the technology's implementation, while smaller businesses and individuals may struggle to keep up (Kend & Nguyen, 2020).

In another manner, Goyal and Aneja (2020) highlight that AI and automation disproportionately impact low and medium-skill jobs, increasing unemployment and widening the income gap between these workers and their higher-skilled counterparts. The study also finds that developing nations experience more severe income inequality from AI advancements, as evidenced by higher Gini coefficients than developed countries.

Industry reports suggest that GAI could raise global gross domestic product (GDP) by 7% and replace 300 million jobs of knowledge workers (Goldman Sachs 2023). Also, In 2021, the Global GAI Market Size reached USD 7.9 Billion, and it is projected at a Compound Annual Growth Rate (CAGR) of 34.3% from 2022 to 2030, reaching a market size of USD 110.8 Billion by 2030 (Acumen Research and Consulting, 2022). McKinsey reports that GAI is set to drive a new wave of productivity, potentially creating between \$2.6 trillion and \$4.4 trillion in value across various industries (Chui et al., 2023).

Environmental level

To contextualize the environmental implications of AI and GAI, it is essential to examine the infrastructural underpinnings of these technologies within the Information and Communication Technology (ICT) and Internet of Things (IoT) sectors. The increased computational demand in the ICT sector considerably escalates its carbon footprint, with estimates from Freitag et al. (2021) attributing between 1.8 to 3.9 % of global carbon emissions to this sector. Additionally, Das (2019) projected that the energy consumption from manufacturing semiconductors for the Internet of Things (IoT) was around 556 TWh in 2016, with an expected increase to 722 TWh by 2025—an eighteenfold rise. These foundational impacts set the stage for understanding advanced AI systems' even more intensive demands.

The energy-intensive operations required to power advanced AI systems contribute significantly to global climate change. The computational processes underpinning AI, particularly those used in large-scale machine learning, demand substantial amounts of electricity generated from carbon-emitting sources. Schwartz et al. (2020) note that the computational demands for deep learning research have rapidly escalated, doubling every few months and leading to an astounding 300,000-fold increase from 2012 to 2018. This surge in

computations contributes to a significantly large carbon footprint. The study by Strubell et al. (2018) highlighted the significant environmental impact of NLP algorithms during the training phase, discovering that the greenhouse gas (GHG) emissions were equivalent to 300 flights between New York and San Francisco. Later, Patterson et al. (2021) updated these findings, showing a substantial decrease in GHG emissions—by a factor of 100—when considering factors like the data center's location in a low-carbon energy area and sparser deep network architectures. However, it is essential to note that their analysis only accounted for the GHG emissions from operating computers and data centers, omitting the broader lifecycle impacts, including production and disposal.

Blockchain technology, which is decentralized to prevent reliance on a central authority or single point of failure, includes applications like cryptocurrencies. A study highlighted that Bitcoin's electricity usage reached 68.7 TWh in 2020, sufficient to power 7 million U.S. households, and is responsible for emitting 44 MtCO₂. Further research by de Vries (2023) noted that the water consumed by a single Bitcoin transaction could equal that of a backyard swimming pool. The growing use of AI and GAI models raises significant environmental concerns, particularly regarding their substantial water consumption. Sankarn (2023) reports that training an AI model comparable to human brain capacity for one year requires approximately 126,000 liters of water. Additionally, the GPT-3 model, developed by OpenAI, used about 700,000 liters during its training. This level of water use is alarming, especially given the vast number of AI models operational today and the significant water requirements of data centers, which intensify the strain on water resources in already water-scarce regions (George et al., 2023). Thus, the escalating computational demands of AI and GAI significantly contribute to global warming, highlighting the urgent need for integrating sustainable practices within the technology sector.

Beyond its energy consumption, AI's broader role in technological innovation and infrastructure, especially in critical sectors such as nuclear energy, contributes to generating hazardous waste. Poornima (2022) discusses the additional environmental burdens and concerns posed by AI-driven industries, including the physical waste from hardware components and the potential for increased cybersecurity risks that could lead to ecological disasters if critical systems are compromised.

Each of these environmental challenges underscores the importance of considering the ecological costs of AI. It compels the need to integrate sustainability into AI technologies' core

design and application to mitigate adverse environmental impacts. This includes investing in energy-efficient computing, exploring renewable energy sources for data centers, and developing AI systems that optimize rather than exacerbate resource use. As AI advances, it will be paramount to ensure its compatibility with environmental preservation and sustainability.

5.2 Methodology

5.2.1 Data collection

The critical data source for our research was the OECD AI Incidents Monitor (AIM). This decision was motivated by the AIM's comprehensive nature, managed by an expert group and funded by the Patrick J. McGovern Foundation. The AIM is distinguished by its systematic methodology for categorizing AI incidents and its use of modern data collection technologies, making it a dependable source of real-time AI incident data. Including a vast range of instances from many domains enables a comprehensive investigation of AI's effects on society.

5.2.2 Period of study and incident selection

The study examines incidents recorded from 1 November 2022 to 26 November 2023. This time period was chosen to include the post-launch era of major GAI systems such as ChatGPT, offering an overview of evolving GAI technologies and their societal ramifications. By focusing on this timeframe, the study captures a vital period in the rapid development and deployment of GAI technology, ensuring the findings are relevant and timely.

5.2.3 Computational framework

5.2.3.1 Programming language and libraries

Python was selected as the primary programming language due to its robustness and flexibility in handling complex data analysis tasks. Python's compatibility with multiple data analysis libraries and its prevalent use in academic research (Garner et al., 2022) supports its choice:

- **Pandas:** Utilized for data manipulation and cleaning, Pandas are essential for preparing datasets for analysis. It provides robust data structure handling capabilities, crucial for large-scale data operations (Nelli, 2015).

- **Openpyxl:** This library was used to facilitate interaction with Excel files where our incident data was initially categorized and prepared. Its functionality is critical for maintaining data accuracy and integrity throughout the analytical process (Dunn et al., 2016).

5.2.3.2 Integration and utilization of AI models

- **API integration with OpenAI's GPT-4:** The study involves integrating Python scripts with OpenAI's GPT-4 model through a registered API key. This integration enables the utilization of GPT4's advanced analytical capabilities to accurately classify and label AI incidents.
- **Model guidance and data labeling:** A custom 'system message' was coded into our Python scripts to guide the GPT-4 model during auto-labeling. This method ensures that the model adheres to our specific research parameters and enhances the reliability of the output.

5.2.3.3 Validation of AI outputs

To reduce potential errors and improve the accuracy of AI-generated labels, the 'temperature' setting of the GPT-4 model was set to zero. This setting helps prevent the model from generating overly creative or irrelevant responses, thereby ensuring the precision of data categorization.

5.2.3.4 Statistical analysis framework

- **Binary logit model:** This model was employed to analyze the probability of outcomes based on binary dependent variables, represented by the presence or absence of specific incident types. The dependent variables include human trust incidents, disinformation incidents, physical safety incidents, privacy incidents, psychological well-being incidents, autonomy & freedom incidents, quality of life incidents, intellectual property incidents, e-commerce incidents, customer service incidents, reputation incidents, business model incidents, organizational innovation incidents, strategic decision incidents, national security incidents, political stability incidents, economic stability incidents, and social cohesion incidents. The independent variables used in the analysis are data governance issues, content-generated issues, and social and environmental issues..

- **Forward stepwise regression:** This statistical technique was chosen for its efficacy in identifying significant predictors in exploratory research scenarios. As Field (2024) recommended, this approach helps unveil complex relationships within the data, guiding the development of theoretical understandings of the impacts of GAI.
- **Normalization technique:** Min-Max scaling was applied to standardize the input data for logistic regression. This statistical technique adjusts the scale of independent variables to a range of 0 to 1, which is crucial for comparative analysis across different variables and ensuring that the logistic regression model functions optimally.

5.3 Results

5.3.1 Incidents distribution

In Figure 25, the diagram shows how often different risks related to GAI risks derived from the dataset. The main areas from this analysis indicate significant focus areas and their relative impacts:

- **Data governance (77.2%):** This area, which encompasses the policies and practices governing data management and usage within GAI systems, emerged as a critical aspect. The high frequency indicates the importance of robust data governance frameworks to mitigate risks.
- **Content generation (32.8%):** Risks associated with generating digital content via GAI tools highlight the challenges of ensuring authenticity and preventing misuse, such as the spread of misinformation.
- **Societal issues (24.2%):** The impact of GAI on broader societal dynamics, including issues of bias, discrimination, and societal disruption, represents a significant concern.
- **Individual impact (92.8%) and organizational impact (93.1%):** These are the most frequently observed impacts, suggesting that GAI technologies significantly affect personal and organizational domains, warranting thorough scrutiny and tailored risk mitigation strategies.
- **Societal (58.7%) and Environmental (2.1%):** While societal impacts are moderately high, environmental considerations are notably low, indicating a potential area for future research and policy development.

There have been various types of research concerning the incidents of GAI illustrating the impact on individuals and organizations, highlighting the need to better comprehend these issues at personal and professional levels and from community and ecosystem perspectives. In short, due to the complexity of data governance and content generation, all these are linked strongly within socio-environmental contexts; an active approach will be required when dealing with potential risks associated with GAI.

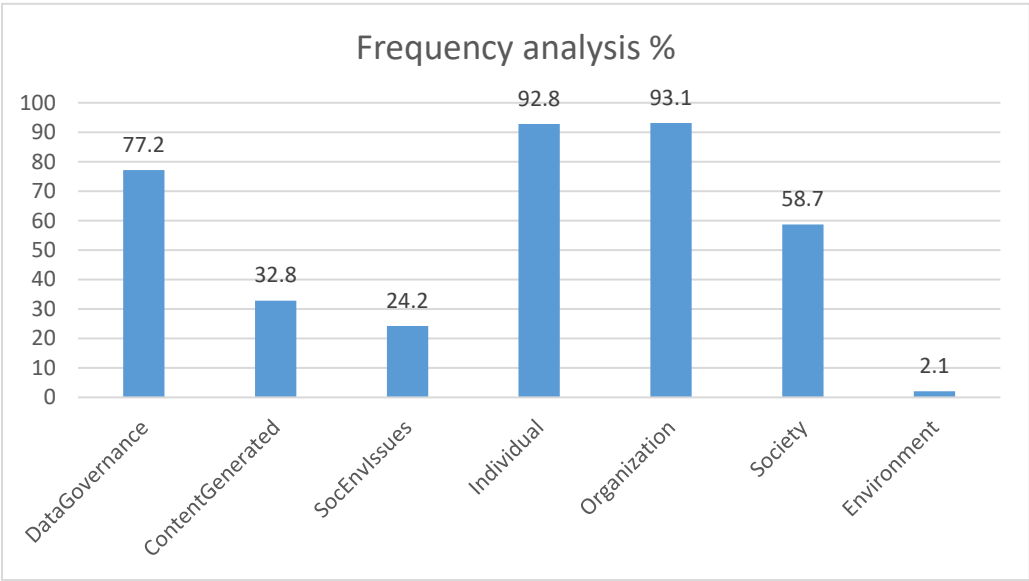


Figure 25: Incidents distribution

5.3.2 Descriptive Analysis

Several key metrics like the number of observations (or frequency of incidents), the mean (average impact per incident on a to 1 scale), standard deviation (variation in impact), and range (minimum and maximum effects) have been included in Table 23.

1. **High-frequency categories:** Categories of ‘observed reputation’ and ‘business models’ show the highest number of observations, with 644 and 628 incidents, respectively, and substantial mean values of 0.751 and 0.732. This suggests that GAI significantly influences organizational reputation and business strategies, reflecting their importance and potential for disruption in these areas.
2. **Significant impact on decision-making:** As shown in Table 23, the category of decision-making has a mean of 0.598 across 513 observations, indicating that GAI

incidents frequently influence key organizational decisions, potentially leading to major strategic shifts or disruptions.

3. **Personal and social implications:** Categories of ‘psychological well-being’ and ‘social cohesion’ with means of 0.375 and 0.484, respectively. These categories highlight the personal and social dimensions where GAI incidents have a noticeable impact, affecting individuals' mental health and the fabric of social interactions.
4. **Environmental and lesser-focused areas:** Global warming and resource depletion categories show minimal incidents (only 8 and 2, respectively) with low mean values, indicating a significant gap in the current assessment and reporting of environmental impacts related to GAI.
5. **Variability in impact:** Categories like disinformation and privacy show high standard deviations (0.451 and 0.439, respectively), indicating a wide variability in the impact of incidents within these categories. This suggests that the severity of effects can vary greatly depending on specific circumstances.

Categories	Observations	Mean	Std. Deviation	Min/Max
Human trust	669	0.779	0.415	0/1
Disinformation	243	0.283	0.451	0/1
Physical safety	170	0.198	0.399	0/1
Privacy	224	0.261	0.439	0/1
Psychological well-being	322	0.375	0.484	0/1
Autonomy and freedom	337	0.393	0.489	0/1
Quality of life	303	0.353	0.478	0/1
Intellectual property violation	187	0.218	0.413	0/1
E-commerce	84	0.098	0.297	0/1
Customer service	100	0.117	0.321	0/1
Reputation	644	0.751	0.433	0/1
Business models	628	0.732	0.443	0/1
Organization innovative capacity	223	0.260	0.439	0/1
Strategic decision-making	513	0.598	0.491	0/1
National security	120	0.140	0.347	0/1
Political stability	101	0.118	0.322	0/1
Economic stability	131	0.153	0.360	0/1
Social cohesion	415	0.484	0.500	0/1
Global warming	8	0.009	0.096	0/1
Resource depletion	3	0.003	0.059	0/1
Waste generation	8	0.009	0.096	0/1
Normlized data governance	239.80	0.279	0.21555	0/1
Normlized content generated	109.67	0.128	0.19750	0/1
Normalized social issues	104.50	0.122	0.21610	0/1

Tableau 23 : Descriptive analysis

5.3.3 Binary logistic regression

According to Andy (2021), the “more crucial to the interpretation of logistic regression is the value of the odds ratio (Exp(B) in the SPSS output), which is an indicator of the change in odds resulting from a unit change in the predictor (p. 270). Thus, in this study, we focus on Exp(B). Andy (2021) further explains that “If the value is greater than 1 then it indicates that as the predictor increases, the odds of the outcome occurring increase. Conversely, a value less than 1 indicates that as the predictor increases, the odds of the outcome occurring decrease” (p. 289).

5.3.3.1 Individual level

Regarding GAI incidents, “Normalized Data Governance issues” signifies the character or cause of danger associated with data governance issues. This examination explores the impact of this risk source on individuals at differing levels, including but not limited to the trust human beings put in each other's disinformation and privacy. Several insights were revealed from the first binary logistic regression with eight dependent variables (human trust, disinformation, physical safety, psychological wellbeing, autonomy and freedom, quality of life, and intellectual property) and three independent variables (normalized data governance, normalized content generation issues, and normalized social and environmental issues), as illustrated in Table 4.

1. Influence of data governance issues

Data governance issues *significantly* impact various dependent variables at the individual level, as demonstrated in Table 24. Specifically, a one-unit increase in data governance issues increases the probability of human trust incidents by 6.47 (Exp(B) = 6.470, $p < 0.001$), disinformation incidents by 13.76 (Exp(B) = 13.759, $p < 0.001$), physical safety incidents by 4.06 (Exp(B) = 4.059, $p < 0.01$), privacy breach incidents by 217.08 (Exp(B) = 217.079, $p < 0.001$), and intellectual property violations by 12.66 (Exp(B) = 12.660, $p < 0.001$). However, this increase shows a *non-significant* relationship with psychological well-being incidents (Exp(B) = 1.287, $p = 0.530$).

Furthermore, a one-unit increase in data governance issues demonstrates a negative relationship with autonomy & freedom incidents ($\text{Exp}(B) = 0.536$, $p = 0.097$), and quality of life incidents ($\text{Exp}(B) = 0.623$, $p = 0.236$). However, these relationships are not *statistically significant*.

2. Influence of Content-generated Issues

Content-generated issues *significantly* influence multiple dependent variables at the individual level, as shown in Table 5. Specifically, a one-unit increase in content-generated issues has a significant relationship with an increase in the probability of human trust incidents by 10.61 ($\text{Exp}(B) = 10.610$, $p < 0.001$), disinformation incidents by 36.93 ($\text{Exp}(B) = 36.925$, $p < 0.001$), physical safety incidents by 6.76 ($\text{Exp}(B) = 6.757$, $p < 0.001$), psychological wellbeing incidents by 86.82 ($\text{Exp}(B) = 86.821$, $p < 0.001$), and quality of life incidents by 5.123 ($\text{Exp}(B) = 5.123$, $p < 0.001$). Conversely, a one-unit increase in content-generated issues significantly decreases the probability of privacy breach incidents by 85.3% ($1 - \text{Exp}(B) = 1 - 0.147 = 0.853$) and intellectual property violations by 0.20 ($\text{Exp}(B) = 0.199$, $p < 0.001$) or 80% ($1 - 0.20$). However, with a one-unit increase in content-generated issues, the probability of autonomy and freedom incidents ($\text{Exp}(B) = 1.856$) increases by 1.856, but the regression is *not significant* ($p = 0.118$).

3. Influence of social and environmental issues

Social and environmental issues significantly affect several dependent variables at the individual level, as illustrated in Table 5. A one-unit increase in social and environmental issues significantly increases the probability of psychological well-being incidents by 6.62 ($\text{Exp}(B) = 6.616$, $p < 0.001$) and autonomy & freedom incidents by 4.54 ($\text{Exp}(B) = 4.542$, $p < 0.001$). However, a one-unit increase in social and environmental issues significantly decreases the probability of disinformation incidents by 63.4% ($1 - \text{Exp}(B) = 1 - 0.366 = 0.634$; $0.634 * 100\% = 63.4\%$) and privacy breach incidents by 0.34 ($\text{Exp}(B) = 0.339$, $p = 0.025$).

Social and environmental issues do *not significantly* affect the probability of human trust incidents ($\text{Exp}(B) = 0.939$, $p = 0.872$), physical safety incidents ($\text{Exp}(B) = 2.117$, $p = 0.081$), quality of life incidents ($\text{Exp}(B) = 23.805$, $p = 0.766$), and intellectual property violations ($\text{Exp}(B) = 0.881$, $p = 0.766$).

Table 24: Results of binary binary logistic regression model at the “individual level”

Y X	Human trust				Disinformation				Physical safety				Privacy				Psychological well-being				Autonomy & freedom				Quality of life				Intellectual property			
	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)
Normalized Data Governance	1.867	0.485	0.000*	6.470	2.622	0.45	0.000*	13.76	1.401	0.448	0.002*	4.059	5.380	0.498	0.000*	217.08	0.252	0.402	0.530	1.29	-0.624	0.376	0.097	0.54	-0.473	0.399	0.236	0.623	2.538	0.438	0.000*	12.660
Normalized Content Generated	2.362	0.610	0.000*	10.610	3.609	0.45	0.000*	36.93	1.911	0.446	0.000*	6.757	-1.917	0.477	0.000*	0.147	4.464	0.458	0.000*	86.8	0.619	0.396	0.118	1.86	1.635	0.417	0.000*	5.132	-1.616	0.488	0.000*	0.199
Normalized social and environmental issues	-0.063	0.389	0.872	0.939	-1.005	0.47	0.033*	0.366	0.750	0.429	0.081	2.117	-1.080	0.483	0.025*	0.339	1.890	0.367	0.000*	6.62	1.513	0.337	0.000*	4.54	3.170	0.358	0.000*	23.805	-0.127	0.425	0.766	0.881
Constant	0.587	0.155	0.000*	1.800	-2.198	0.18	0.000*	0.111	-2.212	0.185	0.000*	0.11	-2.426	0.191	0.000*	0.088	-1.433	0.157	0.000*	0.24	-0.536	0.136	0.000*	0.59	-1.12	0.148	0.000*	0.326	-1.836	0.171	0.000*	0.159
# observation	669				243				170				224				322				337				303				187			
LL (=2LL/-2)	-424.965				-415.562				-404.261				-407.816				-495.674				-560.660				-507.608				-430.769			
Cox R2	0.062				0.200				0.051				0.17941				0.154				0.033				0.109				0.044			
Nagergelle R2	0.095				0.287				0.082				0.26274				0.211				0.044				0.150				0.067			

Tableau 24 : Results of binary binary logistic regression model at the “individual level”

(* $p < 0.05$, indicating statistical significance)

5.3.3.2 Organization Level

In our investigation of the impact of GAI incidents on the organizational level, we observed differentiated effects across various organizational aspects, as illustrated in Table 6.

1. Influence of data governance issues

The influence of data governance issues on various dependent variables at the organizational level is significant, as depicted in Table 6. Specifically, a one-unit increase in data governance issues increases the probability of organizational innovation incidents by 0.15 ($\text{Exp}(B) = 0.149$, $p < 0.001$). However, this increase shows a non-significant

relationship with e-commerce ($\text{Exp}(B) = 2.690$, $p = 0.092$), customer service ($\text{Exp}(B) = 1.011$, $p = 0.984$), reputation ($\text{Exp}(B) = 1.896$, $p = 0.142$), business models ($\text{Exp}(B) = 1.895$, $p = 0.124$), and strategic decision ($\text{Exp}(B) = 1.782$, $p = 0.121$).

2. Influence of content-generated issues

As evidenced in Table 25, content-generated issues significantly influence several dependent variables at the organizational level. A one-unit increase in content-generated issues increases the probability of reputation incidents by 5.11 ($\text{Exp}(B) = 5.106$, $p = 0.002$). Conversely, a one-unit increase in content-generated issues significantly decreases the probability of customer service incidents by 82.6% ($1 - 0.174 = 0.826$, $p = 0.014$), business model incidents by 78.6% ($1 - 0.214$, $p < 0.001$), and organizational innovation incidents by 66.8% ($1 - 0.332$, $p = 0.034$). However, content-generated issues do not significantly affect the probability of e-commerce incidents ($\text{Exp}(B) = 0.442$, $p = 0.229$) and strategic decision incidents ($\text{Exp}(B) = 0.534$, $p = 0.106$).

3. Influence of social and environmental issues

Social and environmental issues significantly influence dependent variables at the organizational level, as shown in Table 25. Specifically, a one-unit increase in social and environmental issues increases the probability of e-commerce incidents by 3.49 ($\text{Exp}(B) = 3.488$, $p = 0.015$), business model incidents by 2.84 ($\text{Exp}(B) = 2.844$, $p = 0.011$), organizational innovation incidents by 4.85 ($\text{Exp}(B) = 4.846$, $p < 0.001$), and strategic decision incidents by 3.00 ($\text{Exp}(B) = 2.997$, $p = 0.002$). Conversely, a one-unit increase in social and environmental issues significantly decreases the probability of reputation incidents by 0.28 ($\text{Exp}(B) = 0.279$, $p < 0.001$). However, social and environmental issues do not significantly affect the probability of customer service incidents ($\text{Exp}(B) = 1.362$, $p = 0.530$).

Table 25: Results of binary logistic regression model at the “organizational level

Y X	E-commerce				Customer service				Reputation				Business models				Organization innovation				Strategic decision			
	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)
Normalized Data Governance	0.990	0.588	0.092	2.690	0.011	0.568	0.984	1.011	0.640	0.436	0.142	1.896	0.639	0.415	0.124	1.895	-1.901	0.457	0.000*	0.149	0.578	0.372	0.121	1.782
Normalized Content Generated	-0.817	0.679	0.229	0.442	-1.748	0.71	0.014*	0.174	1.630	0.522	0.002*	5.106	-1.541	0.417	0.000*	0.214	-1.103	0.521	0.034*	0.332	-0.628	0.388	0.106	0.534
Normalized social and environmental issues	1.249	0.511	0.015*	3.488	0.309	0.492	0.530	1.362	-1.278	0.361	0.000*	0.279	1.045	0.409	0.011*	2.844	1.578	0.358	0.000*	4.846	1.098	0.353	0.002*	2.997
Constant	-2.585	0.233	0.000*	0.075	-1.889	0.203	0.000*	0.151	0.940	0.153	0.000*	2.560	0.925	0.147	0.000*	2.523	-0.681	0.151	0.000*	0.506	0.189	0.133	0.156*	1.208
# observation	84				100				644				628				223				513			
LL (=2LL/-2)	-270.944				-304.376				-461.302				-487.686				-454.383				-571.536			
Cox R2	0.009				0.010				0.047				0.026				0.083				0.015			
Nagergelle R2	0.020				0.02				0.07				0.037				0.122				0.021			

Tableau 25 : Results of binary logistic regression model at the “organizational level

* $p < 0.05$, indicating statistical significance

5.3.3.3 Societal level

1. Influence of data governance issues

The impact of data governance issues on several dependent variables at the societal level as illustrated in Table 26. Specifically, a one-unit increase in data governance issues increases the probability of national security incidents by 3.40 ($\text{Exp}(B) = 3.399$, $p = 0.019$) and social cohesion incidents by 2.23 ($\text{Exp}(B) = 2.227$, $p = 0.042$). However, this increase shows a non-significant relationship with political stability ($\text{Exp}(B) = 0.663$, $p = 0.474$) and economic stability ($\text{Exp}(B) = 0.255$, $p = 0.010$).

2. Influence of content-generated Issues

As shown in Table 26, content-generated issues significantly influence multiple dependent variables at the societal level. A one-unit increase in content-generated issues has a significant relationship with an increase in the probability of national security incidents by 7.79 (Exp(B) = 7.792, $p < 0.001$), political stability incidents by 22.87 (Exp(B) = 22.865, $p < 0.001$), economic stability incidents by 4.45 (Exp(B) = 4.446, $p = 0.004$), and social cohesion incidents by 86.32 (Exp(B) = 86.315, $p < 0.001$).

3. Influence of social and environmental issues

Social and environmental issues significantly influence multiple dependent variables at the societal level, as highlighted in Table 26. A one-unit increase in social and environmental issues significantly decreases the probability of national security incidents by 90% (1-0.9) and political stability incidents by 91% (1-0.09). Conversely, a one-unit increase in social and environmental issues significantly increases the probability of economic stability incidents by 2.79 (Exp(B) = 2.791, $p = 0.018$) and social cohesion incidents by 4.12 (Exp(B) = 4.117, $p < 0.001$).

Table 26: Results of binary logistic regression model at the “societal level”

Y X	National security				Political stability				Economic stability				Social cohesion			
	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)	B	SE	Sign	Exp(B)
Normalized Data	1.223	0.519	0.019*	3.399	-0.411	0.574	0.474	0.663	-1.366	0.531	0.010*	0.255	0.801	0.394	0.042*	2.227
Governance																
Normalized Content	2.053	0.495	0.000*	7.792	3.13	0.536	0.000*	22.865	1.492	0.521	0.004*	4.446	4.458	0.477	0.000*	86.315
Generated																
Normalized social and environmental issues	-2.344	0.73	0.001*	0.096	-2.429	0.777	0.02*	0.088	1.026	0.432	0.018*	2.791	1.415	0.356	0.000*	4.117
Constant	-2.375	0.207	0.000*	0.093	-2.26	0.212	0.000*	0.104	-1.705	0.184	0.000*	0.182	-0.988	0.147	0.000*	0.372
# observation	120				101				131				415			
LL (-2LL/-2)	631.068				-281.437				-357.435				-522.083			
Cox R2	0.071				0.066				0.021				0.155			
Nagelkerke R2	0.128				0.129				0.037				0.207			

Tableau 26 : Results of binary logistic regression model at the “societal level”

* $p < 0.05$, indicating statistical significance

5.3.3.4 Environmental level

Because of the low level of observations of environmental level incidents in the OECD AIM database (8 observations), it is impossible to run binary regression analysis on this object at risk dimension.

5.3.3.5 Model fit and summary statistics

Table 27 summarizes all levels' log-likelihood values along with Cox & Snell R^2 and Nagelkerke R^2 . These metrics help in assessing the fit and explanatory power of the binary logistic regression models used in the study. At the individual level, we can interpret the following: The Log-Likelihood (LL) values are negative, with values ranging from -404.261 (Physical Safety) to -560.660 (autonomy and freedom). Lower log-likelihood values indicate a better fit of the model to the data. Physical safety has the lowest LL value (-404.261), suggesting the best fit among the categories, while autonomy and freedom have the highest LL value (-560.660), indicating the least fit among the models. Cox & Snell R^2 is a pseudo R^2 measure, which means the proportion of variance the model explains. Values range from 0.033 (autonomy and freedom) to 0.200 (disinformation). The highest Cox & Snell R^2 value is for disinformation (0.200), suggesting that the model explains 20% of the variance in disinformation incidents. On the other hand, the lowest Cox & Snell R^2 value is for autonomy and freedom (0.033), indicating that the model explains only 3.3% of the variance in these incidents. Nagelkerke R^2 is an adjusted version of Cox & Snell R^2 , scaled to cover the full range from 0 to 1. Values range from 0.044 (autonomy and freedom) to 0.287 (disinformation). The highest Nagelkerke R^2 value is for disinformation (0.287), indicating that the model explains approximately 28.7% of the variance in disinformation incidents. The lowest Nagelkerke R^2 value is for autonomy and freedom (0.044), suggesting that the model explains about 4.4% of the variance in these incidents.

At the organizational level, the best-fit model is for e-commerce as it has the lowest log-likelihood value (-270.944), indicating the best fit. However, strategic decision-making has the highest LL value (-571.536), indicating the least fit among the models. The model for organizational innovation has the highest Cox & Snell R^2 (0.083) and Nagelkerke R^2 (0.122), suggesting that it explains the highest proportion of variance among the organizational-level impacts.

On the societal level, the lowest LL value (-270.944) suggests the best fit, while strategic decision-making has the highest LL value (-571.536), indicating the least fit among the models. Cox & Snell R^2 is a pseudo R^2 measure, indicating the proportion of variance the model explains. The highest value is for organizational innovation (0.083), suggesting that the model explains 8.3% of the variance in organizational innovation incidents. The lowest value is for E-commerce (0.009), indicating that the model explains only 0.9% of the variance in these

incidents. Nagelkerke R² is an adjusted version of Cox & Snell R², scaled to cover the full range from 0 to 1. The highest value is for organizational innovation (0.122), indicating that the model explains approximately 12.2% of the variance in organizational innovation incidents

Models	Log-Likelihood (LL)	Cox & Snell R²	Nagelkerke R²
<i>Individual level</i>			
Human trust	-424.965	0.062	0.095
Disinformation	-415.562	0.200	0.287
Physical safety	-404.261	0.051	0.082
Privacy	-407.816	0.179	0.263
Psychological wellbeing	-495.674	0.154	0.211
Autonomy and freedom	-560.660	0.033	0.044
Quality of life	-507.608	0.109	0.150
Intellectual property	-430.769	0.044	0.067
<i>Organizational level</i>			
E-commerce	-270.944	0.009	0.020
Customer service	-304.376	0.010	0.020
Reputation	-461.302	0.047	0.070
Business models	-487.686	0.026	0.037
Organizational innovation	-454.383	0.083	0.122
Strategic decision-making	-571.536	0.015	0.021
<i>Societal level</i>			
National security	-631.068	0.071	0.128
Political stability	-281.437	0.066	0.129
Economic stability	-357.435	0.021	0.037
Social cohesion	-522.083	0.155	0.207

Tableau 27 : Model fit and summary statistics

5.4 Discussion

5.4.1 Theoretical contributions

Theoretically, this study complements Wach's (2023) work by exploring under-researched risk dimensions in GAI literature, such as the impact on individuals, including psychological well-being, autonomy, and human trust, as well as specific organizational impacts, such as customer service, strategic decision-making, innovation, and business reputation. This research also investigates unique GAI risks, such as hallucinations and biased content, thus broadening the understanding of GAI's multifaceted implications. Additionally, it extends the relational theory of risk (Boholm & Corvellec, 2011) by applying it specifically to GAI, identifying risk

objects (the nature or sources of risk) and objects at risk (entities vulnerable to these risks) pertinent to this technology. Consequently, this article aids organizations and policymakers in developing tailored risk mitigation strategies based on identified vulnerabilities.

The first managerial implication is that this study provides significant empirical contributions across three key risk objects: data governance, content-generated issues, and social and environmental concerns. By offering a theoretical framework to categorize GAI risks, it differentiates risk objects and objects at risk. Based on a literature review of GAI risks, this framework classifies the nature of risks into three categories with ten variables and their impact into four levels with twenty-one variables. The empirical study confirms the benefit of this categorization, demonstrating that the four levels of objects at risk (individual, organization, society, and environment) and the twenty-one variables are influenced by one or more types of risks (data governance, content-generated, or social and environmental issues).

The second managerial implication is that data governance issues are linked to various individual harm incidents, such as privacy violations, erosion of human trust, and psychological well-being issues. The significant odds ratios observed for privacy reflect broader socio-economic dynamics described as data capitalism. As West (2019) articulated, data capitalism involves commoditizing data and redistributing power in the information age. Our findings highlight the need for ethical data practices in GAI (trained on billions of data points) to prevent personal data exploitation and safeguard individual privacy rights within this evolving economic structure.

The third managerial implication is that this study expands the understanding of factors affecting human trust in GAI, mainly through the lens of data governance issues. Building on Chhillar and Aguilera's (2022) work, which identified vital features influencing human trust in algorithms, our research highlights data governance as a crucial additional dimension. This includes standards for algorithmic and human work, the subjective nature of trust influenced by both human and AI characteristics, and the role of transparency and scrutability in enhancing trust.

The fourth managerial implication is that the strong relationship between content-generated issues and their impact on individuals, organizations, and society underscores the broad risks linked to GAI hallucinations, biased or dangerous content. These findings suggest that content-generated issues could become systemic risks with the widespread adoption of GAI. Further

research is needed to explore this level of risk, potential regulation, or new insurance services to address GAI risks.

The fifth managerial implication is that this study provides detailed insights into how data governance, content-generated, and social and environmental issues affect different organizational activities. While GAI is generally seen as a disruptive technology that significantly enhances productivity, companies often lack awareness of GAI risks within their operations and the necessity of implementing measures to mitigate these risks.

The sixth managerial implication is that this study presents empirical evidence on how different dimensions of GAI influence various aspects of human life and organizations. The findings offer new insights into several under-explored areas: the strong relationship between data governance risk and privacy indicates a significant GAI risk transcending traditional data security concerns, redefining privacy in the GAI era; the link between content-generated risk and psychological well-being suggests that GAI content production can have severe mental health implications; and the strong relationship between content-generated risk and social cohesion indicates that GAI's impact on individuals' feelings of connectedness to society and mutual trust requires further exploration of GAI ethics.

This study also contributes to understanding social and environmental issues. One notable finding is the low level of environmental incidents in the AIM database, raising challenges such as a lack of awareness and measurement of GAI's negative environmental impact and insufficient media reporting on GAI's environmental consequences. Another significant finding is the strong relationship between social and environmental issues and social cohesion incidents, highlighting the need to explore GAI ethics further. Additionally, the study reveals a substantial effect on national security, where GAI-related social issues can compromise public safety and stability, underscoring the importance of addressing these risks through comprehensive policy and regulatory frameworks. Finally, the influence of social issues on political stability suggests that GAI-related social disruptions can exacerbate political tensions and conflicts, necessitating proactive measures to mitigate these risks.

By extending these areas of inquiry, our study contributes to the academic discourse and provides actionable insights for policymakers and technology developers to consider the broader implications of GAI technologies. This holistic approach ensures that the advancement

of GAI can be aligned with ethical standards and societal needs, fostering technology that benefits all sectors of society.

5.4.2 Managerial and policy implications

The practical implications of our findings are profound and diverse, offering critical guidance for policymakers, industry executives, and organizational leaders. The detailed analysis highlights opportunities and challenges, suggesting strategic adjustments to harness the full potential of GAI while mitigating its risks.

The first managerial implication is that organizations should strengthen their data protection policies in response to our data governance and privacy findings. This can be achieved by adopting more robust encryption techniques, refining user consent mechanisms, and promoting transparency in GAI operations. Strengthening these policies will ensure that data is adequately protected and privacy concerns are addressed effectively.

The second managerial implication is that, given the link between GAI content and psychological well-being, businesses must develop practices to ensure employees feel secure and supported in a GAI-integrated workplace. This includes conducting regular mental health assessments and offering training programs. Organizations can foster a more resilient and productive workforce by prioritizing mental well-being.

The third managerial implication is that the influence of GAI content on social cohesion underscores the necessity for companies to adopt and rigorously implement ethical GAI principles. This should primarily focus on fairness and inclusiveness. Companies must develop and enforce comprehensive ethical AI guidelines defining these concepts within GAI development. It is crucial to conduct regular bias audits on GAI models to identify and rectify any biases, evaluating both the training data and the outputs. Ensuring that teams designing and developing GAI systems are diverse can help identify and address potential biases and inclusiveness issues. Additionally, engaging with various stakeholders, including community groups, ethicists, and affected populations, in developing and reviewing GAI is essential.

The fourth managerial implication is that the influence of GAI on strategic decisions necessitates that organizations incorporate GAI risk assessments into their strategic planning to ensure long-term resilience and security. These assessments should evaluate compliance with relevant laws, such as the AI Act, and confirm that using GAI positively impacts the 17

Sustainable Development Goals (SDGs). Haidar (2024) elaborates on a framework for RAI that should be applied to responsible GAI. This framework outlines four main dimensions: the technical dimension involves applying ethical AI principles such as privacy, fairness, explainability, and security. The responsible innovation management dimension embeds elements like anticipation and reflexivity into corporate culture. The legal dimension examines AI-specific laws from the European Union, such as the AI Act and the United States, providing a comparative perspective on legal frameworks governing AI. Lastly, the framework emphasizes that GAI should positively impact the SDGs.

The fifth managerial implication is that to maintain consumer trust. Organizations must engage in transparent practices regarding their use of GAI and adherence to the law, providing clarity and accountability to avoid misrepresentations and build a reputable brand. It is also recommended that every company designing or implementing GAI establishes a section within its hierarchy dedicated to fact-checking and quality control, explicitly addressing misinformation.

Chapter 6: Conclusion

Chapter 1: What theoretical frameworks currently exist to guide the development and management of RDI/RAI? What would be an effective framework for RAI? What are the risks of GAI and their impact on individual society organizational and environment?



Chapter 2: What are the dimensions of Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?



Chapter 3: What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?



Chapter 4: What are the essential dimensions of Responsible AI (RAI) and how do they collectively contribute to fostering ethical and responsible development and deployment of AI technologies in business contexts?



Chapter 5: What are the specific generative AI risks? How do these GAI risks affect different aspects of society, business, and the environment?



Chapter 6: What comprehensive insights and conclusions can be drawn from this dissertation, and how do these findings inform managerial practices, policy development, address research limitations, and suggest future research directions?

6.1 Summary of the key findings for each chapter

Chapter 1: Introduction

Chapter 1 of the thesis establishes the groundwork for a more in-depth examination of AI in the dissertation, focusing on the complex relationship between DI and AI. The chapter starts with an overview of AI and its various types, like narrow AI and superintelligent AI. It traces the historical development of AI, illustrating its progression from early conceptual theories to modern, sophisticated applications. This historical perspective emphasizes significant milestones and contextualizes the current state of AI, demonstrating its transformative potential across various industries: healthcare, finance, retail, manufacturing, information environment, and education. By doing so, the chapter lays a solid foundation for comprehending AI's current state and future potential. The chapter then delves into digital innovation, emphasizing its role as a modern transformation catalyst. By examining specific examples, the chapter illustrates how AI facilitates innovative solutions and drives efficiencies across different sectors:

Nevertheless, the chapter also highlights the intrinsic risks linked to AI. The text outlines and classifies the risks into various significant domains, including bias and discrimination, privacy infringements, the possible weaponization of autonomous systems, cyberattacks, manipulation, social engineering, and socio-economic consequences like unemployment and economic inequality. These risks highlight the necessity of tackling the crucial obstacles using AI. Also, the chapter concludes by articulating the research questions that guide the dissertation, the epistemological positioning, and the research methodologies employed. It also outlines the structure of the dissertation, providing a roadmap for the subsequent chapters.

By categorizing and explaining the various risks associated with AI, the chapter highlights the critical challenges that require attention to ensure RAI deployment. This risk assessment sets the stage for the subsequent exploration of ethical principles and frameworks for RAI.

Chapter 2: A Framework for responsible digital innovation

In this chapter, we delved into digital innovation's definitions, dimensions, and barriers through the lens of *responsibility*. The chapter starts by defining DI and its dark side implementation on the three pillars of sustainability: environmental, social, and economic. Our investigation was guided by the following research question: "What are the dimensions of

Responsible Digital Innovation (RDI), and what challenges should companies consider to successfully implement RDI?”

To address this question, we conducted a qualitative study, interviewing 18 digital innovation managers from a diverse set of organizations, including both SMEs and large companies. This approach allowed us to gather in-depth insights into the practical challenges and strategies related to RDI from those at the forefront of digital innovation.

Our findings led to the development of a comprehensive framework RDI, structured around several key dimensions identified from the qualitative interviews:

- **RDI strategy:** This dimension highlights the importance of long-term strategic planning for digital innovation. It emphasizes that companies need to align their DI efforts with broader sustainability goals, ensuring that their innovation strategies contribute positively to environmental, social, and economic outcomes.
- **Responsible by design digital innovation:** This dimension addresses the challenges that digital innovation-orientated companies should address: limiting the negative impact on the environment, data privacy, building ethical AI, data and information system security, digital sovereignty, and the inclusion of different stakeholders.
- **Organizational challenges and KPI on RDI:** This dimension addresses companies’ internal hurdles when implementing RDI. It emphasizes the value of external collaborations, highlighting the significance of innovative teams and digital partnerships. Successful RDI efforts depend on extensive employee training and the careful selection of partners and suppliers, regardless of their RDI strategy. Additionally, this dimension underscores the need for specific human resources and organizational structures to support RDI activities. Developing specific KPIs is crucial for measuring the effectiveness and impact of RDI efforts, providing tangible metrics for tracking progress, and making necessary adjustments.
- **Digital responsibility towards end Users:** Ensuring user-centric ethical considerations is crucial. This dimension advocates for the protection of end users’ rights and interests, promoting transparency, privacy, and fairness in the use of digital technologies.
- **Catalysts:** Identifying driving forces for implementing RDI is essential for fostering an environment conducive to responsible innovation. These catalysts can include Top management support, benchmarking, regulation, trial-and-error approach, and funding.

Each key dimension is further divided into twenty-five sub-dimensions, providing a detailed and granular view of RDI's practical aspects. This detailed framework offers companies a robust guide for implementing RDI, addressing both strategic and operational levels. Furthermore, this chapter highlighted several barriers to implementing RDI, such as difficulties in scaling responsible practices across large organizations and the resistance to change within established corporate cultures. To overcome these challenges, we suggested policy implications, including the need for supportive regulatory frameworks, incentives for sustainable innovation, and the promotion of cross-sector collaboration.

Chapter 3: Ethics in AI: Analyzing ethical AI declarations and future directions

The ethics of AI, a crucial and rapidly developing field, spans a wide range of topics with significant ethical stakes. Key issues include algorithmic biases, fairness, automated decision-making, accountability, privacy, and regulation. Algorithmic bias can infiltrate systems in numerous ways. The most prevalent view is that bias is embedded within the historical data used to train AI systems. For example, Amazon discontinued its use of an AI hiring and recruitment tool because the algorithm favored male candidates over female ones. This bias arose because the system was trained on data collected over ten years, predominantly from male candidates. As a result, the algorithms learned biased patterns from the historical data, predicting that these types of candidates were most likely to succeed. Consequently, the AI system's recruitment decisions were biased against female and minority candidates. Friedman and Nissenbaum (1996) identify three categories of bias in computer systems: existing bias, technical bias, and emergent bias. In natural language processing, issues can stem from the text corpus—the source material the algorithm uses to learn about relationships between different words. This can result in biased language models if the training data contains biased or unrepresentative text.

According to a 2019 report from the Center for the Governance of AI at the University of Oxford, 82% of Americans believe that robots and AI should be carefully managed. The concerns highlighted include the use of AI in surveillance, the spread of fake content online (including deep fakes, which are doctored videos and audio generated with the help of AI), cyberattacks, infringements on data privacy, hiring bias, autonomous vehicles, and drones that operate without human controllers. Similarly, a 2021 study by KPMG and the University of Queensland Australia across five countries found that 66-79% of citizens in each country view

the impact of AI on society as uncertain and unpredictable, with 96% of respondents expecting AI governance challenges to be handled carefully.

Based on these risks raised by AI, chapter 3 addresses the question: “*What ethical principles underpin current ethical AI declarations, what opportunities and limitations do they present, and to what extent do these declarations fulfill the requirements for responsible AI?*” The methodology employs frequency analysis of 27 AI declarations worldwide to examine the distribution and occurrence of ethical principles within AI ethics declarations and an integrative literature review. We summarize the frequency findings with the following:

- **Dominant principles:** These are the principles most commonly cited (over 50%) across various declarations, signifying a solid consensus or significant concern within the field. The key principles are transparency (81%), fairness (67%), accountability (63%), and privacy (56%).
- **Intermediate principles:** Principles in this category are mentioned with moderate frequency (40-50%). They include security (48%), safety (44%), explainability (44%), and beneficence/well-being (41%). The less frequent mention of these principles could be due to several factors: the complexities involved in putting these principles into practice, the fast-paced evolution of AI, and the ongoing ethical discourse, which indicates that consensus is still developing in these areas.
- **Less emphasized areas:** These principles are cited less often (below 40%) in the declarations, including legislation or legal compliance (15%), data governance (11%), and disinformation (4%). This lower frequency suggests potential gaps in the current ethical frameworks that need to be addressed.

The second outcome of this chapter is the identification of the strengths and weaknesses of these declarations. The strengths identified include global participation in crafting these declarations, their comprehensive coverage, addressing ethical issues in AI, promoting AI for social good, and preserving trust. However, there are also notable weaknesses and limitations, such as the lack of a standardized framework for ethical AI, divergence in definitions and terminology in AI ethics, challenges in translating principles into practice, integration with existing legal frameworks, erosion of trust due to limited focus on impact, usability and practical implementation of ethical tools, conflicting objectives in ethical AI declarations, and geopolitical tensions. This critical assessment provides a foundation for enhancing future AI ethical frameworks and contributes to the ongoing discourse on developing robust, practical, and universally accepted ethical standards for AI. This chapter provided valuable insights for

businesses, policymakers, and scholars, emphasizing the practical challenges of implementing ethical AI principles.

Chapter 4: A framework for responsible AI

This chapter detailed an integrative framework for RAI, focusing on four key dimensions:

The *technical dimension* of machine learning emphasizes the critical importance of data collection in AI research, focusing primarily on analyzing tabular data through supervised learning. This method involves predicting outputs from inputs based on labeled examples, and its success heavily depends on data quality and manipulation. Despite the potential of newer data types, the field faces several technical issues, such as data quality problems, sampling bias, data drift, overfitting, underfitting, and data leakage. While advancements in semi-supervised learning exist, obtaining high-quality labeled data remains resource-intensive and often requires human effort. Sharing labeled datasets can reduce redundancy, but competition in research hampers collaboration. To address these challenges and ensure AI's ethical and reliable use, this chapter suggests implementing ethical principles such as accountability, privacy, transparency, safety, fairness, and autonomy. These principles can help mitigate technical issues and promote trust and integrity in AI applications.

Responsible innovation (RI) management emphasizes the need for firms to align societal values and behaviors with socio-ethical considerations in innovation. The novelty of RI lies in linking innovation explicitly with responsibility, aiming to include all stakeholders early in the research and development process to anticipate benefits and prevent negative consequences. However, current RI approaches often focus narrowly on science and technological development, overlooking other innovation types and commercialization stages, posing challenges in business contexts. Documented implementations of RI in business are scarce, but insights from social and sustainable innovations, which also address societal challenges and involve multiple stakeholders, can be informative. Stilgoe et al.'s (2013) influential RI framework includes anticipation, reflexivity, inclusion and deliberation, responsiveness, and knowledge management, providing practical business guidelines. For instance, anticipation involves assessing AI's impact on roles and responsibilities, while reflexivity requires continuous evaluation of AI systems. Inclusion demands transparent communication policies, responsiveness involves adaptive management using dashboards, and knowledge management focuses on integrating new information into the innovation process. This chapter advocates for implementing ethical principles—accountability, privacy, transparency, safety, fairness, and

autonomy—to address technical challenges such as data quality issues, sampling bias, data drift, overfitting, underfitting, and data leakage, thereby promoting ethical and reliable AI applications.

The legal dimension examined the evolving landscape of AI regulation, comparing approaches in the European Union (EU) and the United States (US). The EU's proactive stance on AI regulation, with robust frameworks like the GDPR and the updated Data Governance Act, prioritizes data protection and RAI development. These regulations emphasize transparency, accountability, and anti-discrimination measures, setting a stringent standard for AI practices. In contrast, the US's decentralized approach, with diverse state laws, offers flexibility but can lead to uncertainties and complexities for businesses operating across multiple states. This thesis details specific laws and regulations in each region, analyzes their effects on AI practices, and highlights the challenges and advantages of each approach.

The final dimension explored the impact of AI on the three pillars of sustainability: economic, social, and environmental. AI has the potential to drive economic growth by increasing efficiency, productivity, and innovation across various industries. Socially, AI can enhance quality of life by improving healthcare, education, and accessibility, though it also raises concerns about job displacement and privacy. Environmentally, AI can contribute to sustainability by optimizing resource use, reducing emissions, and supporting climate change mitigation efforts. However, balancing these benefits with ethical considerations is crucial to avoid exacerbating inequalities or causing unintended harm. The chapter highlighted the need for sustainable AI practices that align with moral principles and promote long-term societal well-being.

The framework underscores the importance of a holistic approach to AI ethics, combining technical, legal, and managerial perspectives to foster RAI development and deployment.

Chapter 5: Generative AI: A quantitative study on emerging risks and impacts

Chapter 5 conducts a quantitative study on GAI's risks and analyzes their impacts across societal, organizational, environmental, and individual dimensions.

This study used a comprehensive dataset from the OECD AI Incidents Monitor, encompassing 858 recorded incidents, employing Python for data analysis, SPSS for statistical evaluation, and binary logistic regression analysis as a method. The objective is to elucidate the relationships between different aspects of GAI risks and their potential impacts.

The analysis uncovers several critical insights. One of the key findings is that normalized data issues significantly affect all examined levels, with the most profound impact observed at the individual level. Issues such as privacy violations and the dissemination of disinformation emerge as particularly detrimental, highlighting the need for robust measures to protect individuals from these risks.

Moreover, the study identifies that normalized content generation issues, while broadly influential across various levels, have the most significant effects on individuals and society. These issues impact human trust, psychological well-being, and social cohesion, underscoring the broader societal implications of GAI technologies.

Additionally, the research highlights that normalized social and environmental concerns exert a notable impact on environmental aspects, with a particular emphasis on global warming. This finding points to the urgent need for GAI technologies to incorporate environmental considerations into their design and deployment processes.

The chapter concludes by advocating for future research aimed at developing a dynamic framework for responsible GAI risk management. Such a framework would address the immediate impacts identified in this study and adapt to the evolving risks as GAI technologies continue to advance. The proposed framework aims to ensure that GAI development aligns with broader societal, environmental, and ethical standards, thereby mitigating potential negative outcomes and enhancing the positive influence of GAI.

6.2 Contribution of the thesis

This thesis offers a comprehensive and multifaceted contribution to the field of RAI, advancing our understanding and application of ethical principles in AI deployment. The research conducted across the four chapters integrates theoretical, qualitative, and quantitative methodologies to provide a robust framework and critical insights that address key challenges in achieving RAI. The main contributions of the thesis are summarized as follows:

The first major contribution of this thesis is developing a comprehensive framework for RDI. This framework serves as a foundational model for understanding and implementing responsible innovation within digital technologies. Derived from qualitative insights, the framework delineates essential dimensions, including strategic planning, ethical design, organizational challenges, user-centric considerations, and catalysts for responsible innovation. In contrast to previous literature, which often focuses on isolated aspects of DI like strategic planning (Nylén & Holmström, 2015) and ethical design alone (van den Hoven, 2017), this

thesis offers an integrated perspective that combines these disparate elements into a cohesive whole. This integrated approach addresses internal organizational hurdles and provides practical guidelines for companies to embed responsibility into their DI processes. Additionally, this thesis extends the discourse on “responsibility” by designing an RAI framework. This framework integrates technical dimensions, sustainable development principles, responsible innovation management practices, and legislative requirements. Doing so bridges the gap between theoretical discussions and practical implementation, offering a holistic approach to RAI grounded in current academic discussions and the practical needs of the industry. This comprehensive approach to RDI and RAI advances theoretical understanding and offers actionable insights for practitioners aiming to navigate the complex landscape of DI responsibly.

The second contribution of this thesis is the critical evaluation of ethical AI principles. This research conducts an in-depth analysis of various ethical AI declarations, identifying and evaluating the principles of transparency, fairness, accountability, and privacy. Previous literature has extensively discussed these principles (Fukuda-Parr & Gibbons, 2021; Hagendorff, 2020; Du & Xie, 2021). However, this thesis offers a nuanced understanding by critically examining the opportunities and limitations inherent in these principles. While transparency, fairness, accountability, and privacy are frequently highlighted in the existing discourse, this thesis reveals gaps in their practical implementation. By scrutinizing these principles, the research underscores the necessity for additional dimensions and practical mechanisms to meet RAI's requirements. For instance, the analysis shows that transparency alone is insufficient without accompanying mechanisms to ensure comprehensibility and accessibility of information to diverse stakeholders. Moreover, this thesis contributes to the ongoing discourse on AI ethics by proposing enhancements to these established principles. It advocates for integrating supplementary dimensions such as inclusivity, adaptability, and resilience, which are critical for addressing AI systems' dynamic and complex nature. These additions enrich the theoretical framework of ethical AI and provide actionable insights for policymakers and practitioners striving to implement ethical AI in real-world scenarios. By offering a comprehensive critique and proposing practical enhancements, this contribution advances the theoretical understanding of AI ethics and supports the development of more robust and effective ethical AI guidelines.

This thesis's third significant contribution is identifying and managing risks associated with GAI. This research quantitatively examines emerging harms linked to GAI, addressing substantial gaps in the current understanding of these harms. Previous literature has primarily focused on narrow AI, often overlooking GAI's unique challenges. This thesis fills this critical gap by offering a structured framework for managing GAI risks. By classifying risks by their nature (risk object) and their impact (object at risk), the research provides a systematic approach to understanding and mitigating GAI risks. This classification helps identify specific vulnerabilities and potential consequences, which is essential for developing targeted risk management strategies. For instance, risks to privacy and security, ethical misuse, and unintended consequences are categorized and analyzed for their potential impact on individuals, organizations, and society. Furthermore, this thesis presents practical guidelines for developers, industry leaders, and policymakers to navigate the complexities of GAI. The structured framework facilitates a deeper understanding of GAI risks and provides actionable strategies for their management. This includes implementing robust monitoring systems, establishing ethical guidelines, and fostering collaboration between stakeholders to ensure the ethical deployment of GAI.

The fourth significant contribution of this thesis is the practical implications it offers for both businesses and policymakers. The comprehensive frameworks and critical insights developed throughout this research provide actionable guidelines and performance indicators that enterprises can utilize to integrate responsible innovation into their strategic and operational processes. Previous literature has often highlighted the need for responsible innovation, but this thesis goes further by detailing specific steps and metrics that companies can adopt to achieve this goal.

For businesses, the practical implications include strategies for embedding ethical considerations into every stage of the innovation process, from ideation to implementation. The thesis outlines how companies can overcome organizational challenges, align innovation efforts with ethical and sustainable goals, and measure progress through clearly defined performance indicators. These guidelines are essential for businesses aiming to navigate the ethical complexities of DI and maintain a competitive edge in a rapidly evolving market.

For policymakers, the findings of this thesis offer a robust foundation for developing regulatory frameworks and policies that promote ethical AI practices. The research underscores the necessity of supportive regulations that encourage responsible innovation while also protecting

societal interests. This thesis contributes to the ongoing efforts to shape effective AI governance by providing a detailed analysis of the current policy gaps and proposing targeted solutions. Moreover, the thesis emphasizes the importance of international collaboration in addressing the global challenges posed by AI. It advocates for a coordinated approach among nations to establish common standards and practices, thereby enhancing the effectiveness of AI governance and ensuring that ethical principles are upheld across borders.

The fifth essential contribution of this thesis is its advancement of the academic discourse on RAI. This research bridges theoretical concepts and practical applications in RAI, enriching the scholarly understanding of responsible innovation. Previous studies have often treated theoretical and practical aspects in isolation. In contrast, this thesis employs an integrative approach that combines qualitative interviews, literature reviews, and quantitative analyses to provide a comprehensive perspective. Through a detailed exploration of ethical principles, frameworks, and risk assessments, the thesis offers a robust resource for future research and development in RAI. The qualitative interviews provide nuanced insights into the real-world challenges and opportunities associated with responsible innovation, while the literature review situates these findings within the broader academic context. The quantitative analyses further validate these insights, offering empirical evidence to support the proposed frameworks and guidelines. This comprehensive approach enhances the theoretical foundations of RAI and offers practical tools and methodologies that researchers can utilize in their work. By integrating diverse methodologies and perspectives, the thesis demonstrates the value of a multidisciplinary approach to addressing the complex issues surrounding AI ethics and responsibility. Moreover, the thesis contributes to the academic discourse by identifying areas where current understanding is lacking and proposing new directions for future research. This includes the need for more empirical studies on the implementation of ethical AI principles and the development of new frameworks that can adapt to the evolving technological landscape.

Overall, this thesis makes significant contributions by taking a comprehensive approach to RAI. It includes the creation of integrated frameworks, a thorough assessment of ethical principles, identifying and mitigating new hazards, and practical recommendations for businesses and politicians. This research contributes to the area by providing a strong basis for ethical and sustainable AI practices, guaranteeing that technology progress positively impacts environmental, social, and economic sustainability.

6.3 Managerial implications

The first managerial implication is that integrating RAI into business operations provides companies with opportunities to create new competitive advantages. By adopting RAI, companies can enhance their reputation, foster consumer trust, and differentiate themselves in the market. For example, businesses prioritizing ethical AI practices can attract customers who are increasingly concerned about privacy and ethical considerations. This commitment can also increase customer loyalty and a stronger brand image.

The second managerial implication is that RAI can create new business models. For instance, companies can develop AI-driven services emphasizing transparency and fairness, appealing to a growing segment of consumers and businesses that value ethical considerations. An example is Microsoft's AI for Good initiative, which focuses on using AI to address societal challenges, thus aligning business objectives with ethical goals.

The third managerial implication is that to implement RAI effectively, companies must undergo significant organizational changes. This involves cultivating a corporate culture that places a high emphasis on accountability and moral principles in the process of technology creation. The rise of new positions, such as Chief Data Officers and Ethics Officers, highlights the necessity for specialized roles that supervise practices related to RAI. Moreover, the role of traditional managers evolves; they must now possess a deep understanding of responsible technology, the workings of AI, and the associated cost-benefit dynamics. Managers should also be familiar with the underlying algorithms and develop strategies to integrate RAI effectively.

The fourth managerial implication is that the shift towards RAI also changes organizations' responsibility dynamics. Developers play a crucial role in assuring the ethical development of AI and often take on managerial responsibilities in this regard. They must possess technical skills, comprehend moral ramifications, and be capable of making responsible decisions autonomously.

The fifth managerial implication is that managers must also cultivate a flexible and creative workforce. As routine tasks become automated through AI, the value of human employees increasingly lies in their creativity and problem-solving abilities. Managers should focus on motivating employees, inspiring innovation, and building a culture of responsibility.

Techniques such as fostering a collaborative environment, recognizing and rewarding innovative ideas, and providing continuous learning opportunities can help achieve this.

The sixth managerial implication is that leadership styles vary between high-growth and low-growth companies. In high-growth companies, AI is often used more extensively, freeing senior executives to focus more on leadership tasks rather than operational ones. This leadership time allows executives to spend more time motivating their teams and driving organizational culture. For example, companies like Google and Amazon utilize AI to streamline operations, enabling their leaders to focus on strategic initiatives and employee engagement.

The seventh managerial implication is that CEOs and senior executives often find themselves in an “accelerating trap,” where the pace of technological advancements pressures them to innovate and adapt continually. This relentless pace can lead to strategic decisions driven more by the need to keep up with competition than by considering long-term implications. Leaders must balance the need for rapid innovation with the principles of RAI, ensuring that their decisions are competitive and ethically sound.

The eighth managerial implication is that the rise of GAI further amplifies these changes in managerial practices. GAI can produce content, designs, and decision-making models, necessitating heightened privacy protection practices. Managers must ensure that generative AI systems comply with stringent data privacy regulations and ethical guidelines to protect user data and maintain trust.

The ninth managerial implication is that GAI also introduces new dimensions of creativity and efficiency in business processes. It can automate complex tasks such as content creation, product design, and strategic planning, allowing managers to focus on higher-order decision-making and innovation. However, this also means managers must understand how generative AI works and its implications. They must be equipped to oversee AI-driven projects, ensure ethical use, and manage the balance between human and AI contributions.

Additionally, managers must be aware of regulatory frameworks like the European Union's AI Act, which establishes a common regulatory and legal framework for AI within the EU. The AI Act, proposed by the European Commission on April 21, 2021, passed the European Parliament on March 13, 2024, and was unanimously approved by the EU Council on May 21, 2024. Understanding and complying with such regulations is crucial for ensuring

that AI implementations are ethical and legally compliant. This awareness will help managers navigate the complex regulatory landscape and align their strategies with legal and ethical standards.

6.4 Policy-making implications

The first policy implication is the necessity for comprehensive regulatory frameworks that support the integration of RAI. Policymakers must develop regulations that address the multi-dimensional aspects of RAI, including technical, ethical, legal, and sustainable development considerations.

The European Union's AI Act exemplifies the second policy implication. This regulation establishes a common regulatory and legal framework for AI within the EU, promoting national cooperation and ensuring compliance by creating the European Artificial Intelligence Board. The Act applies extraterritorially, similar to the EU's General Data Protection Regulation (GDPR), meaning it affects AI providers outside the EU if they have users within the EU.

The third policy implication is that the AI Act covers all types of AI across various sectors, except for AI systems used solely for military, national security, research, and non-professional purposes. As product regulation, it governs AI providers and entities using AI in a professional context without conferring direct rights on individuals. Notably, the Act addresses the rise of generative AI systems, such as ChatGPT, by proposing more restrictive regulations for powerful generative AI with systemic impacts.

The fourth policy implication concerns the institutional governance framework established by the AI Act. According to the European Parliament Legislative Resolution of March 13, 2024, the AI Act includes creating several new institutions to implement and enforce its regulations. These institutions include:

- **AI office:** Attached to the European Commission, this authority will coordinate the implementation of the AI Act in all Member States and oversee the compliance of general-purpose AI providers.
- **European AI board:** Composed of one representative from each Member State, this Board will advise and assist the Commission and Member States to facilitate the consistent and practical application of the AI Act. Its tasks include gathering and sharing

technical and regulatory expertise, providing recommendations, written opinions, and other advice.

- **Advisory forum:** This forum will represent a balanced selection of stakeholders, including industry, start-ups, SMEs, civil society, and academia, ensuring that a broad spectrum of opinions is represented during the implementation and application process.
- **Scientific panel of independent experts:** This panel will provide technical advice and input to the AI Office and national authorities, enforce rules for general-purpose AI models, and ensure that the regulations and implementations of the AI Act correspond to the latest scientific findings.

Member States will also designate "national competent authorities" responsible for ensuring the AI Act's application and implementation and conducting "market surveillance." These authorities will verify that AI systems comply with the regulations, notably by checking the proper performance of conformity assessments and appointing third parties to conduct external conformity assessments.

The fifth policy implication is that policymakers should focus on harmonizing regulations across geopolitical contexts to facilitate international collaboration in RAI development. This could involve creating international standards and guidelines that ensure global consistency and interoperability in AI systems. Additionally, there should be mechanisms for continuous monitoring and assessment of AI systems to ensure ongoing compliance with ethical standards.

The sixth policy implication is the need for policies that promote education and training in AI ethics and responsible innovation. Policymakers can support initiatives by providing resources and training for businesses and professionals to effectively understand and implement RAI principles. This will help build a workforce that is not only technically proficient but also ethically aware and capable of making responsible decisions in AI development and deployment.

In conclusion, integrating RAI into business practices requires a multifaceted approach involving comprehensive regulatory frameworks, continuous monitoring, and education initiatives. The EU AI Act is a pioneering example of how policymakers can create robust regulations promoting ethical AI development while fostering innovation. By addressing these

policy-making implications, governments can ensure the ethical and sustainable advancement of AI technologies.

6.5 Limitation of the thesis

1. Methodological limitations

A notable methodological constraint is the substantial reliance on qualitative research techniques, namely in Chapter 2, where interviews took place with 18 digital innovation managers from both small and large firms. Although qualitative insights are valuable as they encompass several sectors, the sample size limits achieving comprehensive knowledge and framework. It may fail to encompass various experiences and viewpoints from multiple areas. Also, the proposed framework for RAI in Chapter 4 was suggested based on a thorough literature review. However, a significant limitation is that this framework has not been empirically tested in real-world applications. Applying this framework in various business contexts would provide valuable insights into its practicality, effectiveness, and potential challenges or limitations during implementation. Lastly, Chapter 5 employed a quantitative study on GAI risks relying on available data and the current nature of these technologies. However, the rapidly changing landscape of GAI means that new risks and impacts may not have been fully captured in this analysis.

2. Contextual limitations

The contextual focus of the research also presents limitations. The frameworks for RDI and RAI were developed based on specific contexts and industries, primarily within the European Union and Western business settings. The suitability of these frameworks in various cultural, economic, and legal contexts remains unknown. For example, firms that operate in locations with less strict regulations or different cultural perspectives on technology and ethics may have distinct obstacles and opportunities while adopting RDI and RAI.

While this thesis evaluates current ethical AI principles and their roles, the practical implementation of these principles remains challenging. The gap between theoretical ethical guidelines and their real-world application is an area that requires further exploration. This thesis primarily identifies the opportunities and limitations but does not provide exhaustive solutions for bridging this gap.

3. Scope limitations

Another limitation of the thesis is its scope. Although the research offers extensive frameworks for RDI and RAI, it does not extensively explore the specific applications of these frameworks in different sectors or the subtle variations between industries. Various industries, including healthcare, banking, and manufacturing, may have distinct requirements and difficulties related to responsible innovation that are not adequately covered by the broad frameworks suggested in this thesis.

4. Technological evolution

The rapid pace of technological advancement in AI and DI presents another limitation. This thesis bases its frameworks and insights on current technologies and trends. However, AI and DI continually evolve, and new ethical, legal, practical, and managerial challenges will emerge. The dynamic nature of the frameworks requires regular updates to remain relevant and valuable.

5. Implementation challenges

Finally, the thesis acknowledges the theoretical development of the RAI framework and provides limited empirical evidence on the practical implementation of RDI. The transition from theory to practice involves significant challenges, including organizational inertia, resource constraints, lack of funding, and potential resistance to change.

In summary, while this thesis substantially contributes to the understanding and practice of RAI, it is vital to recognize its limitations. Addressing these limitations through future research will be crucial in refining and advancing the frameworks and principles discussed, ensuring that AI technologies are developed and deployed responsibly and sustainably.

6.6 Recommendation for future research

The field of RAI is dynamic and evolving, presenting numerous opportunities for innovative and impactful research. Building on the comprehensive findings of this thesis, the following recommendations aim to push the boundaries of current understanding and practice in RAI, fostering new insights and advancements.

The first recommendation for future research is the exploration of AI ethics through emerging technologies. Further investigation is needed to explore the impact of emerging technologies, such as quantum computing and neuromorphic computing, on the ethical aspects

of AI. These technologies have the potential to transform computational capabilities completely, and it is crucial to comprehend their ethical implications when designing RAI systems. Examining situations where these sophisticated technologies intersect with AI in business operations may uncover novel ethical dilemmas and prospects, thereby providing deeper insights into how businesses can integrate these advanced technologies responsibly.

The second recommendation for future research is the development of AI ethics simulation environments. Developing virtual environments to assess and analyze ethical AI concepts in real time would represent a revolutionary advancement. These simulation platforms can simulate intricate situations that include the implementation of artificial intelligence in diverse industries. This enables researchers and professionals to see the results of various ethical strategies and make necessary adjustments. These settings would offer a secure setting for testing out creative and ethical rules before their actual application in the real world.

The third recommendation for future research is to investigate GAI-driven solutions for global challenges. Investigating sustainable General AI (GAI) is vital for future research. This includes exploring two key ideas: designing GAI sustainably and ensuring GAI positively impacts sustainability. Future studies should focus on leveraging GAI to address some of the world's most pressing issues, such as climate change, poverty, and healthcare access. Researchers can explore how to responsibly design and implement GAI-driven solutions to maximize positive social and environmental impacts.

The fourth recommendation for future research is the integration of cultural and societal values in AI development. A deeper exploration into how cultural and societal values can be integrated into AI systems is crucial. Future research should develop methodologies for incorporating diverse cultural perspectives and ethical norms into AI algorithms and practices. This approach would help create globally applicable AI that respects local traditions and values.

The fifth recommendation for future research is the exploration of ethical AI in the context of human-AI collaboration. As AI systems become more integrated into everyday life, understanding the dynamics of human-AI partnership will be essential. Research should explore the ethical implications of these interactions, focusing on areas such as trust, dependency, and decision-making. Developing frameworks that ensure ethical collaboration between humans and AI will be vital for the responsible advancement of these technologies.

The sixth recommendation for future research is to explore the concept of corporate digital responsibility (CDR). Future research should explore how companies can integrate CDR into their comprehensive business strategies, focusing on areas such as data privacy, cybersecurity, digital inclusion, and the ethical use of AI. Research could investigate optimal practices for implementing CDR, the impact of CDR initiatives on corporate reputation and stakeholder trust, and the challenges and advantages of aligning CDR with traditional CSR endeavors. Additionally, exploring case studies of organizations successfully implementing CDR can offer significant insights and guidelines for other businesses.

The seventh recommendation for future research is participatory design and citizen involvement in AI development. Encouraging participatory design processes where citizens actively develop AI systems can lead to more ethical and socially acceptable technologies. Future research should investigate strategies for involving diverse populations in the AI design process, guaranteeing that their perspectives and apprehensions are acknowledged and resolved. Implementing this collaborative method can improve the credibility and adoption of AI technologies. Several tools and methods can be utilized to integrate people from society into the AI development process. First, arranging workshops and focus groups can allow community members to express their viewpoints, experiences, and concerns about AI technologies. Second, diverse citizen panels can provide ongoing feedback and monitoring during AI development. Third, online and in-person public consultations engage the community.

The eighth recommendation for future research is to explore the ethical implications of AI in creativity and art. GAI's role in creative fields such as art, music, and literature raises unique ethical questions. Future studies should delve into the implications of AI-generated content, exploring issues related to authorship, originality, and cultural impact. Understanding how to integrate GAI ethically into creative processes will be crucial for preserving artistic integrity and cultural heritage.

In summary, future study recommendations emphasize innovative and transformative approaches to advancing RAI. Future research can ensure that AI continues evolving ethically, transparently, and efficiently by exploring new technologies, engaging diverse perspectives, and creating dynamic and participatory frameworks. These forward-thinking recommendations inspire researchers and practitioners to push the boundaries of what is possible in RAI, fostering a future where technology and ethics are in harmonious alignment.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (pp. 265–283).
- Abedin, B. (2022). Managing the tension between opposing effects of explainability of artificial intelligence: a contingency theory perspective. *Internet Research*, 32(2), 425-453. <https://doi.org/10.1108/INTR-05-2020-0300>
- Abelkop, A. D. K., & Carlson, J. C. (2012). Reining in Phaëthon's chariot: Principles for the governance of geoengineering. *Transnational Law & Contemporary Problems*, 21, 101. University of Iowa Legal Studies Research Paper No. 12-27. <https://ssrn.com/abstract=2165372>
- Acemoglu, D., Manera, A., & Restrepo, P. (2020). *Does the US tax code favor automation?* (No. w27052). National Bureau of Economic Research.
- Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: an empirical analysis. *The American Economic Review*, 678-690. <https://www.jstor.org/stable/1811167>
- Acumen Research and Consulting. (2022). GAI Market Size Will Achieve USD 110.8 Billion by 2030 growing at 34.3% CAGR - Exclusive Report by Acumen Research and Consulting. <https://www.globenewswire.com/news-release/2022/12/14/2574140/0/en/Generative-AI-Market-Size-Will-Achieve-USD-110-8-Billion-by-2030-growing-at-34-3-CAGR-Exclusive-Report-by-Acumen-Research-and-Consulting.html>
- Acumen Research and Consulting. (2023). AI in Healthcare Market is forecasted to Reach USD 427.5 Billion by 2032, growing at a CAGR of 44.0% Over the Forecast Period 2023 to 2032. <https://www.globenewswire.com/en/news-release/2023/10/10/2757664/0/en/AI-in-Healthcare-Market-is-forecasted-to-Rreach-USD-427-5-Billion-by-2032-growing-at-a-CAGR-of-44-0-Over-the-Forecast-Period-2023-to-2032.html>
- Adeusi, S. O. (2020). Roles and Importance of Ethics. In *Research Anthology on Rehabilitation Practices and Therapy* (Vol. 1, pp. 01-11). IGI Global. <https://leen.loc.gov/2019057243>
- Adorf, H. M. (2005). Connectionism and neural networks. *Knowledge-Based Systems in Astronomy*, 213-245. https://doi.org/10.1007/3-540-51044-3_25
- Aggarwal, S., Gupta, S., & Ahuja, R. (2021). A review on protein subcellular localization prediction using microscopic images. In *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)* (pp. 72-77). IEEE.
- Aghion, P., Antonin, C., & Bunel, S. (2020). Artificial Intelligence, Growth and Employment: The Role of Policy. *Economie et Statistique / Economics and Statistics*, 510-511–512, 149–164. <https://doi.org/10.24187/ecostat.2019.510t.1994>

Ahmad, A., Liew, A. X., Venturini, F., Kalogeras, A., Candiani, A., Di Benedetto, G., ... & Martos, V. (2024). AI can empower agriculture for global food security: challenges and prospects in developing nations. *Frontiers in Artificial Intelligence*, 7, 1328530. <https://doi.org/10.3389/frai.2024.1328530>

Ahmad, S., Mallick, D. N., & Schroeder, R. G. (2013). New product development: Impact of project characteristics and development practices on performance: NPD project characteristics and practices. *Journal of Product Innovation Management*, 30(2), 331-348. <https://doi.org/10.1111/j.1540-5885.2012.01002.x>

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status quo, challenges and opportunities. *Journal of Cleaner Production*, 289, 125834. <https://doi.org/10.1016/j.jclepro.2021.125834>

Ait-Sahalia, Y., & Sağlam, M. (2024). High frequency market making: The role of speed. *Journal of Econometrics*, 239(2), 105421. <https://doi.org/10.1016/j.jeconom.2022.12.015>

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. <https://doi.org/10.1016/j.ijinfomgt.2021.102387>

Ali, S. M. (2019). White Crisis” And/As “Existential Risk,” Or The Entangled Apocalypticism Of Artificial Intelligence: With Robert M. Geraci And Simon Robinson, “Introduction To The Symposium On Artificial Intelligence And Apocalypticism”; Beth Singler, “Existential Hope And Existential Despair In Ai Apocalypticism And Transhumanism”; Michael Morelli, “The Athenian Altar And The Amazonian Chatbot: A Pauline Reading Of Artificial Intelligence And Apocalyptic Ends”; Victoria Lorrimar, “Mind Uploading And Embodied Cognition: A *Zygon®*, 54(1), 207-224. <https://doi.org/10.1111/zygo.12498>

Alizadehsalehi, S., & Yitmen, I. (2019). A concept for automated construction progress monitoring: Technologies adoption for benchmarking project performance control. *Arabian Journal for Science and Engineering*, 44, 4993-5008. <https://doi.org/10.1007/s13369-018-3669-1>

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*, 15(2). <https://doi.org/10.7759/cureus.35179>

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-236. <http://dx.doi.org/10.1257/jep.31.2.211>

Almgren, R., & Skobelev, D. (2020). Evolution of technology and technology governance. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(2), 22. <http://dx.doi.org/10.3390/joitmc6020022>

Alneyadi, S., Sithirasenan, E., & Muthukkumarasamy, V. (2016). A survey on data leakage prevention systems. *Journal of Network and Computer Applications*, 62, 137–152. <https://doi.org/10.1016/j.jnca.2016.01.008>

Alofi, A., Bokhari, M. A., Bahsoon, R., & Hendley, R. (2022). Optimizing the energy consumption of blockchain-based systems using evolutionary algorithms: A new problem

formulation. *IEEE Transactions on Sustainable Computing*, 7(4), 910-922. <https://doi.org/10.1109/TSUSC.2022.3160491>

Al-Shetwi, A. Q., Hannan, M. A., Jern, K. P., Mansur, M., & Mahlia, T. M. I. (2020). Grid-connected renewable energy sources: Review of the recent integration requirements and control methods. *Journal of Cleaner Production*, 253, 119831. <https://doi.org/10.1016/j.jclepro.2019.119831>

Alsuwat, E., Solaiman, S., & Alsuwat, H. (2023). Concept drift analysis and malware attack detection system using secure adaptive windowing. *Computers, Materials & Continua*, 75(2). <https://doi.org/10.32604/cmc.2023.035126>

Alwan, A. A., Ciupala, M. A., Brimicombe, A. J., Ghorashi, S. A., Baravalle, A., & Falcarin, P. (2022). Data quality challenges in large-scale cyber-physical systems: A systematic review. *Information Systems*, 105, 101951. <https://doi.org/10.1016/j.is.2021.101951>

Alzubaidi, L., Al-Sabaawi, A., Bai, J., Dukhan, A., Alkenani, A. H., Al-Asadi, A., ... & Gu, Y. (2023). Towards risk-free trustworthy artificial intelligence: Significance and requirements. *International Journal of Intelligent Systems*. <https://doi.org/10.1155/2023/4459198>

Amariles, D.R., & Baquero, P.M. (2023). Promises and limits of law for a human-centric artificial intelligence. *Computer Law & Security Review*, 48, 1-9. <http://dx.doi.org/10.1016/j.clsr.2023.105795>

Amatriain, X., & Basilico, J. (2015). Recommender systems in industry: A netflix case study. In *Recommender Systems Handbook* (pp. 385-419). Boston, MA: Springer US.

Amazon Web Services (AWS). (n.d). Transformez l'IA responsable de la théorie en pratique. <https://aws.amazon.com/fr/machine-learning/responsible-ai/>

Ambrose, T. (2021). David Attenborough urges leaders at Cop26 to be ‘motivated by hope not fear.’ *The Guardian*. <https://www.theguardian.com/environment/2021/nov/01/david-attenborough-urges-leaders-at-cop26-to-be-motivated-by-hope-not-fear>

Amit, R., & Han, X. (2017). Value creation through novel resource configurations in a digitally enabled world. *Strategic Entrepreneurship Journal*, 11(3), 228-242. <https://doi.org/10.1002/sej.1256>

Anand, V., & Saxena, D. (2013). Comparative study of modern web browsers based on their performance and evolution. In *2013 IEEE International Conference on Computational Intelligence and Computing Research* (pp. 1-5). IEEE. <http://dx.doi.org/10.1109/ICCIC.2013.6724273>

Anderson, J., Rainie, L., & Luchsinger, A. (2018). Artificial intelligence and the future of humans. *Pew Research Center*, 10(12). <http://tony-silva.com/eslefl/miscstudent/downloadpagearticles/AIhumanfuture-pew.pdf>

Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge University Press.

Andrade, M., Soares, A. M., Solé, M., Pereira, E., & Freitas, R. (2023). Threats of Pollutants Derived from Electronic Waste to Marine Bivalves: The Case of the Rare-Earth Element Yttrium. *Environmental Toxicology and Chemistry*, 42(1), 166-177. <https://doi.org/10.1002/etc.5508>

Andy, F. (2024). *Discovering Statistics Using IBM SPSS Statistics, Sixth d Edition*. Sage.

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. *arXiv preprint arXiv:2007.03051*.

Antunes, M. G., Mucharreira, P. R., Justino, M. R. T., & Texeira-Quirós, J. (2021). Effects of Total Quality Management (TQM) dimensions on innovation—evidence from SMEs. *Sustainability*, 13(18), 10095.

Anzoategui, D., Comin, D., Gertler, M., & Martinez, J. (2019). Endogenous technology adoption and R&D as sources of business cycle persistence. *American Economic Journal: Macroeconomics*, 11(3), 67-110. DOI: 10.1257/mac.20170269

Ardichvili, A., & Jondle, D. (2009). Integrative literature review: Ethical business cultures: A literature review and implications for HRD. *Human Resource Development Review*, 8(2), 223-244. <http://dx.doi.org/10.1177/1534484309334098>

Arif, E. M., Sarwo, S., Soderi, A., Rohman, A., & Fauzan, T. R. (2023). The Role of ChatGPT in Improving Cross-Cultural Team Management Performance. *Journal Minfo Polgan*, 12(1), 1464-1472. <https://doi.org/10.33395/jmp.v12i1.12816>

Arman, M., & Lamiya, U. R. (2023). Exploring the Implication of ChatGPT AI for Business: Efficiency and Challenges. *Journal of Innovation Information Technology and Application (JINITA)*, 5(1), 52-64. <http://dx.doi.org/10.31098/ijmadic.v1i2.1872>

Arora, A. (2020). Conceptualising artificial intelligence as a digital healthcare innovation: an introductory review. *Medical Devices: Evidence and Research*, 223-230. <https://doi.org/10.2147%2FMDER.S262590>

Arthur, W. B. (2010). *The nature of technology: What it is and how it evolves*. Penguin UK.

Arthur, W. B., & Polak, W. (2006). The evolution of technology within a simple computer model. *Complexity*, 11(5), 23-31. <https://doi.org/10.1002/cplx.20130>

Arumugam, D., Prasanna, P. K., & Marathe, R. R. (2023). Do algorithmic traders exploit volatility?. *Journal of Behavioral and Experimental Finance*, 37, 100778. <https://doi.org/10.1016/j.jbef.2022.100778>

Arvan, M. (2018). Mental time-travel, semantic flexibility, and A.I. ethics. *AI & Society*. 1-20. <https://doi.org/10.1007/s00146-018-0848-2>

Ashraf, A. R., Mackey, T. K., & Fittler, A. (2024). Search Engines and Generative Artificial Intelligence Integration: Public Health Risks and Recommendations to Safeguard Consumers Online. *JMIR Public Health and Surveillance*, 10(1), e53086. <https://doi.org/10.2196/53086>

Asimov, I. (1941). Three laws of robotics. *Asimov, I. Runaround*, 2.

Atlam, H. F., & Wills, G. B. (2020). IoT security, privacy, safety and ethics. In M. Farsi, A. Daneshkhah, A. Hosseinian-Far, & H. Jahankhani (Eds.), *Digital twin technologies and smart cities*. Internet of Things. Springer. https://doi.org/10.1007/978-3-030-18732-3_8

Attard-Frost, B., De los Ríos, A., & Walters, D. R. (2023). The ethics of AI business practices: a review of 47 AI ethics guidelines. *AI and Ethics*, 3(2), 389-406. <https://doi.org/10.1007/s43681-022-00156-6>

Augusto, L. M. (2021). From symbols to knowledge systems: A. Newell and HA Simon's contribution to symbolic AI. *Journal of Knowledge Structures and Systems*, 2(1), 29 – 62 <http://dx.doi.org/10.1177/154193120605000904>

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2), 645-709. <https://doi.org/10.1093/qje/qjaa004>

Aven, T. (2010). On how to define, understand and describe risk. *Reliability Engineering & System Safety*, 95(6), 623-631. <https://doi.org/10.1016/j.res.2010.01.011>

Avom, D., Nkengfack, H., Fotio, H. K., & Totouom, A. (2020). ICT and environmental quality in Sub-Saharan Africa: Effects and transmission channels. *Technological Forecasting and Social Change*, 155, 120028. <https://doi.org/10.1016/j.techfore.2020.120028>

Ayamga, M., Akaba, S., & Nyaaba, A. A. (2021). Multifaceted applicability of drones: A review. *Technological Forecasting and Social Change*, 167, 120677. <https://doi.org/10.1016/j.techfore.2021.120677>

Bächle, T. C., & Bareis, J. (2022). “Autonomous weapons” as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies. *European Journal of Futures Research*, 10(1), 1-18. <https://doi.org/10.1186/s40309-022-00202-w>

Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving the sustainable development goals. *Sustainability*, 14(5), 2497. <https://doi.org/10.3390/su14052497>

Bachura, E., Valecha, R., Chen, R., & Rao, H. R. (2022). The OPM data breach: An investigation of shared emotional reactions on Twitter. *MIS Quarterly*, 46(2), 881–910. <https://doi.org/10.25300/MISQ/2022/15596>

Bailey, A. H., Williams, A., & Cimpian, A. (2022). Based on billions of words on the internet, people men. *Science Advances*, 8(13), eabm2463. <https://doi.org/10.1126/sciadv.abm2463>

Ballardini, R. M., He, K., & Roos, T. (2019). AI-generated content: authorship and inventorship in the age of artificial intelligence. In *Online Distribution of Content in the EU* (pp. 117-135). Edward Elgar Publishing. <http://dx.doi.org/10.4337/9781788119900.00015>

- Ballouk, H., Meftah-Wali, S., Tabbah, G., & Ben Jabeur, S. (2023). Institutional investors and public authority ownership impact on green bonds issue: Evidence from France. *Journal of Innovation Economics & Management*, 41(2), 51-73. <https://doi.org/10.3917/jie.pr1.0130>
- Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative ai: A review of requirements, models, input-output formats, evaluation metrics, and challenges. *Future Internet*, 15(8), 260. <https://doi.org/10.3390/fi15080260>
- Barbut, S. (2014). Review: Automation and meat quality-global challenges. *Meat Science*, 96(1), 335–345. <https://doi.org/10.1016/j.meatsci.2013.07.002>
- Baregheh, A., Rowley, J., & Sambrook, S. (2009). Towards a multidisciplinary definition of innovation. *Management decision*, 47(8), 1323-1339. <https://doi.org/10.1108/00251740910984578>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California law review*, 671-732. <https://www.jstor.org/stable/24758720>
- Barometer Edelman Trust. (2019). Edelman Trust Barometer Global Report. Retrieved from: https://www.edelman.com/sites/g/files/aatuss191/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report.pdf.
- Baron, B., Musolesi, M. (2017). Interpretable machine learning for privacy-preserving pervasive systems. arXiv (pp. 1–10).
- Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics*, 143(1), 30-56. <https://doi.org/10.1016/j.jfineco.2021.05.047>
- Bashar, A. (2019). Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence*, 1(02), 73-82. <https://doi.org/10.36548/jaicn.2019.2.003>
- Bastidas, V., Oti-Sarpong, K., Nochta, T., Wan, L., Tang, J., & Schooling, J. (2023). Leadership for responsible digital innovation in the built environment: A socio-technical review for re-establishing competencies. *Journal of Urban Management*, 12(1), 57-73. <https://doi.org/10.1016/j.jum.2023.01.004>
- Baudier, P., Kondrateva, G., Ammi, C., Chang, V., & Schiavone, F. (2023). Digital transformation of healthcare during the COVID-19 pandemic: Patients' teleconsultation acceptance and trusting beliefs. *Technovation*, 120, 102547. <http://dx.doi.org/10.1016/j.technovation.2022.102547>
- Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9).
- Beijing Academy of Artificial Intelligence. (2019). Beijing AI Principles. <https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/>

- Belharar, O., Laamrani, B., & Chakor, A. (2023). Epistemological Paradigms in the social sciences: A Guide for Researchers. *International Journal of Global Community*, 6(2), 229-244. <http://dx.doi.org/10.5281/zenodo.8229124>
- Benbunan-Fich, R., Desouza, K. C., & Andersen, K. N. (2020). IT-enabled innovation in the public sector: introduction to the special issue. *European Journal of Information Systems*, 29(4), 323-328. <https://doi.org/10.1080/0960085X.2020.1814989>
- Benbunan-Fich, R., Desouza, K. C., & Andersen, K. N. (2020). IT-enabled innovation in the public sector: introduction to the special issue. *European Journal of Information Systems*, 29(4), 323-328. <https://doi.org/10.1080/0960085X.2020.1814989>
- Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, 22(2), 10. DOI: 10.17705/1jais.00662
- Benzaghta, M. A., Elwalda, A., Mousa, M. M., Erkan, I., & Rahman, M. (2021). SWOT analysis applications: An integrative literature review. *Journal of Global Business Insights*, 6(1), 55-73. <https://www.doi.org/10.5038/2640-6489.6.1.1148>
- Berthon, P. R., & Pitt, L. F. (2018). Brands, Truthiness and Post-Fact: Managing Brands in a Post-Rational World. *Journal of Macromarketing*, 38(2), 218–227. <https://doi.org/10.1177/0276146718755869>
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First monday*, 21(11-7). <https://ssrn.com/abstract=2982233>
- Beswick, J., & Gandhi, T. (2023). *Build Responsible Generative AI Applications: Introducing the RAFT Framework*. Dataiku. <https://content.dataiku.com/responsible-generative-ai>
- Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: A systematic literature review. *Journal of Banking and Financial Technology*, 4, 111–138. <https://doi.org/10.1007/s42786-020-00020-3>
- Biddle, C., & Schafft, K. A. (2015). Axiology and anomaly in the practice of mixed methods work: Pragmatism, valuation, and the transformative paradigm. *Journal of Mixed Methods Research*, 9, 320-334. <https://doi.org/10.1177/1558689814533157>
- Biggi, G., & Giuliani, E. (2022). The noxious consequences of innovation: what do we know? In *The Dark Side of Innovation* (pp. 12-34). Routledge.
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* (pp. 149-159). PMLR.
- Bird, S., Hutchinson, B., Kenthapadi, K., Kıcıman, E., & Mitchell, M. (2019). Fairness-Aware Machine Learning: Practical Challenges and Lessons Learned. *Companion Proceedings of The 2019 World Wide Web Conference*, 1297–1298. <https://doi.org/10.1145/3308560.3320086>
- Bitsch, L., Salles, A., Evers, K., Changeux, J.-P., Stahl, B., Aicardi, C., Burton Datta, S., Mahfoud, T., Reinsborough, M., Rose, N., Klüver, L., Ladegaard, S. F., Alves, E., Nordfalk,

F., Bådum, N., Eke, D., Knight, W., Grasenick, K., Romero, P. F., ... Rommelfanger, K. (2023). Ethics and Society in Brain Research: Implementing Responsible Research and Innovation (RRI) in the Human Brain Project (HBP). Zenodo. <https://doi.org/10.5281/zenodo.7736402>

Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., Conde-Torres, D., Antelo-Riveiro, P., Pineiro, A., & Garcia-Fandino, R. (2023). The role of AI in drug discovery: challenges, opportunities, and strategies. *Pharmaceuticals*, 16(6), 891. <https://doi.org/10.3390/ph16060891>

Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. *IEEE Access*, 10, 77110–77122. <https://doi.org/10.1109/ACCESS.2022.3191790>

Blichfeldt, H., & Faullant, R. (2021). Performance effects of digital technology adoption and product & service innovation—A process-industry perspective. *Technovation*, 105, 102275. <https://doi.org/10.1016/j.technovation.2021.102275>

Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In B.-J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. van den Hoven (Eds.), *Responsible innovation 2: Concepts, approaches, and applications* (pp. 19–35). Cham: Springer. DOI: 10.1007/978-3-319-17308-5_2

Blowfield, M., & Murray, A. (2014). *Corporate responsibility*. Oxford University Press, USA.

Blum, R., & Beyer, B. (2019). Achieving digital permanence. *Communications of the ACM*, 62(5), 36–42. <https://doi.org/10.1145/3303872>

Boar, B. H. (1997). *Strategic thinking for information technology: How to build the IT organization for the information age*. John Wiley & Sons, Inc..

Bocquet, R., & Dubouloz, S. (2020). Firm openness and managerial innovation: Rebalancing deliberate actions and institutional pressures. *Journal of Innovation Economics & Management*, 32(2), 43–74. <https://doi.org/10.3917/jie.032.0043>.

Boholm, Å., & Corvellec, H. (2011). A relational theory of risk. *Journal of risk research*, 14(2), 175–190. <https://doi.org/10.1080/13669877.2010.515313>

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in Healthcare* (pp. 25–60). Elsevier. <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>

Bolte, L., Vandemeulebroucke, T., & Van Wynsberghe, A. (2022). From an ethics of carefulness to an ethics of desirability: Going beyond current ethics approaches to sustainable AI. *Sustainability*, 14(8), 4472. <https://doi.org/10.3390/su14084472>

Bolte, L., Vandemeulebroucke, T., & van Wynsberghe, A. (2022). From an ethics of carefulness to an ethics of desirability: Going beyond current ethics approaches to sustainable AI. *Sustainability*, 14(8), 4472. <https://doi.org/10.3390/su14084472>

Bonfanti, M. E. (2022). Artificial intelligence and the offence-defence balance in cyber security. *Cyber Security: Socio-Technological Uncertainty and Political Fragmentation*. London: Routledge, 64-79.

Bonina, C., López-Berzosa, D., & Scarlata, M. (2021). Social, commercial, or both? An exploratory study of the identity orientation of digital social innovations. *Information Systems Journal*, 31(5), 695-716. <https://doi.org/10.1111/isj.12290>

Bonire, G., & Gbenga-Ilori, A. (2021). Towards artificial intelligence-based reduction of greenhouse gas emissions in the telecommunications industry. *Scientific African*, 12, e00823. <https://doi.org/10.1016/j.sciaf.2021.e00823>

Bookstaber, R. (2017). The end of theory: Financial crises, the failure of economics, and the sweep of human interaction. *Princeton University Press*. <https://doi.org/10.1515/9781400884964>.

Bordot, F. (2022). Artificial intelligence, robots and unemployment: evidence from OECD countries. *Journal of Innovation Economics & Management*, (1), 117-138. <http://dx.doi.org/10.3917/jie.037.0117>

Bostrom, N. (2014). *Superintelligence, Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bostrom, N., & Yudkowsky, E. (2018). The ethics of artificial intelligence. In *Artificial intelligence safety and security* (pp. 57-69). Chapman and Hall/CRC.

Bosu, M. F., & MacDonell, S. G. (2019). Experience: Quality benchmarking of datasets used in software effort estimation. *Journal of Data and Information Quality*, 11(4), 1–38.

Boustani, N. M. (2022). Artificial intelligence impact on banks clients and employees in an Asian developing country. *Journal of Asia Business Studies*, 16(2), 267–278. <https://doi.org/10.1108/JABS-09-2020-0376>

Bowen, G. (2020). Cybersecurity and AI: Reputational risks to firms and society. In N. Delener & C. Schweikert (Eds.), *Growth and sustainability in global markets: Theory, evidence, and practice* (pp. 1-5). Global Business and Technology Association.

Boyes, C. (2021, October 12). Supermarket opens chat checkouts to combat loneliness. *The New Daily*. Available at <https://thenewdaily.com.au/news/good-ews/2021/10/12/supermarket-chat-checkouts/>

Bracci, E. (2023). The loopholes of algorithmic public services: an “intelligent” accountability research agenda. *Accounting, Auditing & Accountability Journal*, 36(2), 739-763. <https://doi.org/10.1108/AAAJ-06-2022-5856>

Braithwaite, J., & Drahos, P. (2000). *Global business regulation*. Cambridge university press.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *interactions*, 25(5), 38-43. <https://doi.org/10.1145/3236669>

Bremer, S., Millar, K., Wright, N., & Kaiser, M. (2015). Responsible techno-innovation in aquaculture: Employing ethical engagement to explore attitudes to GM salmon in Northern Europe. *Aquaculture*, 437, 370–381. doi:[10.1016/j.aquaculture.2014.12.031](https://doi.org/10.1016/j.aquaculture.2014.12.031).

Brewer, R. N., & Kameswaran, V. (2018). Understanding the Power of Control in Autonomous Vehicles for People with Vision Impairment. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 185–197. <https://doi.org/10.1145/3234695.3236347>

Bringsjord, S. (2011). Psychometric artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, 23(3), 271–277. <https://doi.org/10.1080/0952813X.2010.502314>

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitsoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S. ... Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv. <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>

Bryson, J. J., & Malikova, H. (2021). Is there an AI cold war? *Global Perspectives*, 2(1), 24803. <https://doi.org/10.1525/gp.2021.24803>

Bubukayr, M. A. S., & Almaiah, M. A. (2021, July). Cybersecurity concerns in smart-phones and applications: A survey. In *2021 international conference on information technology (ICIT)* (pp. 725-731). IEEE. <https://doi.org/10.1109/ICIT52682.2021.9491691>

Buccella, A. (2023). “AI for all” is a matter of social justice. *AI and Ethics*, 3(4), 1143-1152. <https://doi.org/10.1007/s43681-022-00222-z>

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., ... & Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, 33(3), 606-659. <http://dx.doi.org/10.1111/1748-8583.12524>

Bughin, J., & Van Zeebroeck, N. (2018). Artificial intelligence: Why a digital base is critical. *The McKinsey Quarterly*. <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Artificial%20intelligence%20Why%20a%20digital%20base%20is%20critical/Artificial-intelligence-Why-a-digital-base-is-critical.ashx>

Bunduchi, R., & Candi, M. (2021). Technology legitimization: A product-level examination across the technology lifecycle. *British Journal of Management*, 33(4), 2052-2069. <https://doi.org/10.1111/1467-8551.12562>

Burget, M., Bardone, E., and Pedaste, M. (2017). Definitions and conceptual dimensions of responsible research and innovation: A literature review. *Science and Engineering Ethics*, 23, 1–19. <https://doi.org/10.1007/s11948-016-9782-1>

Burnham, K., & Anderson, D. (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). Springer-Verlag.

- Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law & Security Review*, 34(2), 257–268. <https://doi.org/10.1016/j.clsr.2018.01.004>
- Calo, R. (2017). Artificial intelligence policy: a primer and roadmap. *UCDL Rev.*, 51, 399.
- Campbell, C., Sands, S., Ferraro, C., Tsao, H. Y. J., & Mavrommatis, A. (2020). From data to action: How marketers can leverage AI. *Business Horizons*, 63(2), 227–243. <https://doi.org/10.1016/j.bushor.2019.12.002>
- Campbell, S. (2005). Determining overall risk. *Journal of risk research*, 8(7-8), 569–581. <https://doi.org/10.1080/13669870500118329>
- Campos, S., & Laurent, R. (2023). A Definition of General-Purpose AI Systems: Mitigating Risks from the Most Generally Capable Models. *Available at SSRN 4423706*.
- Canhoto, A. I., Quinton, S., Pera, R., Molinillo, S., & Simkin, L. (2021). Digital strategy aligning in SMEs: A dynamic capabilities perspective. *The Journal of Strategic Information Systems*, 30(3), 101682. <https://doi.org/10.1016/j.jsis.2021.101682>
- Capgemini Research Institute. (2023). Future-ready education: Empowering secondary school students with digital skills. <https://www.capgemini.com/insights/research-library/digital-skills-in-education/>
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4), 190–202. <http://dx.doi.org/10.1109%2FTMMS.1970.299942>
- Carle, A. (2024). Implementation challenges of innovation policies fostering sustainability: Evidence from a French public grant for technological startups. *Journal of Innovation Economics & Management*, 43, 39–67. <http://dx.doi.org/10.3917/jie.pr1.0152>
- Carroll, A. B., & Shabana, K. M. (2010). The business case for corporate social responsibility: A review of concepts, research and practice. *International journal of management reviews*, 12(1), 85–105. <https://doi.org/10.1111/j.1468-2370.2009.00275.x>
- Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180080. <https://doi.org/10.1098/rsta.2018.0080>
- Ceicyte, J., & Petraite, M. (2018). Networked responsibility approach for responsible innovation: Perspective of the firm. *Sustainability*, 10(6), 1720. <https://doi.org/10.3390/su10061720>
- Çelebi, T. Z. (2023). Environmental impact of cloud computing: an examination on energy consumption and carbon footprint. In Filiz Varol Gulduz. *Current Studies in Management Information Systems* (p. 77–100).

Cellan-Jones, R. (2014). Stephen Hawking warns artificial intelligence could end mankind. *BBC news*. Available at: <https://www.bbc.com/news/technology-30290540>

Chan, A. J., Hooi, L. W., & Ngui, K. S. (2021). Do digital literacies matter in employee engagement in digitalised workplace? *Journal of Asia Business Studies*, 15(3), 523-540. <https://doi.org/10.1108/JABS-08-2020-0318>

Chan, A., Okolo, C. T., Turner, Z., & Wang, A. (2021). The limits of global inclusion in AI development. *arXiv preprint arXiv:2102.01265*.

Chang, X. (2023). Gender Bias in Hiring: An Analysis of the Impact of Amazon's Recruiting Algorithm. *Advances in Economics Management and Political Sciences*, 23(1), 134-140 <http://dx.doi.org/10.54254/2754-1169/23/20230367>

Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., & Yeung, K. (2021). Trustworthy ai. *Reflections on Artificial Intelligence for Humanity*, 13-39. http://dx.doi.org/10.1007/978-3-030-69128-8_2

Chatterjee, S., Chaudhuri, R., & Vrontis, D. (2022). AI and digitalization in relationship management: Impact of adopting AI-embedded CRM system. *Journal of Business Research*, 150, 437–450. <https://doi.org/10.1016/j.jbusres.2022.06.033>

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, 314-347. <https://doi.org/10.1016/j.ins.2014.01.015>

Chen, D., & Zhao, H. (2012). Data security and privacy protection issues in cloud computing. In *2012 international conference on computer science and electronics engineering* 1, 647-651. IEEE. <https://doi.org/10.1109/ICCSEE.2012.193>

Chen, J., Yin, X., & Mei, L. (2018). Holistic innovation: An emerging innovation paradigm. *International Journal of Innovation Studies*, 2(1), 1-13. <https://doi.org/10.1016/j.ijis.2018.02.001>

Chen, L., Jiang, M., Jia, F., & Liu, G. (2022). Artificial intelligence adoption in business-to-business marketing: Toward a conceptual framework. *Journal of Business & Industrial Marketing*, 37(5), 1025–1044. <https://doi.org/10.1108/JBIM-09-2020-0448>

Chen, Y., & Esmailzadeh, P. (2024). GAI in medical practice: in-depth exploration of privacy and security challenges. *Journal of Medical Internet Research*, 26, e53008. <http://dx.doi.org/10.2196/53008>

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, 107, 1753. <https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security>

Chiang, C.-W., & Yin, M. (2022). Exploring the Effects of Machine Learning Literacy Interventions on Laypeople's Reliance on Machine Learning Models. *27th International Conference on Intelligent User Interfaces*, 148–161. <https://doi.org/10.1145/3490099.3511121>

- Chinchanchokchai, S., Thontirawong, P., & Chinchanchokchai, P. (2021). A tale of two recommender systems: The moderating role of consumer expertise on artificial intelligence based product recommendations. *Journal of Retailing and Consumer Services*, 61, 102528. <https://doi.org/10.1016/j.jretconser.2021.102528>
- Chirumalla, K., Reyes, L. G., & Toorajipour, R. (2022). Mapping a circular business opportunity in electric vehicle battery value chain: A multi-stakeholder framework to create a win-win-win situation. *Journal of Business Research*, 145, 569–582. <https://doi.org/10.1016/j.jbusres.2022.02.070>
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2), 153-163. <https://doi.org/10.1089/big.2016.0047>
- Chowdhury, N., Deka, G. C., & Sufian, M. A. (2024). Developing Trading Strategies in Decentralized Market Prediction by Using AI, ML, and Blockchain Technology. In *Blockchain and AI* (pp. 58-122). CRC Press.
- Chu, A. C., & Cozzi, G. (2018). Effects of patents versus R&D subsidies on income inequality. *Review of economic dynamics*, 29, 68-84. <https://doi.org/10.1016/j.red.2017.12.006>
- Chua, W. F., James, R., King, A., Lee, E., & Soderstrom, N. (2022). Task force on climate-related financial disclosures (TCFD) implementation: An overview and insights from the Australian Accounting Standards Board Dialogue Series. *Australian Accounting Review*, 32(3), 396–405. <https://doi.org/10.1111/auar.12388>
- Chui, M., Reoberts, R., Yee, L., Hazan, E., Singla, A., (2023). The economic potential of GAI: The next productivity frontier. *Mckinsey*. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#business-value>
- Ciriello, R. F., Richter, A., & Schwabe, G. (2018). Digital innovation. *Business & Information Systems Engineering*, 60, 563-569. <https://doi.org/10.1007/s12599-018-0559-8>
- Ciulla, J.B. (2022). The Use of Casuistry to Humanize Business. In: *Dion, M., Freeman, R.E., Dmytriiev, S.D. (eds) Humanizing Business. Issues in Business Ethics*, vol 53. Springer, Cham. https://doi.org/10.1007/978-3-030-72204-3_10
- Civan, S., Dogerlioglu, O., & Koksall, T. (2017). Can Information And Communication Technologies Change Individual's Values? *International Journal of Information, Business and Management*, 9(2), 104. <http://dx.doi.org/10.1027/1864-9335/a000123>
- Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*, 209, 118190. <https://doi.org/10.1016/j.eswa.2022.118190>
- Claessens, S., Frost, J., Turner, G., & Zhu, F. (2018). Fintech credit markets around the world: size, drivers and policy issues. *BIS Quarterly Review September*. <https://ssrn.com/abstract=3288096>

Clark, C. M. A., & Gevorkyan, A. V. (2020). Artificial Intelligence and Human Flourishing. *The American Journal of Economics and Sociology*, 79(4), 1307–1344. <https://doi.org/10.1111/ajes.12356>

Clements, R. (2021). Built to Fail: The Inherent Fragility of Algorithmic Stablecoins. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3952045>

Cloarec, J. (2022). Privacy controls as an information source to reduce data poisoning in artificial intelligence-powered personalization. *Journal of Business Research*, 152, 144-153. <https://doi.org/10.1016/j.jbusres.2022.07.045>

Cloarec, J., Meyer-Waarden, L., & Munzel, A. (2024). Transformative privacy calculus: Conceptualizing the personalization-privacy paradox on social media. *Psychology & Marketing*. <https://doi.org/10.1002/mar.21998>

Coad, A., Nightingale, P., Stilgoe, J., & Vezzani, A. (2022). The dark side of innovation. In *The Dark Side of Innovation* (pp. 1-11). Routledge.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The impact of artificial intelligence on innovation* (Vol. 24449). Cambridge, MA, USA: National bureau of economic research.

Colagrossi, A., Pesce, V., Silvestrini, S., Gonzalez-Arjona, D., Hermosin, P., & Battilana, M. (2023). Sensors. In V. Pesce, A. Colagrossi, & S. Silvestrini (Eds.), *Modern spacecraft guidance, navigation, and control* (pp. 253–336). Elsevier. <https://doi.org/10.1016/B978-0-323-90916-7.00006-8>

Collen, M. F., & Kulikowski, C. A. (2015). The development of digital computers. The History of Medical Informatics in the United States, 3-73. https://doi.org/10.1007/978-1-4471-6732-7_1

Contreras-Koterbay, S. (2020). In anxious anticipation of our imminent obsolescence. In C. Neill (Ed.), *Lacanian perspectives on Blade Runner 2049* (pp. 167–187). Palgrave Macmillan Cham. <https://doi.org/10.1007/978-3-030-56754-5>

Cookson, C. (2018, September 6). Artificial Intelligence faces public backlash, warns scientist. *Financial Times*. Retrieved from <https://www.ft.com/content/0b301152-b0f8-11e8-99ca-68cf89602132>

Correani, A., De Massis, A., Frattini, F., Petruzzelli, A. M., & Natalicchio, A. (2020). Implementing a digital strategy: Learning from the experience of three digital transformation projects. *California Management Review*, 62(4), 37–56. <https://doi.org/10.1177/0008125620934864>

Costello, K. (2019). Gartner Survey Shows 37 Percent of Organizations Have Implemented AI in Some Form. *Gartner*. <https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have>

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12. <https://doi.org/10.1080/14703297.2023.2190148>

Courtland, R. (2018). The bias detectives. *Nature*, 558(7710), 357-360. <https://doi.org/10.1038/d41586-018-05469-3>

Couture, S., & Toupin, S. (2019). What does the notion of “sovereignty” mean when referring to the digital?. *New Media & Society*, 21(10), 2305-2322. <https://doi.org/10.1177/1461444819865984>

Crane, A., Palazzo, G., Spence, L. J., & Matten, D. (2014). Contesting the value of “creating shared value.” *California Management Review*, 56(2), 130-153. <https://doi.org/10.1525/cmr.2014.56.2.130>

Crotty, M. (1998). *The foundations of social science research: meaning and perspective in the research process*. New South Wales: Allen & Unwin.

Cui, H., Zhang, H., Ganger, G. R., Gibbons, P. B., & Xing, E. P. (2016). GeePS: Scalable deep learning on distributed GPUs with a GPU-specialized parameter server. In *Proceedings of the eleventh European conference on computer systems* (pp. 1–16). ACM. <https://doi.org/10.1145/2901318.2901323>

Culnan, M. J., & Williams, C. C. (2009). How ethics can enhance organizational privacy: lessons from the choice point and TJX data breaches. *MIS Quarterly*, 673-687. <https://doi.org/10.2307/20650322>

Culp III, K., Eastwood, C., Turner, S., Goodman, M., & Ricketts, K. G. (2016). Using a SWOT analysis: Taking a look at your organization. *Community and Economic Development Publications*, 40546. https://uknowledge.uky.edu/ced_reports/3

D’Angelo, A. (2012). Innovation and export performance: a study of Italian high-tech SMEs. *Journal of management & governance*, 16, 393-423. <https://doi.org/10.1007/s10997-010-9157-y>

D’Cruz, P., Du, S., Noronha, E., Parboteeah, K. P., Trittin-Ulbrich, H., & Whelan, G. (2022). Technology, megatrends and work: Thoughts on the future of business ethics. *Journal of Business Ethics*, 180(3), 879-902. <https://doi.org/10.1007/s10551-022-05240-9>

Dąbrowska, J., Almpantopoulou, A., Brem, A., Chesbrough, H., Cucino, V., Di Minin, A., Giones, F., Hakala, H., Marullo, C., Mention, A. L., Mortara, L., Nørskov, S., Nylund, P. A., Oddo, C. M., Radziwon, A., & Ritala, P. (2022). Digital transformation, for better or worse: A critical multi-level research agenda. *R&D Management*, 52(5), 930–954. <http://dx.doi.org/10.1111/radm.12531>

Dana, L. P., Salamzadeh, A., Mortazavi, S., & Hadizadeh, M. (2022). Investigating the impact of international markets and new digital technologies on business innovation in emerging markets. *Sustainability*, 14(2), 983. <https://doi.org/10.3390/su14020983>

Dana, L. P., Salamzadeh, A., Mortazavi, S., & Hadizadeh, M. (2022). Investigating the impact of international markets and new digital technologies on business innovation in emerging markets. *Sustainability*, 14(2), 983. <https://doi.org/10.3390/su14020983>

Danielsson, J. (2022). *The Illusion of Control: Why Financial Crises Happen, and What We Can (and Can't) Do About It*. Yale University Press.

Danish government. (2019). National Strategy for Artificial Intelligence. *Ministry of Finance and Ministry of Industry, Business and Financial Affairs*. https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf

Das, S. (2019). *Global Energy Footprint of IoT Semiconductors*. Technical report.

Dash, R., McMurtrey, M., Rebman, C., & Kar, U. K. (2019). Application of artificial intelligence in automation of supply chain management. *Journal of Strategic Innovation and Sustainability*, 14(3), 43-53. <https://articlearchives.co/index.php/JSIS/article/view/4867>

Datasciencedojo. (2022). Machine learning 101: Supervised, unsupervised, reinforcement learning explained. <https://datasciencedojo.com/blog/machine-learning-101/>

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 17, 7-28. <https://doi.org/10.1007/s10833-015-9264-2>

Davila, T., Epstein, M., & Shelton, R. (2012). *Making innovation work: How to manage it, measure it, and profit from it*. FT press.

Davis, M., & Laas, K. (2014). “Broader impacts” or “Responsible research and innovation”? A comparison of two criteria for funding research in science and engineering. *Science and Engineering Ethics*, 20, 963-983. <https://doi.org/10.1007/s11948-013-9480-1>

Day, J. C., Janus, A., & Davis, J. (2005). Computer and Internet use in the United States: 2003. *Census Bureau*. <https://www.census.gov/library/publications/2005/demo/p23-208.html>

De Cremer, D., & Kasparov, G. (2022). The ethical AI—paradox: why better technology needs more and not less human responsibility. *AI and Ethics*, 2(1), 1-4. <https://doi.org/10.1007/s43681-021-00075-y>

De Haes, S., & Van Grembergen, W. (2009). An exploratory study into IT governance implementations and its impact on business/IT alignment. *Information Systems Management*, 26(2), 123–137. <https://doi.org/10.1080/10580530902794786>

De Laat, P. B. (2021). Companies committed to responsible AI: From principles towards implementation and regulation? *Philosophy & Technology*, 34, 1135–1193. <https://doi.org/10.1007/s13347-021-00474-3>

de Vries, A. (2024). Bitcoin’s growing water footprint. *Cell Reports Sustainability*, 1(1). <https://doi.org/10.1016/j.crsus.2023.100004>

De Vynck, G., Lerman, R., & Tiku, N. (2023, February 16). Microsoft’s AI chatbot is going off the rails. *The Washington Post*. <https://www.washingtonpost.com/technology/2023/02/16/microsoft-bing-ai-chatbot-sydney/>

Deldjoo, Y., Noia, T. D., & Merra, F. A. (2021). A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks. *ACM Computing Surveys (CSUR)*, 54(2), 1-38. <https://doi.org/10.1145/3439729>

Dembrower, K., Crippa, A., Colón, E., Eklund, M., & Strand, F. (2023). Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *The Lancet Digital Health*, 5(10), e703-e711. [https://doi.org/10.1016/S2589-7500\(23\)00153-X](https://doi.org/10.1016/S2589-7500(23)00153-X)

Demdoun, Z., Meraghni, O., & Bekkouche, L. (2021). The application of green accounting according to activity-based costing for an orientation towards a green economy: Field study. *International Journal of Digital Strategy, Governance, and Business Transformation*, 11(1), 1–15. <http://doi.org/10.4018/IJDSGBT.20210101.oa3>

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4), 197-387. <http://dx.doi.org/10.1561/20000000039>

Denton, E., Hanna, A., Amironesei, R., Smart, A., Nicole, H., & Scheuerman, M. K. (2020). Bringing the people back in: Contesting benchmark machine learning datasets. *arXiv preprint arXiv:2007.07399*.

Department of Industry, Science, Energy, and Resources, Australia. (2019). Australian Government AI Ethics Framework. <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework>

Department of Trade and Industry (DTI). (2003). *Competing in the global economy: The innovation challenge*. Innovation Report. London: DTI

Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological methods & research*, 50(2), 708-739. <https://doi.org/10.1177/0049124118799377>

Dhieb, N., Ghazzai, H., Besbes, H., & Massoud, Y. (2020). A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement. *IEEE Access*, 8, 58546–58558. <https://doi.org/10.1109/ACCESS.2020.2983300>

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). Supporting Social and Emotional Well-Being with Artificial Intelligence. In *Bridging Human Intelligence and Artificial Intelligence* (pp. 125-138). Cham: Springer International Publishing.

Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124, 329-341. <https://doi.org/10.1016/j.jbusres.2020.11.037>

Di Vaio, A., Palladino, R., Pezzi, A., & Kalisz, D. E. (2021). The role of digital innovation in knowledge management systems: A systematic literature review. *Journal of business research*, 123, 220-231. <https://doi.org/10.1016/j.jbusres.2020.09.042>

Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., ... & Zevenbergen, B. (2017). Principles for accountable algorithms and a social impact statement for algorithms. *FAT/ML*.

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, 99, 101896. <https://doi.org/10.1016/j.inffus.2023.101896>

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer. <http://doi.org/10.1007/978-3-030-30371-6>

Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018). Unsupervised Learning Based On Artificial Neural Network: A Review. *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, 322–327. <https://doi.org/10.1109/CBS.2018.8612259>

Drucker, P.F. (1954). *The practice of management*. London: Heinemann

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, 129, 961-974. <https://doi.org/10.1016/j.jbusres.2020.08.024>

Duarte, R., Mainar, A., & Sánchez-Chóliz, J. (2013). The role of consumption patterns, demand and technological factors on the recent evolution of CO2 emissions in a group of advanced economies. *Ecological Economics*, 96, 1-13. <https://doi.org/10.1016/j.ecolecon.2013.09.007>

Dunn Jr, W. D., Cobb, J., Levey, A. I., & Gutman, D. A. (2016). REDLeTr: Workflow and tools to support the migration of legacy clinical data capture systems to REDCap. *International journal of medical informatics*, 93, 103-110. <https://doi.org/10.1016/j.ijmedinf.2016.06.015>

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

Eitel-Porter, R. (2021). Beyond the promise: implementing ethical AI. *AI and Ethics*, 1, 73-80. <https://doi.org/10.1007/s43681-020-00011-6>

El Atillah, E. (2023). Man ends his life after an AI chatbot ‘encouraged’ him to sacrifice himself to stop climate change. *Euronews*. <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate->

Eling, M., & Schnell, W. (2016). What do we know about cyber risk and cyber risk insurance?. *The Journal of Risk Finance*, 17(5), 474-491. <https://doi.org/10.1108/JRF-09-2016-0122>

Elkington, J. (1998). Accounting for the triple bottom line. *Measuring Business Excellence*, 2(3), 18–22. <https://doi.org/10.1108/eb025539>

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

European Commission (EC) (2023). *Artificial intelligence act*. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

European Commission (EC). (2020). Impact Assessment of the Regulation on Artificial Intelligence. Retrieved from: <https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence>

European Commission (EC). (2021). Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. <https://perma.cc/RWT9-9D97>

European Commission for the Efficiency of Justice (CEPEJ). (2018). CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment. <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>

European Commission. (2011). *A renewed EU strategy 2011-14 for Corporate Social Responsibility*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels, 25.10.2011.

European Commission. (2018). AI for humanity. https://ai-watch.ec.europa.eu/countries/france/france-ai-strategy-report_en

European Commission. (2018). Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee and The Committee of the Regions. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>

European Commission. (2018a). Artificial intelligence for Europe. COM(2018) 237 final. Brussels.

European Commission. (2018b) Artificial Intelligence: A European Perspective. Luxembourg: Publications Office of the European Union.

European Commission. (2019). France AI Strategy Report. https://ai-watch.ec.europa.eu/countries/france/france-ai-strategy-report_en

European Commission. (2020). *Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment*. *Shaping Europe's Digital Future - European Commission*. <https://digital->

strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

European Commission. (2020a). On artificial intelligence – A European approach to excellence and trust. White Paper. COM(2020) 65 final. Brussels.

European Commission. (2020b). White Paper on Artificial Intelligence: a European approach to excellence and trust. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Parliament. (2017). Civil law rules on robotics. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). Official Journal of the European Union, C 252, 239-257.

European Parliament. (2024). *Legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (P9_TA(2024)0138)*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

EY. (2024). The Artificial Intelligence (AI) global regulatory landscape Policy trends and considerations to build confidence in AI. https://www.ey.com/en_gl/ai/how-to-navigate-global-trends-in-artificial-intelligence-regulation

Fagerberg, J., Srholec, M., & Verspagen, B. (2010). The role of innovation in development. *Review of economics and institutions*, 1(2). <http://dx.doi.org/10.5202/rei.v1i2.15>

Farber, H. S., Herbst, D., Kuziemko, I., & Naidu, S. (2021). Unions and inequality over the twentieth century: New evidence from survey data. *The Quarterly Journal of Economics*, 136(3), 1325-1385. <https://doi.org/10.1093/qje/qjab012>

Fawcett, A., Mullan, S., & McGreevy, P. (2018). Application of Fraser’s “Practical” Ethic in Veterinary Practice, and Its Compatibility with a “One Welfare” Framework. *Animals*, 8(7), 109. <https://doi.org/10.3390/ani8070109>

Feast, L., & Melles, G. (2010, June). Epistemological positions in design research: A brief review of the literature. In *2nd international conference on design education, University of New South Wales, Sydney, Australia*.

Fedeli, G. (2020). ‘Fake news’ meets tourism: A proposed research agenda. *Annals of Tourism Research*, 80, 102684. <https://doi.org/10.1016/j.annals.2019.02.002>

Feijóo, C., Kwon, Y., Bauer, J. M., Bohlin, E., Howell, B., Jain, R., ... & Xia, J. (2020). Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy. *Telecommunications Policy*, 44(6), 101988. <https://doi.org/10.1016/j.telpol.2020.101988>

Feldstein, S. (2021). Issues on the Frontlines of Technology and Politics. *Carnegie Endowment for International Peace*. https://carnegieendowment.org/files/202110-Feldstein_Frontlines_final3.pdf

Ferraro, C., Demsar, V., Sands, S., Restrepo, M., & Campbell, C. (2024). The paradoxes of GAI-enabled customer service: A guide for managers. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2024.04.013>

Ferreboeuf, H., (2019). Lean ICT: Towards digital sobriety. *The Shift Project*. Available at: <https://theshiftproject.org/en/article/lean-ict-our-new-report/>

Fetzer, J. H. (1990). *What is artificial intelligence?* (pp. 3-27). Springer Netherlands.

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI: Challenges and opportunities. *Business & information systems engineering*, 62, 379-384. <https://doi.org/10.1007/s12599-020-00650-3>

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). GAI. *Business & Information Systems Engineering*, 66(1), 111-126. <https://doi.org/10.1007/s12599-023-00834-7>

Fichman, R. G., Dos Santos, B. L., & Zheng, Z. (2014). Digital innovation as a fundamental and powerful concept in the information systems curriculum. *MIS quarterly*, 38(2), 329-A15. <http://dx.doi.org/10.25300/MISQ/2014/38.2.01>

Firlej, M., & Taeihagh, A. (2021). Regulating human control over autonomous systems. *Regulation & Governance*, 15(4), 1071–1091. <https://doi.org/10.1111/rego.12344>

Fisher, A. (2014). *Metaethics: an introduction*. Routledge.

FitzGerald, D. (2015). Data Centers and Hidden Water Use. *The Wall Street Journal*. Available at: <https://www.wsj.com/articles/data-centers-1435168386>

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1). <https://dx.doi.org/10.2139/ssrn.3518482>

Floridi L. (2011). *The philosophy of information*. New York: Oxford University Press.

Floridi, L. (1996). Brave. Net. World: the Internet as a disinformation superhighway?. *The Electronic Library*, 14(6), 509-514. <http://dx.doi.org/10.1108/eb045517>

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology*, 33, 369-378. <https://doi.org/10.1007/s13347-020-00423-6>

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 535–545). Wiley. <https://doi.org/10.1002/9781119815075.ch45>

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., [Valcke](#), P., & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and

recommendations. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>

Floridi, L., Cows, 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. https://doi.org/10.1007/978-3-030-81907-1_3

Flowers, J. C. (2019). Strong and Weak AI: Deweyan Considerations. In *AAAI spring symposium: Towards conscious AI systems*, 2287(7).

Flyverbom, M., Deibert, R., & Matten, D. (2019). The governance of digital technology, big data, and the internet: New roles and responsibilities for business. *Business & Society*, 58(1), 3-19. <https://doi.org/10.1177/0007650317727540>

Forsberg, E. M., Quaglio, G., O'Kane, H., Karapiperis, T., Van Woensel, L., & Arnaldi, S. (2015). Assessment of science and technologies: Advising for and with responsibility. *Technology in Society*, 42, 21-27. <https://doi.org/10.1016/j.techsoc.2014.12.004>

Fortes, N., Rita, P., & Pagani, M. (2017). The effects of privacy concerns, perceived risk and trust on online purchasing behaviour. *International Journal of Internet Marketing and Advertising*, 11(4), 307-329. <https://doi.org/10.1504/IJIMA.2017.087269>

Franco, S., Mandla, V. R., & Rao, K. R. M. (2017). Urbanization, energy consumption and emissions in the Indian context A review. *Renewable and Sustainable Energy Reviews*, 71, 898-907. <https://doi.org/10.1016/j.rser.2016.12.117>

Franzke, A. S. (2022). An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*, 20(4), 401-423. <http://dx.doi.org/10.1108/JICES-12-2020-0125>

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Boston: Pitman.

Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G., & Friday, A. (2021). The climate impact of ICT: A review of estimates, trends and regulations. <https://doi.org/10.48550/arXiv.2102.02622>

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>

Friedkin, N. E. (2004). Social cohesion. *Annu. Rev. Sociol.*, 30, 409-425. <http://dx.doi.org/10.1146/annurev.soc.30.012703.110625>

- Friedrich, J., Becker, M., Kramer, F., Wirth, M., & Schneider, M. (2020). Incentive design and gamification for knowledge management. *Journal of Business Research*, 106, 341-352. <https://doi.org/10.1016/j.jbusres.2019.02.009>
- Friend, F. J. (2006). Google Scholar: Potentially good for users of academic information. *Journal of electronic publishing*, 9(1). <https://discovery.ucl.ac.uk/id/eprint/1771>
- Fryer, D. (2020). Race, reform, & progressive prosecution. *The Journal of Criminal Law and Criminology*, 110(4), 769–802. <https://www.jstor.org/stable/48595414>
- Fukuda-Parr, S., & Gibbons, E. (2021). Emerging consensus on ‘ethical AI’: Human rights critique of stakeholder guidelines. *Global Policy*, 12, 32-44. <https://doi.org/10.1111/1758-5899.12965>
- Furrer, F. J. (2022). Principles for Safety. In *Safety and Security of Cyber-Physical Systems: Engineering Dependable Software Using Principle-based Development* (pp. 433-448). Wiesbaden: Springer Fachmedien Wiesbaden.
- Future of Life Institute. (2017). Asilomar AI Principles. <https://futureoflife.org/open-letter/ai-principles/>
- Gallersdörfer, U., Klaaßen, L., & Stoll, C. (2020). Energy consumption of cryptocurrencies beyond bitcoin. *Joule*, 4(9), 1843-1846. <https://doi.org/10.1016/j.joule.2020.08.005>
- Gao, Z., & Hands, D. (2021). The strategic role of design in driving digital innovation: A theoretical foundation. *Journal of technology management & innovation*, 16(1), 58-66. <http://dx.doi.org/10.4067/S0718-27242021000100058>
- Garcia-Blandon, J., Castillo-Merino, D., Argilés-Bosch, J. M., & Ravenda, D. (2020). Board-level ethics committees in large European firms. *Business Ethics: A European Review*, 29(4), 824-841. <https://doi.org/10.1111/beer.12305>
- Gardner, H. (1999). Intelligence reframed: Multiple intelligences for the 21st century. New York, NY: Basic Books.
- Garner, B., Thornton, C., Pawluk, A. L., Cortez, R. M., Johnston, W., & Ayala, C. (2022). Utilizing text-mining to explore consumer happiness within tourism destinations. *Journal of Business Research*, 139, 1366-1377. <http://dx.doi.org/10.1016/j.jbusres.2021.08.025>
- Gault, F. (2018). Defining and measuring innovation in all sectors of the economy. *Research Policy*, 47(3), 617-622. <https://doi.org/10.1016/j.respol.2018.01.007>
- Gavard-Perret, M. L., Gotteland, D., Haon, C., & Jolibert, A. (2008). Méthodologie de la recherche. *Editions Pearson Education France*, 42. ISBN : 978-2-7440-7241-3
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé, III, H., Crawford, K. (2018). Datasheets for datasets. arXiv (pp. 1–17).

- Genus, A., & Iskandarova, M. (2018). Responsible innovation: its institutionalisation and a critique. *Technological Forecasting and Social Change*, 128, 1-9.
<https://doi.org/10.1016/j.techfore.2017.09.029>
- George, A. S., George, A. H., & Martin, A. G. (2023). The environmental impact of ai: A case study of water consumption by chat gpt. *Partners Universal International Innovation Journal*, 1(2), 97-104. <https://doi.org/10.5281/zenodo.7855594>
- Georgieff, A., & Hyee, R. (2022). Artificial intelligence and employment: New cross-country evidence. *Frontiers in Artificial Intelligence*, 5, 832736.
<https://doi.org/10.3389/frai.2022.832736>
- German Banking Industry Committee (2020). Comments on EU consultation on the white paper on artificial intelligence (AI). Retrieved from: <https://die-dk.de/en/topics/article/comments-eu-consultation-white-paper-artificial-intelligence-ai/>
- Ghaffari, S., Yousefimehr, B., & Ghatee, M. (2024, February). Generative-AI in E-Commerce: Use-Cases and Implementations. In *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)* (pp. 1-5). IEEE.
<https://doi.org/10.1109/AISP61396.2024.10475266>
- Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, 115(18), 4613–4618.
<https://doi.org/10.1073/pnas.1716999115>
- Giannakopoulos, K., Kavadella, A., Aaqel Salim, A., Stamatopoulos, V., & Kaklamanos, E. G. (2023). Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: Comparative mixed methods study. *Journal of Medical Internet research*, 25, e51580.
<https://doi.org/10.2196/51580>
- Gibney, E. (2018). The scant science behind Cambridge Analytica's controversial marketing techniques. *Nature*, 10. <http://dx.doi.org/10.1038/d41586-018-03880-4>
- Giffi, C. A., Vitale Jr, J., Schiller, T., & Robinson, R. (2018). A reality check on advanced vehicle technologies. *Insights exploring new automotive business models and consumer preferences*, 8.
- Gil-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R., & Lozano-Quilis, J. A. (2020). Customer relationship management: digital transformation and sustainable business model innovation. *Economic research-Ekonomska istraživanja*, 33(1), 2733-2750.
<https://doi.org/10.1080/1331677X.2019.1676283>
- Giovanola, B., & Tiribelli, S. (2022). Beyond bias and discrimination: Redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & Society*, 38, 549–563.
<https://doi.org/10.1007/s00146-022-01455-6>
- Goggin, B. (2019). Inside Facebook's suicide algorithm: Here's how the company uses artificial intelligence to predict your mental state from your posts. Retrieved from:

<https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12?r=US&IR=T>

Goggin, G. (2013). Changing media with mobiles. A companion to new media dynamics, 191-208. DOI:10.1002/9781118321607

Goldman Sachs. (2023). GAI could raise global GDP by 7%. <https://www.goldmansachs.com/insights/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>

Göllner, S., Tropmann-Frick, M., & Brumen, B. (2024). Towards a Definition of a Responsible Artificial Intelligence. In *Information Modelling and Knowledge Bases XXXV* (pp. 40-56). IOS Press.

Gomes, H. M., Grzenda, M., Mello, R., Read, J., Le Nguyen, M. H., & Bifet, A. (2022). A survey on semi-supervised learning for delayed partially labelled data streams. *ACM Computing Surveys*, 55(4), 1-42. <https://doi.org/10.48550/arXiv.2106.09170>

Gonçalves, D., Bergquist, M., Alänge, S., & Bunk, R. (2022). How digital tools align with organizational agility and strengthen digital innovation in automotive startups. *Procedia Computer Science*, 196, 107-116. <https://doi.org/10.1016/j.procs.2021.11.079>

Gondwe, G. (2023). Exploring the multifaceted nature of generative AI in journalism studies: A typology of scholarly definitions. <https://dx.doi.org/10.2139/ssrn.4465446>

Gonzalez-Jimenez, H. (2018). Taking the fiction out of science fiction: (Self-aware) robots and what they mean for society, retailers and marketers. *Futures*, 98, 49–56. <https://doi.org/10.1016/j.futures.2018.01.004>

González-Tejero, C. B., Ribeiro-Navarrete, B., Cano-Marin, E., & McDowell, W. C. (2023). A systematic literature review on the role of artificial intelligence in entrepreneurial activity. *International Journal on Semantic Web and Information Systems*, 19(1), 1–16. <https://doi.org/10.4018/IJSWIS.318448>

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pp. 2672–2680

Gordon, R. J. (2013). US productivity Growth: The Slowdown has returned after a temporary revival. *International Productivity Monitor*, (25), 13.

Gottschalk, S., Yigitbas, E., Nowosad, A., and Engels, G. (2023). Continuous situation-specific development of business models: Knowledge provision, method composition, and method enactment. *Software and Systems Modeling*, 22, 47–73. <https://doi.org/10.1007/s10270-022-01018-9>

Goujard, C. (2023). Italian privacy regulator bans ChatGPT. *Politico*. <https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/>

Goujard, C. (2023, March 31). Italian privacy regulator bans ChatGPT. *Politico*. <https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/>

Government of Canada. (2023). Global Declaration on Information Integrity Online. https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/peace_security-paix_securite/declaration_information_integrity-integrite.aspx?lang=eng

Government of Ireland. (2020). Irelands' national submission to the public consultation on the EU white paper on artificial intelligence. <https://enterprise.gov.ie/en/publications/national-submission-eu-white-paper-on-ai.html>

Goyal, A., & Aneja, R. (2020). Artificial intelligence and income inequality: Do technological changes and worker's position matter?. *Journal of Public Affairs*, 20(4), e2326. <https://doi.org/10.1002/pa.2326>

Graetz, G., & Michaels, G. (2015). *Robots at work: the impact on productivity and jobs* (No. 447). Centre for Economic Performance, LSE.

Granjou, C., Walker, J., & Salazar, J. F. (2017). The politics of anticipation: On knowing and governing environmental futures. *Futures*, 92, 5-11. <https://doi.org/10.1016/j.futures.2017.05.007>

Greenleaf, G. (2019). G20 makes declaration of 'Data Free Flow With Trust': Support and dissent. <https://ssrn.com/abstract=3514407>

Gregory, R., Wagner, H. T., Tumbas, S., & Drechsler, K. (2019). At the crossroads between digital innovation and digital transformation. *Fortieth International Conference on Information Systems*. <https://aisel.aisnet.org/icis2019>

Grodal, S., Krabbe, A. D., & Chang-Zunino, M. (2023). The evolution of technology. *Academy of Management Annals*, 17(1), 141-180. <https://doi.org/10.5465/annals.2021.0086>

GSMA. (2021). IoT for Development: Use cases delivering impact. Available at: <http://www.gsma.com/mobilefordevelopment/digitalutilities/>

Gunkel, D. J. (2018). The other question: Can and should robots have rights? *Ethics and Information Technology*, 20, 87-99. <https://doi.org/10.1007/s10676-017-9442-4>

Guo, X., Li, M., Wang, Y., & Mardani, A. (2023). Does digital transformation improve the firm's performance? From the perspective of digitalization paradox and managerial myopia. *Journal of Business Research*, 163, 113868. <https://doi.org/10.1016/j.jbusres.2023.113868>

Gupta, B. B., Gaurav, A., Panigrahi, P. K., & Arya, V. (2023). Analysis of artificial intelligence-based technologies and approaches on sustainable entrepreneurship. *Technological Forecasting and Social Change*, 186, 122152. <https://doi.org/10.1016/j.techfore.2022.122152>

Gupta, S., Modgil, S., Bhattacharyya, S., and Bose, I. (2022). Artificial intelligence for decision support systems in the field of operations research: Review and future scope of research. *Annals of Operations Research*, 308, 215-274. <https://doi.org/10.1007/s10479-020-03856-6>

Guthrie, J., Cuganesan, S., & Ward, L. (2006). Legitimacy theory: A story of reporting social and environmental matters within the Australian food and beverage industry. In *Asia Pacific Interdisciplinary Research in Accounting Conference (5th: 2007)* (pp. 1-35). APIRA 2007 Organising Committee. <https://ssrn.com/abstract=1360518>

Gwagwa, A., Kazim, E., Kachidza, P., Hilliard, A., Siminyu, K., Smith, M., & Shawe-Taylor, J. (2021). Road map for research on responsible artificial intelligence for development (AI4D) in African countries: The case study of agriculture. *Patterns*, 2(12), 100381. <https://doi.org/10.1016/j.patter.2021.100381>

Hadid, A., Chakraborty, T., & Busby, D. (2024). When geoscience meets generative AI and large language models: Foundations, trends, and future challenges. *Expert Systems*, e13654.

Haenlein, M., Huang, M.-H., & Kaplan, A. (2022). Guest Editorial: Business Ethics in the Era of Artificial Intelligence. *Journal of Business Ethics*, 178(4), 867–869. <https://doi.org/10.1007/s10551-022-05060-x>

Hagedoorn, J., Carayannis, E., & Alexander, J. (2001). Strange bedfellows in the personal computer industry: technology alliances between IBM and Apple. *Research Policy*, 30(5), 837-849. [http://dx.doi.org/10.1016/S0048-7333\(00\)00125-6](http://dx.doi.org/10.1016/S0048-7333(00)00125-6)

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99-120. <https://doi.org/10.1007/s11023-020-09517-8>

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2125–2126. <https://doi.org/10.1145/2939672.2945386>

Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social bots and the spread of disinformation in social media: the challenges of artificial intelligence. *British Journal of Management*, 33(3), 1238-1253. <https://doi.org/10.1111/1467-8551.12554>

Halder, S., Attili, V. S. P., & Gupta, V. (2022). Information privacy assimilation: An organizational framework. *International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)*, 12(1), 1–17. <http://doi.org/10.4018/IJDSGBT.313954>

Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2022). Medical 4.0 technologies for healthcare: Features, capabilities, and applications. *Internet of Things and Cyber-Physical Systems*, 2, 12-30. <https://doi.org/10.1016/j.iotcps.2022.04.001>

Halim, Z., Sulaiman, M., Waqas, M., & Aydın, D. (2023). Deep neural network-based identification of driving risk utilizing driver dependent vehicle driving features: A scheme for critical infrastructure protection. *Journal of Ambient Intelligence and Humanized Computing*, 14(9), 11747-11765. <http://dx.doi.org/10.1007/s12652-022-03734-y>

Hall, R. F. (2013). Mixed methods: In search of a paradigm. In T. Le, & Q. Le (Eds.), *Conducting research in a changing and challenging world* (pp. 71-78). New York: Nova Science Publishers Inc

- Halme, M., Roome, N., & Dobers, P. (2009). Corporate responsibility: Reflections on context and consequences. *Scandinavian Journal of Management*, 25(1), 1-9. <https://doi.org/10.1016/j.scaman.2008.12.001>
- Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36-S40. <https://doi.org/10.1016/j.metabol.2017.01.011>
- Hamon, R., Junklewitz, H., Sanchez, I., Malgieri, G., & De Hert, P. (2022). Bridging the gap between AI and explainability in the GDPR: towards trustworthiness-by-design in automated decision-making. *IEEE Computational Intelligence Magazine*, 17(1), 72-85. <https://doi.org/10.1109/MCI.2021.3129960>
- Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., ... & Zhu, J. (2021). Pre-trained models: Past, present and future. *AI Open*, 2, 225-250. <https://doi.org/10.1016/j.aiopen.2021.08.002>
- Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies*, 58(5), 1159-1197. <https://doi.org/10.1111/joms.12639>
- Hannan, M. A., Al-Shetwi, A. Q., Ker, P. J., Begum, R. A., Mansor, M., Rahman, S. A., ... & Muttaqi, K. M. (2021). Impact of renewable energy utilization and artificial intelligence in achieving sustainable development goals. *Energy Reports*, 7, 5359-5373. <https://doi.org/10.1016/j.egyr.2021.08.172>
- Hansen, E. G., & Grosse-Dunker, F. (2012). Sustainability-oriented innovation. *Encyclopedia of Corporate Social Responsibility: Heidelberg, Germany*. <https://ssrn.com/abstract=2191679>
- Hansen, S. T., & Moe, E. (2021). Renewable energy expansion or the preservation of national energy sovereignty? Norwegian renewable energy policy meets resource nationalism. *Political Geography*, 99, 102760. <https://doi.org/10.1016/j.polgeo.2022.102760>
- Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, 7(1), 21. <https://doi.org/10.1186/s40537-020-00291-z>
- Hassan, M., Mies, R., & Jochem, R. (2023). Key enablers towards mature company-community collaboration in open source hardware. *Journal of Innovation Economics & Management*, 159-191. <https://doi.org/10.3917/jie.pr1.0135>
- Heller, M. A., & Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. *Science*, 280(5364), 698-701. <https://doi.org/10.1126/science.280.5364.698>
- Hellström, T. (2003). Systemic innovation and risk: technology assessment and the challenge of responsible innovation. *Technology in Society*, 25(3), 369-384. [https://doi.org/10.1016/S0160-791X\(03\)00041-1](https://doi.org/10.1016/S0160-791X(03)00041-1)
- Helsper, E. (2021). The digital disconnect: The social causes and consequences of digital inequalities. *The Digital Disconnect*, 1-232.

- Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), 1-43. <https://doi.org/10.48550/arXiv.2002.05651>
- Henderson, P., Li, X., Jurafsky, D., Hashimoto, T., Lemley, M. A., & Liang, P. (2023). Foundation Models and Fair Use. *Stanford Law and Economics Olin Working Paper*, 584. <https://dx.doi.org/10.2139/ssrn.4404340>
- Henderson, P., Li, X., Jurafsky, D., Hashimoto, T., Lemley, M. A., & Liang, P. (2023). Foundation models and fair use. *Stanford Law and Economics Olin Working Paper*, 584. <https://dx.doi.org/10.2139/ssrn.4404340>
- Herberger, T. A., & Dötsch, J. J. (2021). The Means Justifies the End? Digitalization and Sustainability as a Social Challenge. A Plea for an Integrative View. In *Digitalization, Digital Transformation and Sustainability in the Global Economy: Risks and Opportunities* (pp. 1-8). Springer International Publishing.
- Hercheui, M., & Guerard, S. (2022). The impact of artificial intelligence biases: Algorithmic discrimination. *Global Conference on Business and Finance Proceedings*, 17(1). <https://doi.org/10.2811/25847>
- Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics*, 1(1), 41-47. <https://doi.org/10.1007/s43681-020-00008-1>
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, 332(6025), 60-65. <http://dx.doi.org/10.1126/science.1200970>
- Hilgartner, S. (1992). The social construction of risk objects: Or, how to pry open networks of risk. In J. F. Short Jr. & L. Clarke (Eds.), *Organizations, uncertainties, and risk* (pp. 39–51). Boulder, CO: Westview..
- Hill, K. (2020). The secretive company that might end privacy as we know it. *The New York Times*. Retrieved from: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>
- Hilty, L. M., & Hercheui, M. D. (2010, September). ICT and sustainable development. In *IFIP International Conference on Human Choice and Computers* (pp. 227-235). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. *Information and Organization*, 28(1), 52–61. <https://doi.org/10.1016/j.infoandorg.2018.02.004>
- Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(282-317), 2. <https://doi.org/10.7551/mitpress/3349.003.0005>
- Hirvonen, N., Jylhä, V., Lao, Y., & Larsson, S. (2023). Artificial intelligence in the information ecosystem: Affordances for everyday information seeking. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24860>

HLEGAI [High Level Expert Group on Artificial Intelligence] European Commission. (2019). Ethics Guidelines for Trustworthy AI.

Holder, E. (2014). Attorney General Eric Holder Speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference. *The United States Department of Justice*. <https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th>.

Holm, J. R., & Lorenz, E. (2022). The impact of artificial intelligence on skills at work in Denmark. *New Technology, Work and Employment*, 37(1), 79–101. <https://doi.org/10.1111/ntwe.12215>

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 1-23. <http://dx.doi.org/10.1007/s40593-021-00239-1>

Holmström, J., & Carroll, N. (2024). How organizations can innovate with GAI. *Business Horizons*. <https://doi.org/10.1016/j.bushor.2024.02.010>

Holweg, M., Younger, R., & Wen, Y. (2022). The reputational risks of AI. *California Management Review Insights*. <https://cmr.berkeley.edu/2022/01/the-reputational-risks-of-ai/>

Horner, N. C., Shehabi, A., & Azevedo, I. L. (2016). Known unknowns: indirect energy effects of information and communication technology. *Environmental Research Letters*, 11(10), <https://doi.org/10.1088/1748-9326/11/10/103001>

Horner, N., & Azevedo, I. (2016). Power usage effectiveness in data centers: overloaded and underachieving. *The Electricity Journal*, 29(4), 61-69. <https://doi.org/10.1016/j.tej.2016.04.011>

Hosmer, D. W., Lemeshow, S., & Cook, E. (2000). Applied binary logistic regression 2nd edition. New York: John Wiley and Sons Inc.

Hossain, M. (2015). Crowdsourcing in business and management disciplines: an integrative literature review. *Journal of Global Entrepreneurship Research*, 5, 1-19. <https://doi.org/10.1186/s40497-015-0039-2>

House of Lords, UK Parliament. (2017). House of Lords Report on AI

Howard, P. N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2798311>

Howard, P. N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2798311>

Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81–93. <https://doi.org/10.1080/19331681.2018.1448735>

Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81–93. <https://doi.org/10.1080/19331681.2018.1448735>

<https://ec.europa.eu/digital-single-market/en/news/ethicsguidelines-trustworthy-ai>

Hu, H. H., Cui, W. T., Lin, J., & Qian, Y. J. (2014). ICTs, social connectivity, and collective action: A cultural-political perspective. *Journal of Artificial Societies and Social Simulation*, 17(2), 7. DOI: 10.18564/jasss.2486

Hu, K. (2023). ChatGPT sets record for fastest-growing user base-analyst note. *Reuters*, 12, 2023. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>

Huang, M.-H., & Rust, R. T. (2020). Engaged to a Robot? The Role of AI in Service. *Journal of Service Research*, 24(1), 30–41. <https://doi.org/10.1177/1094670520902266>

Hunt, W., Sarkar, S., & Warhurst, C. (2022). Measuring the impact of AI on jobs at the organization level: Lessons from a survey of UK business leaders. *Research Policy*, 51(2), 104425. <https://doi.org/10.1016/j.respol.2021.104425>

Hussain, M., & Mumtaz, S. (2014). E-waste: impacts, issues and management strategies. *Reviews on environmental health*, 29(1-2), 53-58. <https://doi.org/10.1515/reveh-2014-0016>

Hwang, M. I. (2019). Top management support and information systems implementation success: A meta-analytical replication. *International Journal of Information Technology and Management*, 18(4), 347-361. <https://doi.org/10.1504/IJITM.2019.103050>

Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Review Press. <https://hbr.org/2020/01/competing-in-the-age-of-ai>

IDC (2023, March 7). *Worldwide spending on AI-centric systems forecast to reach \$154 billion in 2023, according to IDC*. <https://www.idc.com/getdoc.jsp?containerId=prUS50454123>

ideo.org. (2015). *The field guide to human-centered design*. Retrieved from: <http://www.designkit.org/resources/1>

IEEE. (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. http://standards.ieee.org/develop/indconn/ec/gieaisv2_guidelines.pdf

Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature Medicine*, 26(4), 463-464. <https://doi.org/10.1038/s41591-020-0832-5>

Involve, & DeepMind. (2019). *How to stimulate effective public engagement on the ethics of Artificial Intelligence*. Available at: <https://www.involve.org.uk/sites/default/files/field/attachemnt/How%20to%20stimulate%20effective%20public%20debate%20on%20the%20ethics%20of%20artificial%20intelligence%20.pdf>

Irlandoust, M. (2016). The renewable energy-growth nexus with carbon emissions and technological innovation: Evidence from the Nordic countries. *Ecological indicators*, 69, 118-125. <https://doi.org/10.1016/j.ecolind.2016.03.051>

ISO 31000. (2009). Risk management-Principles and guidelines. *International Organization for Standardization*, Geneva

Italian government. (2018). Proposte per una Strategia italiana per l'intelligenza artificiale. Ministero dello sviluppo economico. <https://libro-bianco-ia.readthedocs.io/en/latest/>

Jabbar, H., & Khan, R. Z. (2014). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 70, 978–981. http://doi.org/10.3850/978-981-09-5247-1_017

Jacob Leal, S., Napoletano, M., Roventini, A., & Fagiolo, G. (2016). Rock around the clock: An agent-based model of low-and high-frequency trading. *Journal of Evolutionary Economics*, 26, 49-76. <https://doi.org/10.1007/s00191-015-0418-4>

James, D. D. (1987). The economics of technological progress: a comparison of non-institutionalist and institutionalist dissent from the neoclassical position. *Journal of Economic Issues*, 21(2), 733-741. <https://doi.org/10.1080/00213624.1987.11504665>

Janjeva, A., Harris., A., Mercer, S., Kasprzyk, A., & Gausen, A. (2023). The Rapid Rise of GAI: Assessing risks to safety and security. The Alan Turing Institute. <https://cetas.turing.ac.uk/publications/rapid-rise-generative-ai>

Jarosz, L. (2012). Growing inequality: Agricultural revolutions and the political ecology of rural development. *International Journal of Agricultural Sustainability*, 10(2), 192–199. <https://doi.org/10.1080/14735903.2011.600605>

Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*, 2, 1-12. <https://doi.org/10.1016/j.aiia.2019.05.004>

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

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

- John-Mathews, J. M., Cardon, D., & Balagué, C. (2022). From reality to world: A critical perspective on AI fairness. *Journal of Business Ethics*, 945-959. <https://doi.org/10.1007/s10551-022-05055-8>.
- Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics*, 127, 707–715. <https://doi.org/10.1007/s10551-014-2180-1>
- Jonsen, A. R., & Toulmin, S. E. (1988). *The abuse of casuistry: A history of moral reasoning*. Univ of California Press.
- Jotabá, M. N., Fernandes, C. I., Gunkel, M., & Kraus, S. (2022). Innovation and human resource management: A systematic literature review. *European Journal of Innovation Management*. <https://doi.org/10.1108/EJIM-07-2021-0330>.
- Jung, M., Eom, M. T., & Yayla, A. (2022). Investigating the Interaction Effect between IT-enabled Innovation and Corporate Social Responsibility on Firm Performance. *Information Systems Management*, 1-20. <https://doi.org/10.1080/10580530.2022.2031352>
- Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal Of Operations & Production Management*, 37(1), 10-36. <https://doi.org/10.1108/IJOPM-02-2015-0078>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Kamiya, G. (2020). The carbon footprint of streaming video: fact-checking the headlines. *IEA: International Energy Agency*
- Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2023). The GAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. *Review of Managerial Science*, 1-32. <https://doi.org/10.1007/s11846-023-00696-z>
- Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2023). The GAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. *Review of Managerial Science*, 1-32. <https://doi.org/10.1007/s11846-023-00696-z>
- Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). Strategy, not technology, drives digital transformation. *MIT Sloan Management Review*. <http://sloanreview.mit.edu/projects/strategy-drives-digital-transformation/>
- Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*, 9(1), 36–41. https://ssir.org/articles/entry/collective_impact
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>

- Kaplan, S., & Garrick, B. J. (1981). On The Quantitative Definition of Risk. *Risk Analysis*, 1(1), 11–27. <https://doi.org/10.1111/j.1539-6924.1981.tb01350.x>
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4401-4410).
- Kassens-Noor, E., Wilson, M., Kotval-Karamchandani, Z., Cai, M., & Decaminada, T. (2021). Living with autonomy: Public perceptions of an AI-mediated future. *Journal of Planning Education and Research*, 0739456X20984529.
- Kastner, J. K., & Hong, S. J. (1984). A review of expert systems. *European journal of operational research*, 18(3), 285-292. [https://doi.org/10.1016/0377-2217\(84\)90150-4](https://doi.org/10.1016/0377-2217(84)90150-4)
- Katila, R. (2000). Using patent data to measure innovation performance. *International Journal of Business Performance Management*, 2(1-3), 180-193. <https://doi.org/10.1504/IJBPM.2000.000072>
- Kauffman, M. E., & Soares, M. N. (2020). AI in legal services: new trends in AI-enabled legal services. *Service Oriented Computing and Applications*, 14(4), 223-226. <https://doi.org/10.1007/s11761-020-00305-x>
- Kay, L. E. (2001). From Logical Neurons to Poetic Embodiments of Mind: Warren S. McCulloch's Project in Neuroscience. *Science in Context*, 14(4), 591–614. <https://doi.org/10.1017/S0269889701000266>
- Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9), 100314. <https://doi.org/10.1016/j.patter.2021.100314>
- Kazim, F. A. (2021). Digital transformation in communities of Africa. *International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)*, 11(1), 1–23. <http://doi.org/10.4018/IJDSGBT.287100>
- Kearns, G. S., & Sabherwal, R. (2006). Strategic alignment between business and information technology: A knowledge-based view of behaviors, outcome, and consequences. *Journal of Management Information Systems*, 23(3), 129–162. <https://doi.org/10.2753/MIS0742-1222230306>
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, 17, 1-9. <https://doi.org/10.1186/s12916-019-1426-2>
- Kend, M., & Nguyen, L. A. (2020). Big Data Analytics and Other Emerging Technologies: The Impact on the Australian Audit and Assurance Profession. *Australian Accounting Review*, 30(4), 269–282. <https://doi.org/10.1111/auar.12305>
- Khan, S. (2021). University of Illinois to Discontinue Remote-Testing Software After Students Complain of Privacy Violation. Available at: <https://www.techtimes.com/articles/256488/20210129/university-illinois-discontinue-remote-testing-software-students-complain-privacy-violation.htm>

Khan, S. (2021). University of Illinois to Discontinue Remote-Testing Software After Students Complain of Privacy Violation. *Techtimes*. <https://www.techtimes.com/articles/256488/20210129/university-illinois-discontinue-remote-testing-software-students-complain-privacy-violation.htm>

Khan, S. A. R., Piprani, A. Z., & Yu, Z. (2023). The decision-making of internet recycler considering Internet-of-Things application. *International Journal of Retail & Distribution Management*. <https://doi.org/10.1108/IJRDM-03-2023-0177>

Khan, S. A. R., Razzaq, A., Yu, Z., & Miller, S. (2021). Retracted: Industry 4.0 and circular economy practices: A new era business strategies for environmental sustainability. *Business Strategy and the Environment*, 30(8), 4001-4014. <https://doi.org/10.1002/bse.2853>

Khan, S. A., Lederer, A. L., & Mirchandani, D. A. (2013). Top management support, collective mindfulness, and information systems performance. *Journal of International Technology and Information Management*, 22(1), 6. <https://doi.org/10.58729/1941-6679.1005>

Khan, W. Z., Rehman, M. H., Zangoti, H. M., Afzal, M. K., Armi, N., & Salah, K. (2020). Industrial internet of things: Recent advances, enabling technologies and open challenges. *Computers & electrical engineering*, 81, 106522. <https://doi.org/10.1016/j.compeleceng.2019.106522>

Khasreen, M. M., Banfill, P. F., & Menzies, G. F. (2009). Life-cycle assessment and the environmental impact of buildings: a review. *Sustainability*, 1(3), 674-701. <https://doi.org/10.3390/su1030674>

Khattak, S. I., Ahmad, M., Khan, Z. U., & Khan, A. (2020). Exploring the impact of innovation, renewable energy consumption, and income on CO2 emissions: new evidence from the BRICS economies. *Environmental Science and Pollution Research*, 27(12), 13866-13881.

Khattak, S. I., Ahmad, M., Khan, Z. U., & Khan, A. (2020). Exploring the impact of innovation, renewable energy consumption, and income on CO2 emissions: new evidence from the BRICS economies. *Environmental Science and Pollution Research*, 27(12), 13866-13881. <https://doi.org/10.1007/s11356-020-07876-4>

Khattak, S. I., Ahmad, M., Khan, Z. U., & Khan, A. (2020). Exploring the impact of innovation, renewable energy consumption, and income on CO2 emissions: new evidence from the BRICS economies. *Environmental Science and Pollution Research*, 27(12), 13866-13881. <https://doi.org/10.1007/s11356-020-07876-4>

Khther, R. A., & Othman, M. (2013). Cobit framework as a guideline of effective IT governance in higher education: A review. *International Journal of Information Technology Convergence and Services*, 3(1), 21–29. <http://doi.org/10.5121/ijitcs.2013.3102>

Kidder, R. M. (2003). How good people make tough choices: Resolving the dilemmas of everyday living.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135–146. <https://doi.org/10.1016/j.bushor.2019.11.006>

- Kim, J., & Canny, J. (2017). Interpretable learning for self-driving cars by visualizing causal attention. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2942-2950). <https://doi.org/10.1109/ICCV.2017.320>
- Kline, R. (2010). Cybernetics, automata studies, and the Dartmouth conference on artificial intelligence. *IEEE Annals of the History of Computing*, 33(4), 5-16. <https://doi.org/10.1109/MAHC.2010.44>
- Klinke, A., & Renn, O. (2002). A new approach to risk evaluation and management: Risk-based, precaution-based, and discourse-based strategies 1. *Risk Analysis: An International Journal*, 22(6), 1071-1094. <https://doi.org/10.1111/1539-6924.00274>
- Knight, R. (2015). Convincing skeptical employees to adopt new technology. *Harvard Business Review*, 2, 1-7. Available at: <https://hbr.org/2015/03/convincing-skeptical-employees-to-adopt-new-technology>
- Koch, S., & Altinkemer, K. (2021). A value framework for technology potentials: Business adoption of emotion detection capabilities. *International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)*, 11(1), 1–13. <http://doi.org/10.4018/IJDSGBT.302636>
- Koebler, J. (2020). Detroit Police Chief: Facial Recognition Software Misidentifies 96% of the Time. *Vice*. <https://www.vice.com/en/article/dyzykz/detroit-police-chief-facial-recognition-software-misidentifies-96-of-the-time>
- Kohli, C., Suri, R., & Kapoor, A. (2015). Will social media kill branding?. *Business horizons*, 58(1), 35-44. <https://doi.org/10.1016/j.bushor.2014.08.004>
- Kohli, R., & Melville, N. P. (2019). Digital innovation: A review and synthesis. *Information Systems Journal*, 29(1), 200-223. <https://doi.org/10.1111/isj.12193>
- Korinek, A., & Stiglitz, J. E. (2021). *Artificial intelligence, globalization, and strategies for economic development* (No. w28453). National Bureau of Economic Research. DOI 10.3386/w28453
- Kosta, E. (2022). Algorithmic state surveillance: Challenging the notion of agency in human rights. *Regulation & Governance*, 16(1), 212–224. <https://doi.org/10.1111/rego.12331>
- Kosta, E. (2022). Algorithmic state surveillance: Challenging the notion of agency in human rights. *Regulation & Governance*, 16(1), 212–224. <https://doi.org/10.1111/rego.12331>
- Kou, G., Shi, Y., & Wang, S. (2011). Multiple criteria decision making and decision support systems—guest editor's introduction. *Decision Support Systems*, 51(2), 247–249. <https://doi.org/10.1016/j.dss.2010.11.027>
- Kraus, S., Breier, M., Lim, W. M., Dabić, M., Kumar, S., Kanbach, D., ... & Ferreira, J. J. (2022). Literature reviews as independent studies: guidelines for academic practice. *Review of Managerial Science*, 16(8), 2577-2595. <http://dx.doi.org/10.1007/s11846-022-00588-8>

- Krauss, S. E., Hamzah, A., Omar, Z., Suandi, T., Ismail, I. A., Zahari, M. Z., & Nor, Z. M. (2009). Preliminary investigation and interview guide development for studying how Malaysian farmers' form their mental models of farming. *The Qualitative Report*, 14(2), 245. <http://www.nova.edu/ssss/QR/QR14-2/krauss.pdf>
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being?. *American psychologist*, 53(9), 1017.
- Kulkarni, N. D., & Bansal, S. (2022). Revolutionizing Manufacturing: The Integral Role of AI and Computer Vision in Shaping Future Industries. *International Journal of Science and Research (IJSR)*, (13)1, 1183- 1188. <https://dx.doi.org/10.21275/SR24118231838>
- Kumamoto, H., & Henley, E. J. (2000). Probabilistic risk assessment and management for engineers and scientists. IEEE Press, ISBN-13: 978-0780360174
- Kumar, A., Alaraj, M., Rizwan, M., & Nangia, U. (2021). Novel AI based energy management system for smart grid with RES integration. *IEEE Access*, 9, 162530–162542. <https://doi.org/10.1109/ACCESS.2021.3131502>
- Kumar, V., & Grag, L. M. (2018). Predictive Analytics: A Review of Trends and Techniques. *International Journal of Computer Applications*, 182(1), 31–37. <https://doi.org/10.5120/ijca2018917434>
- Kuzey, C., Uyar, A., & Delen, D. (2014). The impact of multinationality on firm value: A comparative analysis of machine learning techniques. *Decision Support Systems*, 59, 127-142. <https://doi.org/10.1016/j.dss.2013.11.001>
- Kwee, T. C., & Kwee, R. M. (2021). Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. *Insights into imaging*, 12, 1-12. <https://doi.org/10.1186/s13244-021-01031-4>
- Lacina, L. (2021). Trust in climate science is strong, but optimism about progress is limited: Global survey. *World Economic Forum*. Available at: <https://www.weforum.org/press/2021/11/trust-in-climate-science-is-strong-but-optimism-about-progress-is-limited-global-survey/>
- Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). *Quantifying the Carbon Emissions of Machine Learning* (arXiv:1910.09700). arXiv. <http://arxiv.org/abs/1910.09700>
- Lair, B. (2020). The Risks of Autonomous Weapons Systems for Crisis Stability and Conflict Escalation in Future U.S.-Russia Confrontations. <https://www.rand.org/blog/2020/06/the-risks-of-autonomous-weapons-systems-for-crisis.html>
- Langkau, S., & Espinoza, L. A. T. (2018). Technological change and metal demand over time: What can we learn from the past?. *Sustainable Materials and Technologies*, 16, 54-59. <https://doi.org/10.1016/j.susmat.2018.02.001>
- Lannelongue, L., Grealey, J., & Inouye, M. (2021). Green Algorithms: Quantifying the Carbon Footprint of Computation. *Advanced Science*, 8(12), 2100707. <https://doi.org/10.1002/advs.202100707>

- Laperche, B., Lefebvre, G., & Langlet, D. (2011). Innovation strategies of industrial groups in the global crisis: Rationalization and new paths, *Technological Forecasting and Social Change*, 78(8), 1319-1331. <http://dx.doi.org/10.1016/j.techfore.2011.03.005>
- Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*, 7(3), 437–451. <https://doi.org/10.1017/als.2020.19>
- Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2). <https://doi.org/10.14763/2020.2.1469>
- Laud, R. L., & Schepers, D. H. (2009). Beyond Transparency: Information Overload and a Model for Intelligibility. *Business and Society Review*, 114(3), 365–391. <https://doi.org/10.1111/j.1467-8594.2009.00347.x>
- Lautenschläger, K. (1985). Controlling military technology. *Ethics*, 95(3), 692-711. <https://www.jstor.org/stable/2381045>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lazona, K. (2019). The new Republic. <https://newrepublic.com/authors/kevin-lozano>
- Lean, T. (2013). “Inside a Day You Will Be Talking to It Like an Old Friend”: The Making and Remaking of Sinclair Personal Computing in 1980s Britain. In *Hacking Europe: From Computer Cultures to Demoscenes* (pp. 49-71). London: Springer London.
- Lecocq, X., Warnier, V., Demil, B., & Plé, L. (2024). Using Artificial Intelligence (AI) Generative Technologies For Business Model Design with IDEATe Process: A Speculative Viewpoint. *Journal of Business Models*, 12(1), 21-35. <https://doi.org/10.54337/jbm.v12i1.8402>
- Lecocq, X., Warnier, V., Demil, B., & Plé, L. (2024). Using Artificial Intelligence (AI) Generative Technologies For Business Model Design with IDEATe Process: A Speculative Viewpoint. *Journal of Business Models*, 12(1), 21-35. <https://doi.org/10.54337/jbm.v12i1.8402>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, 18(1), 271. <https://doi.org/10.3390/ijerph18010271>
- Lee, J. K. (2022). Mission-based strategic management model for national R&D programs in Republic of Korea. *International Journal of Technology Management*, 89(3-4), 253-269. <https://doi.org/10.1504/IJTM.2022.123739>

- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. <https://doi.org/10.1177/2053951718756684>
- Legg, S., & Hutter, M. (2007). *Universal Intelligence: A Definition of Machine Intelligence* arXiv. <http://arxiv.org/abs/0712.3329>
- Lehmann, A. (2005). Innovations in Information and Communication Technologies: Benefits and Threats. In *International Seminar On Nuclear War And Planetary Emergencies—32nd Session* (pp. 143-152).
- Leiting, T., Külschbach, A., & Stich, V. (2023). Development of a platform business model for co-creation ecosystems for sustainable furniture. *Journal of Innovation Economics & Management*, 40(1), 81-107. <https://doi.org/10.3917/jie.pr1.0129>.
- Lettice, F., Pawar, P., & Rogers, H. (2013). Responsible innovation: What challenges does it pose for the new product development process. In *19th International Concurrent Enterprising (ICE) Conference*.
- Levitch, J. (2024). Build a More Sustainable Data Center for the Future. Available at: <https://blog.purestorage.com/perspectives/data-growth-and-environmental-sustainability/>
- Li, J., & Hou, L. (2019). A reflection on the taxi reform in China: Innovation vs. Tradition. *Computer Law & Security Review*, 35(3), 251-262. <https://doi.org/10.1016/j.clsr.2019.02.005>
- Li, X., Han, H., Evangelou, N., Wichrowski, N. J., Lu, P., Xu, W., ... & Tsapatsis, M. (2023). Machine learning-assisted crystal engineering of a zeolite. *Nature communications*, 14(1), 3152. <https://doi.org/10.1038/s41467-023-38738-5>
- Li, X., Wang, X., & Zhang, S. (2022). Impacts of Urban Spatial Development Patterns on Carbon Emissions: Evidence from Chinese Cities. *Land*, 11(11), 2031. <https://doi.org/10.3390/land11112031>
- Lian, J., Freeman, L., Hong, Y., & Deng, X. (2021). Robustness with respect to class imbalance in artificial intelligence classification algorithms. *Journal of Quality Technology*, 53(5), 505–525. <https://doi.org/10.1080/00224065.2021.1963200>
- Lighthill, J. (1973). ‘Artificial Intelligence: A general survey’ in Artificial Intelligence: A paper symposium. Retrieved from UK Science Research Council website: http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm.
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching*, 6(1). <http://dx.doi.org/10.37074/jalt.2023.6.1.32>

- Lin, Y., Hu, Q., Liu, J., & Duan, J. (2015). Multi-label feature selection based on max-dependency and min-redundancy. *Neurocomputing*, 168, 92–103. <https://doi.org/10.1016/j.neucom.2015.06.010>
- Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., ... & Tang, J. (2022). Trustworthy AI: A computational perspective. *ACM Transactions on Intelligent Systems and Technology*, 14(1), 1-59.
- Liu, Q., & Lawell, C. Y. C. L. (2015). The effects of innovation on income inequality in China. *Shandong Province Educational Department*.
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021). Corporate digital responsibility. *Journal of Business Research*, 122, 875-888. <https://doi.org/10.1016/j.jbusres.2019.10.006>
- Lohmann, R., Breivik, K., Dachs, J., & Muir, D. (2007). Global fate of POPs: current and future research directions. *Environmental pollution*, 150(1), 150-165. <https://doi.org/10.1016/j.envpol.2007.06.051>
- Lokshina, I., & Lanting, C. (2020). An attempt to model decision processes for time-constrained applications. In *Proceedings of the Science Fiction Prototyping Conference* (pp. 27-32).
- Lokshina, I., Kniezova, J., & Lanting, C. (2022). On Building Users' Initial Trust in Autonomous Vehicles. *Procedia Computer Science*, 198, 7-14. <https://doi.org/10.1016/j.procs.2021.12.205>
- Long, T. B., Blok, V., Dorrestijn, S., & Macnaghten, P. (2020). The design and testing of a tool for developing responsible innovation in start-up enterprises. *Journal of Responsible Innovation*, 7(1), 45-75. <https://doi.org/10.1080/23299460.2019.1608785>
- López Jiménez, D., Dittmar, E. C., & Vargas Portillo, J. P. (2021). New directions in corporate social responsibility and ethics: Codes of conduct in the digital environment. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-021-04753-z>
- Lu, V. N., Wirtz, J., Kunz, W. H., Paluch, S., Gruber, T., Martins, A., & Patterson, P. G. (2020). Service Robots, Customers and Service Employees: What Can We Learn from the Academic Literature and Where Are the Gaps?. *Journal of Service Theory and Practice*, 30(3), 361-391. <https://doi.org/10.1108/JSTP-04-2019-0088>
- Lubberink, R., Blok, V., Van Ophem, J., and Omta, O. (2017). Lessons for responsible innovation in the business context: A systematic literature review of responsible, social and sustainable innovation practices. *Sustainability*, 9(5), 721. <https://doi.org/10.3390/su9050721>
- Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6), 2824–2838. <https://doi.org/10.1111/bjet.12861>

- Luger, G. F. (2021). Symbol-Based AI and Its Rationalist Presuppositions. *Knowing our World: An Artificial Intelligence Perspective*, 77-104. https://doi.org/10.1007/978-3-030-71873-2_4
- Luzzati, T., Parenti, A., & Rughi, T. (2018). Economic growth and cancer incidence. *Ecological Economics*, 146, 381-396. <https://doi.org/10.1016/j.ecolecon.2017.11.031>
- Lysyakov, M., & Viswanathan, S. (2023). Threatened by AI: Analyzing Users' Responses to the Introduction of AI in a Crowd-sourcing Platform. *Information Systems Research*, 34(3), 1191-1210. <https://doi.org/10.1287/isre.2022.1184>
- Lythreathis, Sophie, Sanjay Kumar Singh, and Abdul-Nasser El-Kassar. "The digital divide: A review and future research agenda." *Technological Forecasting and Social Change* 175 (2022): 121359. <https://doi.org/10.1016/j.techfore.2021.121359>
- M. U., E. (2023). The beneficial and harmful aspects of artificial intelligence in the educational process. *Horizon: Journal of Humanity and Artificial Intelligence*, 2(5), 603–606. <https://univerpubl.com/index.php/horizon/article/view/1774>
- Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International Journal of Research in Marketing*, 37(3), 481–504. <https://doi.org/10.1016/j.ijresmar.2020.04.005>
- Maarouf, H. (2019). Pragmatism as a supportive paradigm for the mixed research approach: Conceptualizing the ontological, epistemological, and axiological stances of pragmatism. *International Business Research*, 12(9), 1-12.
- Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of experimental social psychology*, 19(5), 469-479. [https://doi.org/10.1016/0022-1031\(83\)90023-9](https://doi.org/10.1016/0022-1031(83)90023-9)
- Madsen, E. B., Nielsen, M. W., Bjørnholm, J., Jagsi, R., & Andersen, J. P. (2022). Author-level data confirm the widening gender gap in publishing rates during COVID-19. *Elife*, 11, e76559. <https://doi.org/10.7554/eLife.76559>
- Maican, C., & Lixandriou, R. (2016). A system architecture based on open source enterprise content management systems for supporting educational institutions. *International Journal of Information Management*, 36(2), 207-214. <https://doi.org/10.1016/j.ijinfomgt.2015.11.003>
- Majeed, A. (2023). Attribute-Centric and Synthetic Data Based Privacy Preserving Methods: A Systematic Review. *Journal of Cybersecurity and Privacy*, 3(3), 638-661. <https://doi.org/10.3390/jcp3030030>
- Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262–273. <https://dx.doi.org/10.1016/j.jbusres.2020.07.045>
- Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, 120, 262–273. <https://doi.org/10.1016/j.jbusres.2020.07.045>

- Man, K., & Chahl, J. (2022). A review of synthetic image data and its use in computer vision. *Journal of Imaging*, 8(11), 310. <https://doi.org/10.3390/jimaging8110310>
- Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale JL & Tech.*, 21, 106. <https://ssrn.com/abstract=3273016>
- Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution & Logistics Management*, 38(3), 192–223. <https://doi.org/10.1108/09600030810866986>
- Manyika, J., Lund, S., Chui, M., Bughin, B., Woetzel, L., Batra, P., Ko, R., & Sanghvi, S. (2017). Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages. Available at: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>
- Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255-262. <https://doi.org/10.1177/1365712718807226>
- Mariani, R., Rossi, F., Cucchiara, R., Pavone, M., Simkin, B., Koene, A., & Papenbrock, J. (2023). Trustworthy AI—Part 1. *Computer*, 56(2), 14-18. <https://doi.org/10.1109/MC.2022.3227683>
- Marín-Beltrán, I., Demaria, F., Ofelio, C., Serra, L. M., Turiel, A., Ripple, W. J., Mukul, S. A., & Costa, M. C. (2022). Scientists' warning against the society of waste. *Science of The Total Environment*, 811, 151359. <https://doi.org/10.1016/j.scitotenv.2021.151359>
- Marketandmarkets. (2023). Artificial Intelligence (AI) Market by offering (hardware software), Technology (ML(Deep Learnin (LLM, Transformers (GPT 1, 2, 3, 4), NLP, comuter visison), business function, vertical, and regional – global forecast to 2030. <https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-market-74851580.html>
- Marsoof, A., Luco, A., Tan, H., & Joty, S. (2023). Content-filtering AI systems—limitations, challenges and regulatory approaches. *Information & Communications Technology Law*, 32(1), 64-101. <https://doi.org/10.1080/13600834.2022.2078395>
- Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984-986. <https://doi.org/10.1126/science.aba3758>
- Masih, J., & Rajkumar, R. (2019). Internet addiction disorder and mental health in adolescents. *Depress Anxiety*, 13(002). DOI: 10.4172/2167-1044.S13-002
- Mawson, C. O. Sylvester (1975). "sui generis". *Dictionary of Foreign Terms* (2 ed.). New York: Thomas Y. Crowell Company.
- Mayson, S. G. (2018). Bias in, bias out. *YAle IJ*, 128, 2218. <https://heinonline.org/HOL/LandingPage?handle=hein.journals/ylr128&div=47&id=&page=>

McCurry, J. (2021). South Korean AI chatbot pulled from Facebook after hate speech towards minorities. *The Guardian*. <https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook>

McFadden, D. (1978). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272.

McIntosh, T. R., Liu, T., Susnjak, T., Watters, P., Ng, A., & Halgamuge, M. N. (2023). A culturally sensitive test to evaluate nuanced gpt hallucination. *IEEE Transactions on Artificial Intelligence*, 1-13. <http://dx.doi.org/10.1109/TAI.2023.3332837>

McKinsey. (2017). A future that works: automation, employment, and productivity. Available at: <https://www.mckinsey.com/~media/mckinsey/featured%20insights/digital%20disruption/harnessing%20automation%20for%20a%20future%20that%20works/a-future-that-works-full-report-mgi-january-2017.ashx>

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35. <http://arxiv.org/abs/1908.09635>

Meijer, A. (2014). Transparency. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp. 507–524). Oxford: Oxford University Press.

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, 39(1), 53–63. <https://doi.org/10.1080/10580530.2020.1849465>

Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine*, 6(1), 120.

Meta. (2021). Facebook’s five pillars of Responsible AI. <https://ai.meta.com/blog/facebook-five-pillars-of-responsible-ai/>

Meurier, A.-L., Ghafoor, Z., Foehrenbach, C., Hartmann, C., Herzog, J., Madzou, L., & Narzul, R. (2018). *Mission assigned by the Prime Minister Édouard Philippe*.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434. <https://doi.org/10.1016/j.im.2021.103434>

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and “the dark side” of AI. *European Journal of Information Systems*, 31(3), 257–268. <https://doi.org/10.1080/0960085X.2022.2026621>

Mill, J. S. (2016). Utilitarianism. In *Seven masterpieces of philosophy* (pp. 329-375). Routledge.

Ministry of Foreign Affairs of Japan. (2019). *G20 AI Principles*. <https://oecd.ai/en/work/documents/g20-ai-principles>

Ministry of Science, Technology, and Innovations. (2021). Summary of the Brazilian Artificial Intelligence Strategy -EBIA-. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-summary_brazilian_4-979_2021.pdf

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501–507. <https://doi.org/10.1038/s42256-019-0114-4>

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. https://doi.org/10.1177/205395171667967_9

Mkrttchian, V., & Voronin, V. (2021). Digitalization of lifecycle management of domestic Russian tour products based on problem-oriented digital twins-avatars, supply chain, 3D-hybrid, federated, and coordinated blockchain. *International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)*, 11(1), 1–13. <http://doi.org/10.4018/IJDSGBT.20210101.0a2>

Moen, K., Aggleton, P., Leshabari, M. T., & Middelthon, A. L. (2012). Not at all so hard-to-reach: same-sex attracted men in Dar es Salaam. *Culture, Health & Sexuality*, 14(2), 195-208. <https://doi.org/10.1080/13691058.2011.632825>

Moir, L. (2001). What do we mean by corporate social responsibility?. *Corporate Governance: The international journal of business in society*, 1(2), 16-22. <https://doi.org/10.1108/EUM0000000005486>

Molęda, M., Małysiak-Mrozek, B., Ding, W., Sunderam, V., & Mrozek, D. (2023). From corrective to predictive maintenance—A review of maintenance approaches for the power industry. *Sensors*, 23(13), 5970. <https://doi.org/10.3390/s23135970>

Mollemans, M. (2022). Algorithmic Trading Solutions, 2022: Market and Vendor Landscape. <https://www.chartis-research.com/market-structure/7946246/algorithmic-trading-solutions-2022-market-and-vendor-landscape>

Montiel, I., & Delgado-Ceballos, J. (2014). Defining and measuring corporate sustainability: Are we there yet?. *Organization & Environment*, 27(2), 113-139. <https://doi.org/10.1177/1086026614526413>

Morelli, J. (2011). Environmental Sustainability: A Definition for Environmental Professionals. *Journal of Environmental Sustainability*, 1(1), 1–10. <https://doi.org/10.14448/jes.01.0002>

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and engineering ethics*, 26(4), 2141-2168. <https://doi.org/10.1007/s11948-019-00165-5>

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: barriers, enablers and next steps. *AI & Society*, 1-13. <https://doi.org/10.1007/s00146-021-01308-8>

Morris, M. R., & Brubaker, J. R. (2024). Generative ghosts: Anticipating benefits and risks of AI afterlives. <https://doi.org/10.48550/arXiv.2402.01662>

Mota, Y. (2024). Generative AI in healthcare: Use cases and challenges. *N-iX*.
<https://www.n-ix.com/generative-ai-in-healthcare/>

Mueller, E. T. (2016). *Transparent Computers: Designing Understandable Intelligent Systems*. Erik T. Mueller, San Bernardino, CA.

Müller, C. (2013). Non-Muslims as part of Islamic law: Juridical casuistry in a fifth/eleventh-century law manual. In M. Fierro & J. Tolan (Eds.), *The legal status of ḍimmī-s in the Islamic West (second/eighth-ninth/fifteenth centuries)* (Vol. 1, pp. 21–63). Brepols Publishers. <https://doi.org/10.1484/M.RELMIN-EB.1.101811>

Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 555–572). Springer. https://doi.org/10.1007/978-3-319-26485-1_33

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics*, 167, 209–234. <https://doi.org/10.1007/s10551-019-04407-1>

Murphy, M.H. (2017). Algorithmic Surveillance: The Collection Conundrum. *International Review of Law, Computers & Technology*, 31 (2), 225–242. <http://dx.doi.org/10.1080/13600869.2017.1298497>

Murphy, M.H. (2017). Algorithmic Surveillance: The Collection Conundrum. *International Review of Law, Computers & Technology*, 31 (2), 225–242. <http://dx.doi.org/10.1080/13600869.2017.1298497>

Murphy, M.H. (2017). Algorithmic Surveillance: The Collection Conundrum. *International Review of Law, Computers & Technology*, 31 (2), 225–242. <http://dx.doi.org/10.1080/13600869.2017.1298497>

Murrja, A., Ndrejoni, A., Troka, P., & Shehu, E. (2019). Explanatory and argumentative essay on the risk of entrepreneurship-the definition of business risk. *Business and Economics*, 89.

Murugavalli, S., Jainulabudeen, S. A. K., Kumar, G. S., & Anuradha, D. (2016). Enhancing security against hard AI problems in user authentication using CAPTCHA as graphical passwords. *International Journal of Advanced Computer Research*, 6(24), 93–99. <https://doi.org/10.19101/IJACR.2016.624001>

Mutasa, S., Sun, S., & Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting?. *Clinical imaging*, 65, 96–99. <https://doi.org/10.1016/j.clinimag.2020.04.025>

Naciti, V. (2019). Corporate governance and board of directors: The effect of a board composition on firm sustainability performance. *Journal of Cleaner Production*, 237, 117727. <https://doi.org/10.1016/j.jclepro.2019.117727>

Nadeem, A., Abedin, B., & Marjanovic, O. (2020). *Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies*.

Nahodil, P., & Vítků, J. (2013). How to design an autonomous creature based on original artificial life approaches. In *Beyond artificial intelligence: Contemplations, expectations, applications* (pp. 161-180). Berlin, Heidelberg: Springer Berlin Heidelberg.

NAIS. (2023). Singapore National AI Strategy AI for Public Good For Singapore and the World. <https://www.smartnation.gov.sg/nais/>

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management. *MIS Quarterly*, 41(1), 223-238. <https://www.jstor.org/stable/26629644>

Namugenyi, C., Nimmagadda, S. L., & Reiners, T. (2019). Design of a SWOT analysis model and its evaluation in diverse digital business ecosystem contexts. *Procedia Computer Science*, 159, 1145-1154.

Nathan, L. P. (2012). Sustainable information practice: An ethnographic investigation. *Journal of the American Society for Information Science and Technology*, 63(11), 2254-2268. <https://doi.org/10.1002/asi.22726>

National Nanotechnology Initiative. (2007) 'National Nanotechnology Initiative Strategic Plan', https://www.nano.gov/sites/default/files/pub_resource/nni_strategic_plan_2007.pdf

National Program for Artificial Intelligence. (2018). UAE National Strategy for AI. <https://ai.gov.ae/wp-content/uploads/2021/07/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf>

National Security Commission on Artificial Intelligence (NSCAI). (2021). The NSCAI Final Report presents an integrated national strategy to reorganize the government, reorient the nation, and rally our closest allies and partners to defend and compete in the coming era of AI-accelerated competition and conflict. <https://www.nsc.ai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>

National Security Commission on Artificial Intelligence. (2021). *The NSCAI Final Report presents an integrated national strategy to reorganize the government, reorient the nation, and rally our closest allies and partners to defend and compete in the coming era of AI-accelerated competition and conflict*. <https://www.nsc.ai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>

Navinchandran, M., Sharp, M. E., Brundage, M. P., & Sexton, T. B. (2022). Discovering critical KPI factors from natural language in maintenance work orders. *Journal of Intelligent Manufacturing*, 1-19. <https://doi.org/10.1007/s10845-021-01772-5>

Nazareno, L., & Schiff, D. S. (2021). The impact of automation and artificial intelligence on worker well-being. *Technology in Society*, 67, 101679. <https://doi.org/10.1016/j.techsoc.2021.101679>

Nelli, F., & Nelli, F. (2015). pandas in Depth: Data Manipulation. *Python Data Analytics: Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language*, 131-165.

Nepelski, D. (2019). How to facilitate digital innovation in Europe. *Intereconomics*, 54(1), 47-52. <https://doi.org/10.1007/s10272-019-0791-6>

Newell, A., & Shaw, J. C. (1957). Programming the logic theory machine. *Papers Presented at the February 26-28, 1957, Western Joint Computer Conference: Techniques for Reliability on - IRE-AIEE-ACM '57 (Western)*, 230–240. <https://doi.org/10.1145/1455567.1455606>

Newell, S., & Marabelli, M. (2020). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ‘datification’. In *Strategic Information Management* (pp. 430-449). Routledge.

Ngo, L. V., & O'cass, A. (2013). Innovation and business success: The mediating role of customer participation. *Journal of Business Research*, 66(8), 1134-1142. <https://doi.org/10.1016/j.jbusres.2012.03.009>

Niedderer, K., & Roworth-Stokes, S. (2007). The role and use of creative practice in research and its contribution to knowledge. In *IASDR International Conference* (Vol. 795).

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, 53, 102104. <https://doi.org/10.1016/j.ijinfomgt.2020.102104>

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, 53, 102104. <https://doi.org/10.1016/j.ijinfomgt.2020.102104>

Noack, A., Ahern, I., Dou, D., & Li, B. (2021). An empirical study on the relation between network interpretability and adversarial robustness. *SN Computer Science*, 2, 1-13. <https://doi.org/10.1007/s42979-020-00390-x>

Norris, T. B., & Suomela, T. (2017). Information in the ecosystem: Against the “information ecosystem”. *First Monday*, 22(9). <https://firstmonday.org/ojs/index.php/fm/article/download/6847/6530>

Norton, A. (2017). *Automation and inequality: the changing world of work in the global South*. International Institute for Environment and Development.

Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., & Floridi, L. (2023). Taking AI risks seriously: a new assessment model for the AI Act. *AI & SOCIETY*, 1-5. <https://doi.org/10.1007/s00146-023-01723-z>

Nyland, R. (2007). *The gratification niches of Internet social networking, e-mail, and face-to-face communication*. Brigham Young University.

Nylén, D., & Holmström, J. (2015). Digital innovation strategy: A framework for diagnosing and improving digital product and service innovation. *Business Horizons*, 58(1), 57–67. <https://doi.org/10.1016/j.bushor.2014.09.001>

O’Connell, N., & Binsanad, A. (2023). Saudi Data and Artificial Intelligence Authority Reveals AI Ethics Principles 2.0. <https://www.tamimi.com/news/ai-ethics-principles-version-2-0/>

O’Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M. I., & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15(1), e1968. <https://doi.org/10.1002/rcs.1968>

OECD. (2010). The OECD Innovation Strategy: Getting a Head Start on Tomorrow. Retrieved from: <https://www.oecd-ilibrary.org/docserver/9789264083479-sum-en.pdf?expires=1710637180&id=id&accname=guest&checksum=A3EFA3CDC708301750C215DA59EE93B1>

OECD. (2019). *OECD AI Principles overview*. <https://oecd.ai/en/ai-principles>

Ogie, R., Castilla Rho, J., Clarke, R. J., & Moore, A. (2018, October). Disaster risk communication in culturally and linguistically diverse communities: the role of technology. In *Proceedings* (Vol. 2, No. 19, p. 1256). MDPI. <https://doi.org/10.3390/proceedings2191256>

Oladoyinbo, T. O., Olabanji, S. O., Olaniyi, O. O., Adebiyi, O. O., Okunleye, O. J., & Ismaila Alao, A. (2024). Exploring the challenges of artificial intelligence in data integrity and its influence on social dynamics. *Asian Journal of Advanced Research and Reports*, 18(2), 1-23. <http://dx.doi.org/10.9734/AJARR/2024/v18i2601>

Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., & Liu, S. (2022). Fake news on social media: the impact on society. *Information Systems Frontiers*, 26(2), 443-458. <https://doi.org/10.1007/s10796-022-10242-z>

Olateju, D. J., Olateju, O. A., Adeoye, S. V., & Ilyas, I. S. (2021). A critical review of the application of the legitimacy theory to corporate social responsibility. *International Journal of Managerial Studies and Research (IJMSR)*, 9(3), 01–06. <https://doi.org/10.20431/2349-0349.0903001>

Oluwaseyi, J. (2024). AI in Finance: Applications, Risks, and Regulatory Considerations. *Artificial Satellites*.

Oniani, D., Hilsman, J., Peng, Y., Poropatich, R. K., Pamplin, J. C., Legault, G. L., & Wang, Y. (2023). Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. *NPJ Digital Medicine*, 6(1), 225. <https://doi.org/10.1038/s41746-023-00965-x>

Onik, M. H., Al-Zaben, N., Yang, J., & Kim, C.-S. (2018). Privacy of Things (PoT): Personally Identifiable Information Monitoring System for Smart Homes. In *conference Proceedings of Symposium of the Korean Institute of communications and Information Sciences (KICS)* (256-257).

Oravec, J. A. (2022). “Our Hearts Go Out to the Victim’s Family”: Death by Robot and Autonomous Vehicle. In *Good Robot, Bad Robot: Dark and Creepy Sides of Robotics, Autonomous Vehicles, and AI* (pp. 177-204). Cham: Springer International Publishing.

Orbik, Z., & Zozuláková, V. (2019). Corporate social and digital responsibility. *Management Systems in Production Engineering*, 27, 79-83. DOI 10.1515/mspe-2019-0013

Orzes, G., Moretto, A. M., Ebrahimpour, M., Sartor, M., Moro, M., & Rossi, M. (2018). United Nations Global Compact: Literature review and theory-based research agenda. *Journal of Cleaner Production*, 177, 633–654. <https://doi.org/10.1016/j.jclepro.2017.12.230>

Owczarczuk, M. (2023). Ethical and regulatory challenges amid artificial intelligence development: an outline of the issue. *Ekonomia i Prawo. Economics and Law*, 22(2 (Forthcoming)).

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. *Responsible innovation: managing the responsible emergence of science and innovation in society*, 27-50. <https://doi.org/10.1002/9781118551424.ch2>

Pagallo, U. (2017). The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection. *European Data Protection Law Review*, 3(1), 36-46. DOI: 10.21552/edpl/2017/1/7

Pagani, M. (2013). Digital business strategy and value creation: Framing the dynamic cycle of control points. *Mis Quarterly*, 617-632. <https://www.jstor.org/stable/43825925>

Pagani, M., & Pardo, C. (2017). The impact of digital technology on relationships in a business network. *Industrial Marketing Management*, 67, 185-192. <https://doi.org/10.1016/j.indmarman.2017.08.009>

Paneque-Gálvez, J., McCall, M., Napoletano, B., Wich, S., & Koh, L. (2014). Small Drones for Community-Based Forest Monitoring: An Assessment of Their Feasibility and Potential in Tropical Areas. *Forests*, 5(6), 1481–1507. <https://doi.org/10.3390/f5061481>

Panigrahi, S., Maski, P., & Thondiyath, A. (2021, December). Deep learning based real-time biodiversity analysis using aerial vehicles. In *International conference on robot intelligence technology and applications* (pp. 401-412). Cham: Springer International Publishing.

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers*, 25, 123–141. <https://doi.org/10.1007/s10796-022-10251-y>

Pappadà, A., Chattat, R., Valente, M., & Ottoboni, G. (2021). Assistive technologies in dementia care: an updated analysis of the literature. *Frontiers in psychology*, 12, 644587. <https://doi.org/10.3389/fpsyg.2021.644587>

Pappas, I. O., Mikalef, P., Dwivedi, Y. K., Jaccheri, L., & Krogstie, J. (2023). Responsible digital transformation for a sustainable society. *Information Systems Frontiers*, 25(3), 945-953. <https://doi.org/10.1007/s10796-023-10406-5>

- Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, 63(3), 403–414. <https://doi.org/10.1016/j.bushor.2020.01.003>
- Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2021). Carbon emissions and large neural network training. <https://doi.org/10.48550/arXiv.2104.10350>
- Paul, R., & Elder, L. (2006). *The thinker's guide to understanding the foundations of ethical reasoning: Based on critical thinking concepts & tools*.
- Peek, N., Combi, C., Marin, R., & Bellazzi, R. (2015). Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes. *Artificial intelligence in medicine*, 65(1), 61-73. <https://doi.org/10.1016/j.artmed.2015.07.003>
- Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI—A transformational shift for government: So, what next for research?. *Public Policy and Administration*, 35(1), 24-44. <http://dx.doi.org/10.1177/0952076718780537>
- Pereira, A. M., Moura, J. A. B., Costa, E. D. B., Vieira, T., Landim, A. R., Bazaki, E., & Wanick, V. (2022). Customer models for artificial intelligence-based decision support in fashion online retail supply chains. *Decision Support Systems*, 158, 113795. <https://doi.org/10.1016/j.dss.2022.113795>
- Peretto, P. F., & Seater, J. J. (2013). Factor-eliminating technical change. *Journal of Monetary Economics*, 60(4), 459-473. <https://doi.org/10.1016/j.jmoneco.2013.01.005>
- Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *European radiology experimental*, 2, 1-10. <https://doi.org/10.1186/s41747-018-0061-6>
- Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *European radiology experimental*, 2, 1-10. <http://dx.doi.org/10.1186/s41747-018-0061-6>
- Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for Motivation, Engagement and Wellbeing in Digital Experience. *Frontiers in Psychology*, 9, 797. <https://doi.org/10.3389/fpsyg.2018.00797>
- Peterson, R. (2004). Crafting information technology governance. *Information Systems Management*, 21(4), 7–22. <https://doi.org/10.1201/1078/44705.21.4.20040901/84183.2>
- Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic Information Systems. *Business & Information Systems Engineering*, 61(4), 523–533. <https://doi.org/10.1007/s12599-019-00599-y>
- Phadermrod, B., Crowder, R. M., & Wills, G. B. (2019). Importance-performance analysis based SWOT analysis. *International journal of information management*, 44, 194-203. DOI: 10.1002/9781118785317.weom120103

- Pichlak, M., & Szromek, A. R. (2021). Eco-Innovation, Sustainability and Business Model Innovation by Open Innovation Dynamics. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(2), 149. <https://doi.org/10.3390/joitmc7020149>
- Pierre, A., & Fernandez, A.-S. (2018). Going deeper into SMEs' innovation capacity: An empirical exploration of innovation capacity factors. *Journal of Innovation Economics & Management*, 25(1), 139. <https://doi.org/10.3917/jie.pr1.0019>.
- Pillemer, K., Wells, N. M., Wagenet, L. P., Meador, R. H., & Parise, J. T. (2011). Environmental sustainability in an aging society: a research agenda. *Journal of Aging and Health*, 23(3), 433-453. <https://doi.org/10.1177/0898264310381278>
- Plekhanov, D., Franke, H., & Netland, T. H. (2022). Digital transformation: A review and research agenda. *European Management Journal*. <https://doi.org/10.1016/j.emj.2022.09.007>
- Polit D.S. & Beck C.T. (2010). *Essentials of Nursing Research. Appraising Evidence for Nursing Practice*, 7th edn. Lippincott-Raven Publishers, Philadelphia
- Polykalas, S. E., & Prezerakos, G. N. (2019). When the mobile app is free, the product is your personal data. *Digital Policy, Regulation and Governance*, 21(2), 89–101. <https://doi.org/10.1108/DPRG-11-2018-0068>
- Poornima, B. (2022). Cyber threats and nuclear security in India. *Journal of Asian Security and International Affairs*, 9(2), 183-206. <https://doi.org/10.1177/23477970221099748>
- Porter, M. E., & Kramer, M. R. (2018). Creating shared value: How to reinvent capitalism—And unleash a wave of innovation and growth. In *Managing sustainable business: An executive education case and textbook* (pp. 323-346). Dordrecht: Springer Netherlands.
- Pradhan, A., Mehta, K., & Findlater, L. (2018). “Accessibility Came by Accident”: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3173574.3174033>
- Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1), 12. <https://doi.org/10.3390/publications9010012>
- Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00258-9>
- Prettner, K., & Schaefer, A. (2021). The U-shape of income inequality over the 20th century: The role of education. *The Scandinavian Journal of Economics*, 123(2), 645-675. <https://doi.org/10.1111/sjoe.12424>
- PricewaterhouseCoopers (PwC). (2018). Sizing the prize PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution What's the real value of AI for your business and how can you capitalise?. <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>

- Priyadarshini, I. (2018). Cyber security risks in robotics. In *Cyber security and threats: concepts, methodologies, tools, and applications* (pp. 1235-1250). IGI Global.
- Proctor, J. D. (1998). Ethics in geography: Giving moral form to the geographical imagination. *Area*, 30(1), 8–18. <https://doi.org/10.1111/j.1475-4762.1998.tb00043.x>
- Puaschunder, J. M. (2020). Revising growth theory in the artificial age: Putty and clay labor. *Archives in Business Research*, 8(3), 65–107. <https://dx.doi.org/10.2139/ssrn.3385238>
- PwC (2018). *Sizing the prize: PwC's global artificial intelligence study: Exploiting the AI revolution*. <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>
- Qiang, T., Gao, H., & Ma, X. (2021). Pro-environmental behavior and smartphone uses of on-campus engineering students in Xi'an, China. *Plos one*, 16(11), e0259542.
- Qiang, V., Rhim, J., & Moon, A. (2023). No such thing as one-size-fits-all in AI ethics frameworks: A comparative case study. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-023-01653-w>
- Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of management information systems*, 25(4), 145-182. <https://doi.org/10.2753/MIS0742-1222250405>
- Quintelier, K., Van Speybroeck, L., & Braeckman, J. (2011). Normative Ethics Does Not Need a Foundation: It Needs More Science. *Acta Biotheoretica*, 59(1), 29–51. <https://doi.org/10.1007/s10441-010-9096-7>
- Radu, C., & Smaili, N. (2022). Board Gender Diversity and Corporate Response to Cyber Risk: Evidence from Cybersecurity Related Disclosure. *Journal of Business Ethics*, 177(2), 351–374. <https://doi.org/10.1007/s10551-020-04717-9>
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of management review*, 46(1), 192-210. <https://doi.org/10.5465/amr.2018.0072>
- Raita, E., & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, 23(4), 363-371. <https://doi.org/10.1016/j.intcom.2011.04.002>
- Ramasasthy, A. (2015). Corporate social responsibility versus business and human rights: Bridging the gap between responsibility and accountability. *Journal of Human Rights*, 14(2), 237–259. <https://doi.org/10.1080/14754835.2015.1037953>
- Rampa, R., & Agogu  , M. (2021). Developing radical innovation capabilities: Exploring the effects of training employees for creativity and innovation. *Creativity and Innovation Management*, 30(1), 211-227. <https://doi.org/10.1111/caim.12423>.

Ransbotham, S., Kiron, D., Gerbert, P., and Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT Sloan Management Review*. <https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/>

Rashid, A. B., Kausik, A. K., Al Hassan Sunny, A., & Bappy, M. H. (2023). Artificial intelligence in the military: An overview of the capabilities, applications, and challenges. *International Journal of Intelligent Systems*. <https://doi.org/10.1155/2023/8676366>

Rasskazov, V. E. (2020). Financial and economic consequences of distribution of artificial intelligence as a general-purpose technology. *Finance Theory and Practice*, 24(2), 120-132.

Ray, P. P. (2023). Benchmarking, ethical alignment, and evaluation framework for conversational AI: Advancing responsible development of ChatGPT. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(3), 100136.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>

Re, R. M., & Solow-Niederman, A. (2019). Developing artificially intelligent justice. *Stan. Tech. L. Rev.*, 22, 242.

Republic of Türkiye Ministry of Industry and Technology. (2021). National artificial intelligence strategy 2021-2025. <https://cbddo.gov.tr/SharedFolderServer/Genel/File/TRNationalAIStrategy2021-2025.pdf>

Ribeiro, B. E., Smith, R. D., & Millar, K. (2017). A mobilising concept? Unpacking academic representations of responsible research and innovation. *Science and engineering ethics*, 23, 81-103. <https://doi.org/10.1007/s11948-016-9761-6>

Rich, E. (1985). Artificial intelligence and the humanities. *Computers and the Humanities*, 19(2), 117-122. <http://www.jstor.org/stable/30204398>

Richards, C. E., Lupton, R. C., & Allwood, J. M. (2021). Re-framing the threat of global warming: an empirical causal loop diagram of climate change, food insecurity and societal collapse. *Climatic Change*, 164(3), 49. <https://doi.org/10.1007/s10584-021-02957-w>

Riehm, K. E., Feder, K. A., Tormohlen, K. N., Crum, R. M., Young, A. S., Green, K. M., Pacek, L. R., La Flair, L. N., & Mojtabai, R. (2019). Associations between time spent using social media and internalizing and externalizing problems among US youth. *JAMA Psychiatry*, 76(12), 1266-1273. <https://doi.org/10.1001/jamapsychiatry.2019.2325>

Rigby, M. J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics*, 21(2), 121-124. DOI : 10.1001/amajethics.2019.121.

Rikap, C. (2022). The expansionary strategies of intellectual monopolies: Google and the digitalization of healthcare. *Economy and Society*, 52(1), 1–27. <https://doi.org/10.1080/03085147.2022.2131271>

- Ristuccia, H., & Rossen, M. (2015). Reputation risk as a board concern. *The Corporate Board (January/February)*, 6-10. <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Governance-Risk-Compliance/gx-gccg-reputation-risk-as-a-board-concern.pdf>
- Roberts, H., Hine, E., & Floridi, L. (2023). Digital Sovereignty, Digital Expansionism, and the Prospects for Global AI Governance. In *Quo Vadis, Sovereignty? New Conceptual and Regulatory Boundaries in the Age of Digital China* (pp. 51-75). Cham: Springer Nature Switzerland.
- Roco, M. C., Rejeski, D., Whitesides, G., Dunagan, J., MacDonald, A., Fisher, E., ... & Suchman, M. (2013). Innovative and Responsible Governance of Converging Technologies. *Convergence of Knowledge, Technology and Society: Beyond Convergence of Nano-Bio-Info-Cognitive Technologies*, 433-489.
- Rodriguez, A., Cotran, H., & Stewart, L. S. (2017). Evaluating the effectiveness of sustainability disclosure: Findings from a recent SASB study. *Journal of Applied Corporate Finance*, 29(2), 100–108. <https://doi.org/10.1111/jacf.12237>
- Rogers, E. M. (1983). *Diffusion of Innovations, 3rd Edition*. The free press
- Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change¹. *The journal of psychology*, 91(1), 93-114. <https://doi.org/10.1080/00223980.1975.9915803>
- Roša, A., & Lobanova, L. (2022). Ethical responsibility of a company in the context of digital transformation of work: Conceptual model. *Sustainability*, 14(18), 11365. <https://doi.org/10.3390/su141811365>
- Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International journal of communication*, 10, 27. <https://dx.doi.org/10.2139/ssrn.2686227>
- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386-408. <https://doi.org/10.1037/H0042519>
- Rothenberger, L., Fabian, B., & Arunov, E. (2019). Relevance of Ethical Guidelines for Artificial Intelligence-a Survey and Evaluation. In *ECIS*.
- Roy, A. (2018). National Strategy for Artificial Intelligence #AIForAll. <https://indiaai.gov.in/research-reports/national-strategy-for-artificial-intelligence/>
- Royakkers, L., Timmer, J., Kool, L., & Van Est, R. (2018). Societal and ethical issues of digitization. *Ethics and Information Technology*, 20, 127–142. <https://doi.org/10.1007/s10676-018-9452-x>
- Rozado, D. (2023). The political biases of chatgpt. *Social Sciences*, 12(3), 148. <http://dx.doi.org/10.3390/socsci12030148>

Rubin H.J. & Rubin I.S. (2005). *Qualitative Interviewing: The Art of Hearing the Data*, 2nd edn. SAGE, Thousand Oaks, CA

Russell, S., & Norvig, P. (2010). *Artificial intelligence: a modern approach*. Pearson Education.

Rutten, C. J., Velthuis, A. G. J., Steeneveld, W., & Hogeveen, H. (2013). Invited review: Sensors to support health management on dairy farms. *Journal of dairy science*, 96(4), 1928-1952. <https://doi.org/10.3168/jds.2012-6107>

Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, 19(1), 61-86. <https://doi.org/10.1108/JICES-12-2019-0138>

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.

Sadhu, S. D., Garg, M., & Kumar, A. (2018). Major environmental issues and new materials. In *New polymer nanocomposites for environmental remediation* (pp. 77-97). Elsevier.

Sætra, H. S. (2023). Generative AI: Here to stay, but for good?. *Technology in Society*, 75, 102372. <https://doi.org/10.1016/j.techsoc.2023.102372>

Safavi, K., & Kalis, B. (2019). How AI can change the future of health care. *Harvard Business Review*. Available at: <https://hbr.org/webinar/2019/02/how-ai-can-change-the-future-of-health-care>

Saggar, A., & Nigam, B. (2023). Maximising net zero in energy-intensive industries: An overview of AI applications for greenhouse gas reduction. *Journal of Climate Change*, 9(1), 13-23. <https://doi.org/10.3233/JCC230003>

Saha, L., Kumar, V., Tiwari, J., Rawat, S., Singh, J., & Bauddh, K. (2021). Electronic waste and their leachates impact on human health and environment: Global ecological threat and management. *Environmental Technology & Innovation*, 24, 102049. <https://doi.org/10.1016/j.eti.2021.102049>

Saleh, J. H., Mark, G., & Jordan, N. C. (2009). Flexibility: a multi-disciplinary literature review and a research agenda for designing flexible engineering systems. *Journal of engineering design*, 20(3), 307-323. <https://doi.org/10.1080/09544820701870813>

Samiha, B. (2013). The importance of the 3R principle of municipal solid waste management for achieving sustainable development. *Mediterranean journal of social sciences*, 4(3), 129-135. [https://doi.org/10.1016/S0959-6526\(03\)00053-2](https://doi.org/10.1016/S0959-6526(03)00053-2)

Sandu, S., & Ciocanel, B. (2014). Impact of R&D and innovation on high-tech export. *Procedia Economics and finance*, 15, 80-90. [https://doi.org/10.1016/S2212-5671\(14\)00450-X](https://doi.org/10.1016/S2212-5671(14)00450-X)

- Sankaran, V. (2023). ChatGPT data centres are consuming a staggering amount of water, study warns. *Independent*. <https://www.independent.co.uk/tech/chatgpt-data-centre-water-consumption-b2318972.html>
- Saraji, S. (2023). Introduction to Blockchain. In *Sustainable Oil and Gas Using Blockchain* (pp. 57-74). Cham: Springer International Publishing.
- Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology*, 24, 4. <https://doi.org/10.1007/s10676-022-09624-3>
- Sasan, J. M. V. (2021). The Social Contract Theories of Thomas Hobbes and John Locke: Comparative Analysis. *Shanlax International Journal of Arts, Science and Humanities*, 9(1), 34-45. <https://doi.org/10.34293/sijash.v9i1.4042>
- Sasubilli, M. K., & Venkateswarlu, R. (2021). Cloud computing security challenges, threats, and vulnerabilities. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 476-480. <https://doi.org/10.1109/ICICT50816.2021.9358709>
- Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2), 34–38. <http://dx.doi.org/10.14569/IJARAI.2013.020206>
- Saveliev, A., & Zhurenkov, D. (2021). Artificial intelligence and social responsibility: The case of the artificial intelligence strategies in the United States, Russia, and China. *Kybernetes*, 50(3), 656–675. <https://doi.org/10.1108/K-01-2020-0060>
- Sayler, K. M. (2020). Artificial intelligence and national security. *Congressional Research Service*, 45178. <https://crsreports.congress.gov/product/details?prodcode=R45178>
- Schiff, D. S., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). *Principles to practices for responsible AI: Closing the gap*. arXiv. <https://doi.org/10.48550/arXiv.2006.04707>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Schmidt, A. L., & Van der Sijde, P. (2022). Disruption by design? Classification framework for the archetypes of disruptive business models. *R&D Management*, 52(5), 893–929. <https://doi.org/10.1111/radm.12530>
- Schneider, G. (2022). Legal challenges of AI supported legal services: Bridging principles and markets. *The Italian Law Journal*, 8, 243–291.
- Schultze, U., & Brooks, J. A. M. (2019). An interactional view of social presence: Making the virtual other “real”. *Information Systems Journal*, 29(3), 707-737. <https://doi.org/10.1111/isj.12230>
- Schumpeter, J. (1934). *The theory of economic development*. Harvard University Press, Cambridge, MA

Schumpeter, J. (1939). *Business cycles: a theoretical historical and statistical analysis of the capitalist process* (2vol). McGraw-Hill, New York

Schwab, K. (2017). *The fourth industrial revolution*. Crown Currency.

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63. <https://doi.org/10.48550/arXiv.1907.10597>

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(3), 417-424. <https://doi.org/10.1017/S0140525X00005756>

Sethi, T. S., & Kantardzic, M. (2017). On the reliable detection of concept drift from streaming unlabeled data. *Expert Systems with Applications*, 82, 77–99. <https://doi.org/10.1016/j.eswa.2017.04.008>

Seyitoğlu, F., Atsız, O., Taş, S., & Kaya, F. (2023). Double-edged perspectives on service robots: working with robots and robots' future career impacts. *Journal of Teaching in Travel & Tourism*, 23(1), 1-19. <https://doi.org/10.1080/15313220.2022.2076768>

Shackleton, J. R. (2018). Robocalypse now? Why we shouldn't panic about automation, algorithms, and artificial intelligence. *Institute of Economic Affairs Current Controversies*, (61). <https://dx.doi.org/10.2139/ssrn.3853643>

Shah, V. (2018). Next-Generation Artificial Intelligence for Personalized Medicine: Challenges and Innovations. *International Journal of Computer Science and Technology*, 2(2), 1-15. <https://jest.com.pk/index.php/jest/issue/view/5>

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, 2(1), 24–28. <https://doi.org/10.1016/j.gltp.2021.01.004>

Sharma, P., & Barua, S. (2023). From Data Breach to Data Shield: The Crucial Role of Big Data Analytics in Modern Cybersecurity Strategies. *International Journal of Information and Cybersecurity*, 7(9), 31–59. Retrieved from: <https://publications.dlpress.org/index.php/ijic/article/view/46>

Shehabi, A., Smith, S. J., Masanet, E., & Koomey, J. (2018). Data center growth in the United States: Decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12), 124030. <https://doi.org/10.1088/1748-9326/aaec9c>

Shirono, M. K., Chhabra, E., Das, M. B., Fan, M. Y., & Villanova, M. H. C. (2021). Is mobile money part of money? Understanding the trends and measurement. *IMF Working Paper*

Shivkumar, K. (2023). The ai paradigm: transforming e-commerce dynamics in the digital age. *International Research Journal of Modernization in Engineering Technology and Science*. 06(01). <https://www.doi.org/10.56726/IRJMETS48902>

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495-504. <https://doi.org/10.1080/10447318.2020.1741118>

- Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine*, 296, 114782. <https://doi.org/10.1016/j.socscimed.2022.114782>
- Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management (JDM)*, 31(2), 74-87. <http://dx.doi.org/10.4018/JDM.2020040105>
- Sigalov, K., Ye, X., König, M., Hagedorn, P., Blum, F., Severin, B., ... & Groß, D. (2021). Automated payment and contract management in the construction industry by integrating building information modeling and blockchain-based smart contracts. *Applied sciences*, 11(16), 7653. <https://doi.org/10.3390/app11167653>
- Silva, L. M. D., Bitencourt, C. C., Faccin, K., & Iakovleva, T. (2019). The role of stakeholders in the context of responsible innovation: A meta-synthesis. *Sustainability*, 11(6), 1766. <https://doi.org/10.3390/su11061766>
- Singh, R., Mishra, S., & Tripathi, K. (2021). Analysing acceptability of E-rickshaw as a public transport innovation in Delhi: A responsible innovation perspective. *Technological Forecasting and Social Change*, 170, 120908. <https://doi.org/10.1016/j.techfore.2021.120908>
- Sithole, M. M., & Buchana, Y. (2021). Effects of innovation activities on employment growth in upper-middle-income countries with high unemployment rates. *Development Southern Africa*, 38(3), 371-390. <https://doi.org/10.1080/0376835X.2020.1796595>
- Sivan, R., & Zukarnain, Z. A. (2021). Security and privacy in cloud-based e-health system. *Symmetry*, 13(5), 742. <https://doi.org/10.3390/sym13050742>
- Sivasubramaniam, N., Liebowitz, S. J., & Lackman, C. L. (2012). Determinants of new product development team performance: A meta-analytic review. *Journal of Product Innovation Management*, 29(5), 803-820. <https://doi.org/10.1111/j.1540-5885.2012.00940.x>
- Sjödin, D., Parida, V., Palmié, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research*, 134, 574–587. <https://doi.org/10.1016/j.jbusres.2021.05.009>
- Slater, T., Lawrence, I. R., Otosaka, I. N., Shepherd, A., Gourmelen, N., Jakob, L., ... & Nienow, P. (2021). Earth's ice imbalance. *The Cryosphere*, 15(1), 233-246. <https://doi.org/10.5194/tc-15-233-2021>
- Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International*, 20(4), 97-106.
- Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard Business Review*, 85(11), 68–76. <https://hbr.org/2007/11/a-leaders-framework-for-decision-making>
- Somashekhar, S. P., Kumarc, R., Rauthan, A., Arun, K. R., Patil, P., & Ramya, Y. E. (2017). Abstract S6-07: Double blinded validation study to assess performance of IBM artificial

intelligence platform, Watson for oncology in comparison with Manipal multidisciplinary tumour board—First study of 638 breast cancer cases. *Cancer Research*, 77(4_Supplement), S6-07. <http://dx.doi.org/10.1158/1538-7445.SABCS16-S6-07>

Son, H. (2017). JP Morgan software does in seconds what took lawyers 360,000 hours.' *Bloomberg*. <https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance>.

Song, D., Shi, E., Fischer, I., & Shankar, U. (2012). Cloud data protection for the masses. *Computer*, 45(1), 39-45. <https://doi.org/10.1109/MC.2012.1>

Song, L., Shokri, R., & Mittal, P. (2019). Privacy risks of securing machine learning models against adversarial examples. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (pp. 241-257). <https://doi.org/10.1145/3319535.3354211>

Sorooshian, S., & Panigrahi, S. (2020). Impacts of the 4th Industrial Revolution on Industries. *Walailak Journal of Science and Technology (WJST)*, 17(8), 903-915. <http://dx.doi.org/10.48048/wjst.2020.7287>

Šotić, A., & Rajić, R. (2015). The review of the definition of risk. *Online Journal of Applied Knowledge Management*, 3(3), 17-26.

Sousa, M. J., & Rocha, Á. (2019). Skills for disruptive digital business. *Journal of Business Research*, 94, 257-263. <https://doi.org/10.1016/j.jbusres.2017.12.051>

Spitz, K., & Trudinger, J. (2019). Metals, Their Biological Functions and Harmful Impacts: Metals are Naturally Occurring Elements. In *Mining and the Environment* (pp. 411-432). CRC Press.

Spruit, S. L., Hoople, G. D., & Rolfe, D. A. (2016). Just a cog in the machine? The individual responsibility of researchers in nanotechnology is a duty to collectivize. *Science and engineering ethics*, 22, 871-887. <https://doi.org/10.1007/s11948-015-9718-1>

Stahl, B. C. (2013). Responsible research and innovation: The role of privacy in an emerging framework. *Science and Public Policy*, 40(6), 708-716. <https://doi.org/10.1093/scipol/sct067>

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., ... & Wright, D. (2021). Artificial intelligence for human flourishing—Beyond principles for machine learning. *Journal of Business Research*, 124, 374-388. <https://doi.org/10.1016/j.jbusres.2020.11.030>

Steering group and secretariat of the Artificial Intelligence Programme. (2019). Leading the way into the age of artificial intelligence Final report of Finland's Artificial Intelligence Programme 2019. *Ministry of Economic Affairs and Employment*. ISBN: 978-952-327-437-2

Steimers, A., & Schneider, M. (2022). Sources of risk of AI systems. *International Journal of Environmental Research and Public Health*, 19(6), 3641. <https://doi.org/10.3390/ijerph19063641>

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580. <http://dx.doi.org/10.1016/j.respol.2013.05.008>

Stilgoe, J., Owen, R., & Macnaghten, P. (2020). Developing a framework for responsible innovation. In *The Ethics of Nanotechnology, Geoengineering, and Clean Energy* (pp. 347-359). Routledge.

Stilgoe, J., Owen, R., and Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>

Stilgoe, S., & Guston, D. H. (2017). *Responsible research and innovation*. The handbook of science and technology studies.

Stoll, C., Klaaßen, L., & Gellersdörfer, U. (2019). The carbon footprint of bitcoin. *Joule*, 3(7), 1647-1661. [https://www.cell.com/joule/abstract/S2542-4351\(19\)30255-7&lang=en](https://www.cell.com/joule/abstract/S2542-4351(19)30255-7&lang=en)

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A. L., Shah, J., Tambe, M., & Teller, A. (2016). Artificial Intelligence and Life in 2030." *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, 52.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. <https://doi.org/10.48550/arXiv.1906.02243>

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09), 13693–13696. <https://doi.org/10.1609/aaai.v34i09.7123>

Stuck, R. E., Holthausen, B. E., & Walker, B. N. (2021). The role of risk in human-robot trust. In *Trust in human-robot interaction* (pp. 179-194). Academic Press. <https://doi.org/10.1016/B978-0-12-819472-0.00008-3>

Stupp, C. (2019). Fraudsters used AI to mimic CEO's voice in unusual cybercrime case. *The Wall Street Journal*, 30(08). Retrieved from: <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>

Subhash, V. (2023). Can Large Language Models Change User Preference Adversarially?. <http://dx.doi.org/10.48550/arXiv.2302.10291>

Sugumaran, V., & Gulla, J. (2012). A. "Applied Semantic Web Technologies: Overview and Future Directions." *CRC Press*, 3-17.

Sun, L., Shang, Z., Xia, Y., Bhowmick, S., & Nagarajaiah, S. (2020). Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection. *Journal of Structural Engineering*, 146(5), 04020073. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0002535](https://doi.org/10.1061/(ASCE)ST.1943-541X.0002535)

- Šurdonja, S., Giuffrè, T., & Deluka-Tibljša, A. (2020). Smart mobility solutions – necessary precondition for a well-functioning smart city. *Transportation Research Procedia*, 45, 604–611. <https://doi.org/10.1016/j.trpro.2020.03.051>
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet policy review*, 8(2). <https://ssrn.com/abstract=3420747>
- Sutcliffe, H. (2011). A report on responsible research and innovation. *MATTER and the European Commission*.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press. <https://mitpress.mit.edu/9780262039246/reinforcement-learning/>
- Sverdlik, Y. (2016). Here's How Much Water All US Data Centers Consume. *Data Center Knowledge*. Available at: <https://www.datacenterknowledge.com/archives/2016/07/12/heres-how-much-water-all-us-data-centers-consume>
- Sverdlik, Y. (2016). Here's How Much Water All US Data Centers Consume. *Data Center Knowledge*. Available at: <https://www.datacenterknowledge.com/archives/2016/07/12/heres-how-much-water-all-us-data-centers-consume>
- Swan, M. (2015). *Blockchain: Blueprint for a new economy*. O'Reilly Media, Inc..
- Tabakhi, S., & Moradi, P. (2015). Relevance–redundancy feature selection based on ant colony optimization. *Pattern Recognition*, 48(9), 2798–2811. <https://doi.org/10.1016/j.patcog.2015.03.020>
- Tamasiga, P., Onyeaka, H., Bakwena, M., Happonen, A., & Molala, M. (2023). Forecasting disruptions in global food value chains to tackle food insecurity: The role of AI and big data analytics—A bibliometric and scientometric analysis. *Journal of Agriculture and Food Research*, 14, 100819. <https://doi.org/10.1016/j.jafr.2023.100819>
- Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to enhance business operations. *MIT Sloan Management Review*, 60(4). <https://sloanreview.mit.edu/article/using-ai-to-enhance-business-operations/>
- Taylor, P. (2018). Big data market size revenue forecast worldwide from 2011 to 2027. *Statista*. Available at: <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>
- Teece, D. J. (2014). A dynamic capabilities-based entrepreneurial theory of the multinational enterprise. *Journal of International Business Studies*, 45, 8-37. <https://doi.org/10.1057/jibs.2013.54>
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509-533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z)
- Teoli, D., Sanvictores, T., & An, J. (2019). SWOT analysis. *Study Guide from StatPearls Publishing, Treasure Island (FL)*.

Thangam, D., Muniraju, H., Ramesh, R., Narasimhaiah, R., Khan, N. M. A., Booshan, S., ... & Ganesh, R. S. (2024). Impact of Data Centers on Power Consumption, Climate Change, and Sustainability. In *Computational Intelligence for Green Cloud Computing and Digital Waste Management* (pp. 60-83). IGI Global.

The Economist. (2017, May 6). Regulating the internet giants: The world's most valuable resource is no longer oil, but data. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

The White House. (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, 2(1), 10–12. <https://doi.org/10.1038/s42256-019-0136-y>

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447-464. <https://doi.org/10.1007/s12525-020-00441-4>

Thomann, P. E. (2023). Geopolitical Competition and the Challenges for the European Union of Countering the Malicious Use of Artificial Intelligence. In *The Palgrave Handbook of Malicious Use of AI and Psychological Security* (pp. 453-486). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-22552-9_17

Tian, H., & Tian, J. (2021). The mediating role of responsible innovation in the relationship between stakeholder pressure and corporate sustainability performance in times of crisis: Evidence from selected regions in China. *International Journal of Environmental Research and Public Health*, 18(14), 7277. <https://doi.org/10.3390/ijerph18147277>

Toledo, M. (2020). Video game habits & COVID-19. <https://dx.doi.org/10.2139/ssrn.3676004>

Tolles, J., & Meurer, W. J. (2016). Logistic regression: relating patient characteristics to outcomes. *Jama*, 316(5), 533-534. <https://doi.org/10.1001/jama.2016.7653>

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., ... & Clopath, C. (2020). AI for social good: unlocking the opportunity for positive impact. *Nature Communications*, 11(1), 2468. <https://doi.org/10.1038/s41467-020-15871-z>

Torres-Carrión, González-González, Toledo-Delgado, Muñoz-Cruz, Gil-Iranzo, Reyes-Alonso, & Hernández-Morales. (2019). Improving Cognitive Visual-Motor Abilities in Individuals with Down Syndrome. *Sensors*, 19(18), 3984. <https://doi.org/10.3390/s19183984>

Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., ... & Manser, E. (2019). Considerations for AI fairness for people with disabilities. *AI Matters*, 5(3), 40-63. <https://doi.org/10.1145/3362077.3362086>

- Trittin-Ulbrich, H., & Böckel, A. (2022). Institutional entrepreneurship for responsible digital innovation: The case of corporate digital responsibility. *Creativity and Innovation Management*, 31(3), 447-459. <https://doi.org/10.1111/caim.12513>
- Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2023). Responsible AI for digital health: A synthesis and a research agenda. *Information Systems Frontiers*, 25, 2139–2157. <https://doi.org/10.1007/s10796-021-10146-4>
- Truby, J., Brown, R., & Dahdal, A. (2020). Banking on AI: mandating a proactive approach to AI regulation in the financial sector. *Law and Financial Markets Review*, 14(2), 110-120. <https://doi.org/10.1080/17521440.2020.1760454>
- Tseng, H. H. (2015). The development of the personal and professional values-integrated framework as an aid to ethical decision making. *American Journal of Educational Research*, 3(9), 1163-1167. DOI:10.12691/education-3-9-16
- Tripathi, A., Misra, K., Dhanuka, R., & Singh, J. P. (2023). Artificial intelligence in accelerating drug discovery and development. *Recent Patents on Biotechnology*, 17(1), 9-23. <https://doi.org/10.2174/1872208316666220802151129>
- 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. <https://doi.org/10.1016/j.techfore.2021.121100>
- Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, 9(1), 49–62. <https://doi.org/10.1007/s10676-006-9128-9>
- Turner III, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The qualitative report*, 15(3), 754. <http://www.nova.edu/ssss/QR/QR15-3/qid.pdf>
- Tuukkanen, V., Wolgsjö, E., & Rusu, L. (2022). Cultural values in digital transformation in a small company. *Procedia Computer Science*, 196, 3-12. <https://doi.org/10.1016/j.procs.2021.11.066>
- Tzimas, T. (2021). The Ethics of AI. In: *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective*. *Law, Governance and Technology Series*, vol 46, 69-99. Springer, Cham. https://doi.org/10.1007/978-3-030-78585-7_4
- Uddin, M. S., Chi, G., Al Janabi, M. A., & Habib, T. (2022). Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability. *International Journal of Finance & Economics*, 27(3), 3713-3729. <https://doi.org/10.1002/ijfe.2346>
- Ulnicane, I. (2022). Artificial Intelligence in the European Union: Policy, ethics and regulation. In *The Routledge handbook of European integrations*. Taylor & Francis.
- Undersecretariat of Information Technologies of the Cabinet Chief. (2023). Argentina: Recommendations for a reliable Artificial Intelligence. Retrieved from: <https://insightplus.bakermckenzie.com/bm/data-technology/argentina-recommendations-for-a-reliable-artificial-intelligence>

Université de Montréal. (2018). Montreal Declaration for the Responsible Development of Artificial Intelligence. <https://montrealdeclaration-responsibleai.com/>

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, 18(6), 463-477. <https://doi.org/10.1038/s41573-019-0024-5>

van Buitenlandse Zaken, M. (2023). Global Declaration on Information Integrity Online-Diplomatic statement-Government.

van den Hoven, J. (2017). Ethics for the digital age: Where are the moral specs? Value sensitive design and responsible innovation. In *Informatics in the Future: Proceedings of the 11th European Computer Science Summit (ECSS 2015), Vienna, October 2015* (pp. 65-76). Cham: Springer International Publishing.

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers' Service Experiences. *Journal of Service Research*, 20(1), 43–58. <https://doi.org/10.1177/1094670516679272>

Van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, 144, 93-106. <https://doi.org/10.1016/j.jbusres.2022.01.076>

van Laer, T., Feiereisen, S., & Visconti, L. M. (2019). Storytelling in the digital era: A meta-analysis of relevant moderators of the narrative transportation effect. *Journal of Business Research*, 96, 135–146. <https://doi.org/10.1016/j.jbusres.2018.10.053>

Van Staveren, I. (2007). Beyond utilitarianism and deontology: Ethics in economics. *Review of Political Economy*, 19(1), 21-35. <https://doi.org/10.1080/09538250601080776>

Van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1, 213–218. <https://doi.org/10.1007/s43681-021-00043-6>

Vantage Market Research. (2022). AI in Manufacturing Market. <https://www.vantagemarketresearch.com/industry-report/ai-in-manufacturing-market-1404>

Vara-Miguel, A., Negredo-Bruna, S., & Amoedo-Casais, A. (2017). Digital News Report. ES 2017. <https://doi.org/10.15581/019.007>

Varkey, B. (2021). Principles of clinical ethics and their application to practice. *Medical Principles and Practice*, 30(1), 17-28. <https://doi.org/10.1159/000509119>

Vasisht, D., Kapetanovic, Z., Won, J., Jin, X., Chandra, R., Sinha, S., ... & Stratman, S. (2017). {FarmBeats}: an {IoT} platform for {Data-Driven} agriculture. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)* (pp. 515-529).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017) Attention is all you need. *Advances in neural information processing systems*, 30. <https://doi.org/10.48550/arXiv.1706.03762>

Vergeer, M. (2020). Artificial Intelligence in the Dutch Press: An Analysis of Topics and Trends. *Communication Studies*, 71(3), 373–392. <https://doi.org/10.1080/10510974.2020.1733038>

Verhezen, P. (2020). What to expect from artificial intelligence in business: How wise board members can and should facilitate human–AI collaboration. In M. Y. Kuznetsov & M. I. Nikishova (Eds.), *Challenges and opportunities of corporate governance transformation in the digital era* (pp. 61–90). IGI Global. <https://doi.org/10.4018/978-1-7998-2011-6>

Verma, P., & Oremus, W. (2023). These lawyers used ChatGPT to save time. They got fired. Washingtonpost. <https://www.washingtonpost.com/technology/2023/11/16/chatgpt-lawyer-fired-ai/>

Vervoort, J., & Gupta, A. (2018). Anticipating climate futures in a 1.5 C era: the link between foresight and governance. *Current Opinion in Environmental Sustainability*, 31, 104–111. <https://doi.org/10.1016/j.cosust.2018.01.004>

Vetter, T. R., & Mascha, E. J. (2017). Bias, confounding, and interaction: Lions and tigers, and bears, oh my! *Anesthesia & Analgesia*, 125(3), 1042–1048. <https://doi.org/10.1213/ANE.0000000000002332>

Vigneau, L., Humphreys, M., & Moon, J. (2015). How do firms comply with international sustainability standards? Processes and consequences of adopting the global reporting initiative. *Journal of Business Ethics*, 131, 469–486. <https://doi.org/10.1007/s10551-014-2278-5>

Villar, A. S., & Khan, N. (2021). Robotic process automation in banking industry: a case study on Deutsche Bank. *Journal of Banking and Financial Technology*, 5(1), 71–86. <https://doi.org/10.1007/s42786-021-00030-9>

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-019-14108-y>

Vivarelli, M. (2014). Innovation, employment and skills in advanced and developing countries: A survey of economic literature. *Journal of Economic Issues*, 48(1), 123–154. <https://doi.org/10.2753/JEI0021-3624480106>

Vlasov, A. I., Grigoriev, P. V., Krivoshein, A. I., Shakhnov, V. A., Filin, S. S., & Migalin, V. S. (2018). Smart management of technologies: predictive maintenance of industrial equipment using wireless sensor networks. *Entrepreneurship and Sustainability Issues*, 6(2), 489–502. [https://dx.doi.org/10.9770/jesi.2018.6.2\(2\)](https://dx.doi.org/10.9770/jesi.2018.6.2(2))

Von der Leyen, U. (2019). A Union that strives for more my agenda for Europe. Retrieved from: <https://op.europa.eu/en/publication-detail/-/publication/43a17056-ebf1-11e9-9c4e-01aa75ed71a1>

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607–1622. <https://doi.org/10.1007/s13347-021-00477-0>

- von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. In: *Dusseldorf, M., Beecroft, R. (eds) Technikfolgen abschätzen lehren*. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93468-6_2
- Von Schomberg, R. (2013). A vision of responsible research and innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society*, 51-74. <https://doi.org/10.1002/9781118551424.ch3>
- von Wolff, R. M., Hobert, S., & Schumann, M. (2022). Designing process-based chatbots in enterprises: The case of business travel organization considering the users' perspective and business value. *AIS Transactions on Human-Computer Interaction*, 14(4), 578-623. <https://doi.org/10.17705/1thci.00180>
- Voumik, L. C., & Sultana, T. (2022). Impact of urbanization, industrialization, electrification and renewable energy on the environment in BRICS: fresh evidence from novel CS-ARDL model. *Heliyon*, 8(11). <https://doi.org/10.1016/j.heliyon.2022.e11457>
- 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. <https://doi.org/10.1080/09585192.2020.1871398>
- Vyas, S., Gupta, S., Bhargava, D., & Boddu, R. (2022). Fuzzy logic system implementation on the performance parameters of health data management frameworks. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/9382322>
- Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., ... & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7-30. <http://dx.doi.org/10.15678/EBER.2023.110201>
- wael AL-khatib, A. (2023). Drivers of generative artificial intelligence to fostering exploitative and exploratory innovation: A TOE framework. *Technology in Society*, 75, 102403. <https://doi.org/10.1016/j.techsoc.2023.102403>
- Walkowiak, E., & MacDonald, T. (2023). Generative AI and the Workforce: What Are the Risks?. <https://dx.doi.org/10.2139/ssrn.4568684>
- Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893–1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>
- Wang, J., Shan, Z., Gupta, M., & Rao, H. R. (2019). A longitudinal study of unauthorized access attempts on information systems: The role of opportunity contexts. *MIS Quarterly*, 43(2). <https://doi.org/10.25300/MISQ/2019/14751>
- Wang, W., & Siau, K. (2018). Ethical and moral issues with AI. *Twenty-fourth Americas Conference on Information Systems, New Orleans, 2018*

Wang, X., & Cheng, Z. (2020). Cross-sectional studies: Strengths, weaknesses, and recommendations. *Chest*, 158(1), S65–S71. <https://doi.org/10.1016/j.chest.2020.03.012>

Weichbroth, P., Wereszko, K., Anacka, H., & Kowal, J. (2023). Security of cryptocurrencies: A view on the state-of-the-art research and current developments. *Sensors*, 23(6), 3155. <https://doi.org/10.3390/s23063155>

Weili, L., Khan, H., Khan, I., & Han, L. (2022). The impact of information and communication technology, financial development, and energy consumption on carbon dioxide emission: evidence from the Belt and Road countries. *Environmental Science and Pollution Research*, 1-16. <https://doi.org/10.1007/s11356-021-18448-5>

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45. <https://doi.org/10.1145/357980.357991>

Weller, A. (2019). Transparency: Motivations and Challenges. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L., Müller, KR. (eds) *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science*, vol 11700. Springer, Cham. https://doi.org/10.1007/978-3-030-28954-6_2

Wellmer, F. W., Buchholz, P., Gutzmer, J., Hagelüken, C., Herzig, P., Littke, R., & Thauer, R. K. (2019). *Raw materials for future energy supply*. Cham: Springer International Publishing.

Wendelken, A., Danzinger, F., Rau, C., & Moeslein, K. M. (2014). Innovation without me: Why employees do (not) participate in organizational innovation communities. *R&D Management*, 44(2), 217–236. <https://doi.org/10.1111/radm.12042>

Werden, G. J., Froeb, L. M., & Tardiff, T. J. (1996). The use of the logit model in applied industrial organization. *International Journal of the Economics of Business*, 3(1), 83-105. <http://dx.doi.org/10.1080/758533490>

West, M., Kraut, R., & Ei Chew, H. (2021). *I'd blush if I could: Closing gender divides in digital skills through education*. UNESCO. <https://doi.org/10.54675/RAPC9356>

Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in methods and practices in psychological science*, 2(3), 214-227. <http://dx.doi.org/10.1177/2515245919848684>

Wewege, L., Lee, J., & Thomsett, M. C. (2020). Disruptions and digital banking trends. *Journal of Applied Finance and Banking*, 10(6), 15-56.

Whittlestone, J., Nyrop, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 195-200). <http://dx.doi.org/10.1145/3306618.3314289>

- Wickson, F., & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: Learning from transdisciplinarity. *Journal of Responsible Innovation*, 1(3), 254-273. <https://doi.org/10.1080/23299460.2014.963004>
- Wiener, N. (1961). *Cybernetics: Or Control and Communication in the Animal and the Machine* (2nd ed.). Cambridge, MA: MIT Press.
- Willcock, S., Martínez-López, J., Hooftman, D. A., Bagstad, K. J., Balbi, S., Marzo, A., ... & Athanasiadis, I. N. (2018). Machine learning for ecosystem services. *Ecosystem services*, 33, 165-174. <https://doi.org/10.1016/j.ecoser.2018.04.004>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223-235. <https://doi.org/10.1080/17439884.2020.1798995>
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review*. Available at: <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces>
- Wilson, R. and Crouch, A. (1982). *Risk/Benefit Analysis*. Ballinger Publishing Company, Cambridge.
- Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]. *Proceedings of the IEEE*, 107(3), 509-517. <https://doi.org/10.1109/JPROC.2019.2900622>
- Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, 39(4), 101685. <https://doi.org/10.1016/j.giq.2022.101685>
- Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, 39(4), 101685. <https://doi.org/10.1016/j.giq.2022.101685>
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, 29(5), 907–931. <https://doi.org/10.1108/JOSM-04-2018-0119>
- Wolpert, D. H. (2021). What is important about the no free lunch theorems?. In *Black box optimization, machine learning, and no-free lunch theorems* (pp. 373-388). Cham: Springer International Publishing.
- Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, 61(6), 823-832. <https://doi.org/10.1016/j.bushor.2018.07.001>
- Wu, S. P. J., Straub, D. W., & Liang, T. P. (2015). How information technology governance mechanisms and strategic alignment influence organizational performance. *MIS Quarterly*, 39(2), 497–518. <https://www.jstor.org/stable/26628363>

- Xin, M., Xing, J., Pengfei, W., Houru, L., Mengcheng, W., & Hong, Z. (2018). Online activities, prevalence of Internet addiction and risk factors related to family and school among adolescents in China. *Addictive behaviors reports*, 7, 14-18. <https://doi.org/10.1016/j.abrep.2017.10.003>
- Xu, M., David, J. M., & Kim, S. H. (2018). The fourth industrial revolution: Opportunities and challenges. *International journal of financial research*, 9(2), 90-95. <http://dx.doi.org/10.5430/ijfr.v9n2p90>
- Xu, R., Tao, Y., Lu, Z., & Zhong, Y. (2018). Attention-mechanism-containing neural networks for high-resolution remote sensing image classification. *Remote Sensing*, 10(10), 1602. <https://doi.org/10.3390/rs10101602>
- Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3313831.3376301>
- Yang, Y., Jiang, Y., Jia, F., & Chen, L. (2022). The impact of supplier instability on corporate social responsibility performance over the firm lifecycle: A social systems theory perspective. *British Journal of Management*, 0, 1-23. <https://doi.org/10.1111/1467-8551.12651>.
- Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., ... & Gadekallu, T. R. (2024). Gpt (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3389497>
- Yigitcanlar, T. (2011). Position paper: redefining knowledge-based urban development. *International Journal of Knowledge-Based Development*, 2(4), 340-356. <https://doi.org/10.1504/IJKBD.2011.044343>
- Yigitcanlar, T. (2011). Position paper: redefining knowledge-based urban development. *International Journal of Knowledge-Based Development*, 2(4), 340-356. <https://doi.org/10.1504/IJKBD.2011.044343>
- Yimer, A. H. (2021). Challenging the challenges of e-government: The Ethiopian context. *International Journal of Digital Strategy, Governance, and Business Transformation (IJDSGBT)*, 11(1), 1–12. <http://doi.org/10.4018/IJDSGBT.286772>
- Yong, J. Y., Yusliza, M. Y., Ramayah, T., Chiappetta Jabbour, C. J., Sehnem, S., & Mani, V. (2020). Pathways towards sustainability in manufacturing organizations: Empirical evidence on the role of green human resource management. *Business Strategy and the Environment*, 29(1), 212–228. <https://doi.org/10.1002/bse.2359>
- Yoo, Y., Boland Jr, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science*, 23(5), 1398-1408. <https://doi.org/10.1287/orsc.1120.0771>
- Yu, I. S., & Cheung, W. Y. (2018). Metastatic colorectal cancer in the era of personalized medicine: a more tailored approach to systemic therapy. *Canadian Journal of Gastroenterology and Hepatology*, 2018.

- Yu, R., & Ali, G. S. (2019). What's inside the black box? AI challenges for lawyers and researchers. *Legal Information Management*, 19(1), 2-13. <https://doi.org/10.1017/S1472669619000021>
- Zacklad, M. (2019). Le design de l'information : textualisation, documentarisation, auctorialisation. *Communication & langages*, 199, 37-64. <https://doi.org/10.3917/comla1.199.0037>
- Zacklad, M., Joubert, D. & Gérard, A. (2023). Co-design de dispositifs de prévention collaboratifs au service de la prévention des rixes et de la police de sécurité. *Approches Théoriques en Information-Communication (ATIC)*, 6, 61-87. <https://doi.org/10.3917/atic.006.0061>
- Zajko, M. (2021). Conservative AI and social inequality: Conceptualizing alternatives to bias through social theory. *AI & SOCIETY*, 36(3), 1047–1056. <https://doi.org/10.1007/s00146-021-01153-9>
- Zeide, E. (2017). The Structural Consequences of Big Data-Driven Education. *Big Data*, 5(2), 164–172. <https://doi.org/10.1089/big.2016.0061>
- Zeleznikowl, J. (2017). *Can Artificial Intelligence And Oiine Dispute Resolution Ehance Efhciency And Effectiveness In Courts.*
- Zewde, E. A., Tolossa, T., Tiruneh, S. A., Azanaw, M. M., Yitbarek, G. Y., Admasu, F. T., ... & Melaku, M. D. (2022). Internet addiction and its associated factors among African high school and university students: systematic review and meta-analysis. *Frontiers in psychology*, 13, 847274. <https://doi.org/10.3389/fpsyg.2022.847274>
- Zhanbayev, R. A., Irfan, M., Shutaleva, A. V., Maksimov, D. G., Abdykadyrkyzy, R., & Filiz, Ş. (2023). Demoethical model of sustainable development of society: A roadmap towards digital transformation. *Sustainability*, 15(16), 12478. <https://doi.org/10.3390/su151612478>
- Zhang, J., & Hon, H. W. (2020). Towards responsible digital transformation. *California Management Review Insights*.
- Zhang, K., Qian, Q., & Zhao, Y. (2020). Evolution of Guangzhou biomedical industry innovation network structure and its proximity mechanism. *Sustainability*, 12(6), 2456. <https://doi.org/10.3390/su12062456>
- Zhang, P., & Kamel Boulos, M. N. (2023). GAI in medicine and healthcare: Promises, opportunities and challenges. *Future Internet*, 15(9), 286. <https://doi.org/10.3390/fi15090286>
- Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, 57(2), 102025. <https://doi.org/10.1016/j.ipm.2019.03.004>
- Zhang, Y. (2023). Analysis on the Development Strategy of Hainan's Sports Tourism Informatization in the Digital Era. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 19(1), 1-18. DOI: 10.4018/IJSWIS.325788

Zhang, Y., Song, F., & Sun, J. (2023). QEBVerif: Quantization error bound verification of neural networks. In C. Enea & A. Lal (Eds.), *Computer aided verification* (pp. 413–437). Springer. https://doi.org/10.1007/978-3-031-37703-7_20

Zhong, S., Xin, C., Shen, H., & Chen, X. (2021). Effects of land urbanization and internet penetration on environmental sustainability: A cross-regional study of China. *Environmental Science and Pollution Research*, 28(47), 66751–66771. <https://doi.org/10.1007/s11356-021-15226-1>

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). A survey on ethical principles of AI and implementations. In 2020 *IEEE Symposium Series on Computational Intelligence* (SSCI) (pp. 3010-3017). IEEE. <http://dx.doi.org/10.1109/SSCI47803.2020.9308437>

Zhu, Z., Lin, H., Gu, D., Wang, L., Wu, H., & Fang, Y. (2023). MusREL: A utility-weighted multi-strategy relation extraction model-based intelligent system for online education. *International Journal on Semantic Web and Information Systems*, 19(1), 1–19. <https://doi.org/10.4018/IJSWIS.329965>

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), 479–493. <https://doi.org/10.1002/asi.20508>

Zou, J., & Schiebinger, L. (2021). Ensuring that biomedical AI benefits diverse populations. *EBioMedicine*, 67. <https://doi.org/10.1016/j.ebiom.2021.103358>

Zuboff, S. (2019, January). Surveillance capitalism and the challenge of collective action. In *New labor forum* (Vol. 28, No. 1, pp. 10-29). Sage CA: Los Angeles, CA: SAGE Publications.

Zwart, H., Landeweerd, L., & Van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the European research funding arena from ‘ELSA’ to ‘RRI’. *Life Sciences, Society And Policy*, 10, 1-19. <https://doi.org/10.1186/s40504-014-0011-x>

Annex 1

Interview segments	Questions
Introduction	"Thank you for agreeing to participate in this interview. We are conducting this study in collaboration with Cap Digital, focusing on Responsible Digital Innovation within companies. Your insights will be invaluable to our research."
Defining RDI	"How do you define Responsible Digital Innovation?"
Dimensions of RDI	"Could you describe the key dimensions of Responsible Digital Innovation as you see them in your company's context?"
Company's position	"What is your company position on this concept: responsible digital innovation"?
Practices for RDI	"What specific practices does your company undertake to ensure Responsible Digital Innovation?"
Barriers to RDI	"What are the main barriers your company faces in implementing Responsible Digital Innovation?"
Exploration of unaddressed areas	"In your view, what aspects related to Responsible Digital Innovation are currently missing not only within your company but also in the industry at large?"
Conclusion	"We greatly value the insights you've shared with us today about Responsible Digital Innovation at your company. Your contribution is essential to our research. Please, if you wish to share any additional documents or information, please do not hesitate to send them to us."

Annex 2

Major Global Players			
Categories and sub-categories	Name of the declaration	Principles	Year of Publication
OECD	OECD Principles on Artificial Intelligence	(1) Inclusive growth, sustainable development, and well-being; (2) Human-centered values and fairness; (3) Transparency and explainability; (4) Robustness, security, and safety; (5) Accountability.	2019
European Union	Ethics Guidelines for Trustworthy AI	(1) Human agency and oversight, (2) Technical robustness and safety, (3) Privacy and data governance, (4) Transparency, (5) Diversity, non-discrimination, and fairness, (6) Societal and environmental well-being, (7) Accountability.	2019
G20	G20 AI Principles	(1) Inclusive growth, sustainable development, and well-being; (2) Human-centered values and fairness; (3) Transparency and explainability; (4) Robustness, security, and safety; (5) Accountability.	2019
UNESCO	Recommendation on the Ethics of Artificial Intelligence	(1) Proportionality and Do No Harm; (2) Safety and security; (3) Fairness and non-discrimination; (4) Sustainability; (5) Right to Privacy, and Data Protection; (6) Human oversight and determination; (7) Transparency and explainability; (8) Responsibility and accountability; (9) Awareness and literacy; (10) Multi-stakeholder and adaptive governance and collaboration	2021
United Nations General Assembly High-Level Week	Global Declaration on Information Integrity Online	(1) Abstaining from and condemning state-led disinformation, (2) Respecting and promoting freedom of expression, (3) Implementing legislation compliant with international law, (4) Promoting digital literacy and supporting independent media, (5) Safeguarding human rights in countering disinformation.	2023
United States	Blueprint for an AI Bill of Rights	(1) Safety and effective system, (2) Algorithm discrimination protections, (3) Data Privacy (4) Notice and explanation, (5) Human alternatives, consideration, and fallback	2023

Non-Governmental Organizations			
Categories and sub-categories	Name of the declaration	Principles	Year of Publication
Future of Life	Asilomar AI Principles	(1) Safety, (2) Failure transparency, (3) Judicial transparency, (4) Responsibility, (5) Value alignment, (6) Human value, (7) Personal privacy, (8) Liberty and privacy, (9) Shared prosperity.	2017
Montreal Declaration	Montreal Declaration for the Responsible Development of Artificial Intelligence	(1) Well-being, (2) Respect for autonomy, (3) Protection of privacy and intimacy, (4) Solidarity, (5) Democratic participation, (6) Equity, (7) Diversity inclusion, (8) Prudence, (9) Responsibility, (10) Sustainable development.	2018
CEPEJ	Ethical Charter on the Use of AI in Judicial Systems	(1) Respect for fundamental rights, (2) Non-discrimination, (3) Quality and security, (4) Transparency, impartiality, and fairness, and (5) User control over AI systems.	2018
Institute of Electrical and Electronics Engineers (IEEE)	Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems	(1) Human rights, (2) Well-being, (3) Data agency, (4) Effectiveness, (5) Transparency, (6) Accountability, (7) Awareness of misuse, (8) Competency.	2019

Technology Companies				
Categories and sub-categories		Name of the declaration	Principles	Year of Publication
Google		Google AI principles	Socially beneficial, Avoid creating or reinforcing unfair bias, Built and tested for safety, Accountable to people, Incorporate privacy design principles, Uphold high standards of scientific excellence	2018
Amazon		Transformez l'IA responsable de la théorie en pratique	(1) Equity, (2) Robustness, (3) Governance, (4) Fairness, (5) Transparency, (6) Explainability, (7) Privacy, (8) Security.	n.d
Facebook (Meta)		Facebook's AI Ethics and Principles	(1) Privacy and security, (2) Fairness and inclusion, (3) Robustness and safety, (4) Transparency and control, (5) Accountability and governance.	2021
Apple		Values and Principles of Apple	(1) Privacy, (2) Safety, (3) Accessibility (4) User control, (5) Responsibility, (6) Environment, (7) Education	2023
Microsoft		Empowering responsible AI practices	(1) Fairness, (2) Reliability and safety, (3) Privacy and security, (4) Inclusiveness, (5) Transparency, (6) accountability.	n.d
National and Regional Entities				
Categories and sub-categories		Name of the declaration	Principles	Year of Publication
Western Europe	Italy	Proposte per una Strategia italiana	(1) Transparency, (2) Accountability, and (3) Reliability.	2018
	France	AI for humanity	(1) Transparency, (2) Fairness, (3) Security, (4) Accountability, (5) Promoting ethics education for all actors in the "algorithmic chain" (designers, professionals, and citizens), (6) Societal Well-being, (7) Open data policy.	2018
	Finland	Finland has determined ethical principles for the responsible use of human artificial intelligence	(1) Human-centeredness, (2) Transparency, (3) Explainability, (4) Fairness and non-discrimination, (5) Accountability and trust, (6) Privacy, (7) Safety, (8) Human oversight.	2023
Latin America & The Caribbean	Brazil	Summary of the Brazilian Artificial Intelligence Strategy -EBIA-	(1) Respect for human dignity, freedom, democracy, and equality; (2) Respect for human rights and diversity; (3) The guarantee of protection of privacy and personal data; (4) Transparency, reliability, and audibility; (5) Human oversight.	2021
	Argentina	Recommendations for reliable artificial intelligence	(1) Proportionality and Harmlessness, (2) Safety and Security, (3) Equity and Nondiscrimination, (4) Sustainability, (5) Right to Privacy and Data Protection, (6) Oversight and Human Decision-Making, (7) Transparency and Explainability, (8) Responsibility and Accountability.	2023
Middle East & North Africa	United Arab Emirates	UAE National Strategy For Artificial Intelligence 2031	(1) Fairness, (2) Inclusiveness, (3) Equality, (4) Human Benefit, (5) Human Rights	2018
	Saudi Arabia	Saudi Arabia's AI Ethics Principles 2.0: A New Era of AI Governance	(1) Fairness, (2) Privacy & Security, (3) Humanity, (4) Social & Environmental Benefits, (5) Reliability & Safety, (6) Transparency & Explainability, (7) Accountability & Responsibility	2023
South & Central Asia	India	National Strategy for Artificial Intelligence #AIForAll	(1) Accountability, (2) Data protection, (3) Transparency, (4) Fairness, (5) Inclusive, (6) Reliability, (7) Accountability icon, (8) AI inclusivity principle, (9) Beneficiality, (10) Explainability, (11) Fair and unbiased, (12) Privacy preference center, (13) Security, (14) Bias mitigation, (15) Ethics, (16) Sustainability.	2021
	Turkey	National artificial intelligence strategy 2021-2025	(1) Explainable, (2) Trustworthy, (3) Unbiased, (4) Fairness, (5) Accountability, (6) Transparency.	2021
East Asia	Singapore	Singapore National AI Strategy 2.0	(1) Transparency and Explainability, (2) Fairness, (3) Human-Centricity, (4) Safety and Reliability, (5) Security and Privacy, (6) Responsibility and Accountability.	2023
	China	Beijing AI Principles	(1) Humanity and Beneficence, (2) Ethical Awareness and Responsibility, (3) Diversity and Inclusiveness, (4) Transparency and Openness, (5) Accountability and Governance, (6) Informed Consent and Stakeholder Rights, (7) Education and Empowerment, (8) Long-term Safety and Existential Risk Awareness.	2019
Australia	Australia	Australian Government AI Ethics Framework	(1) Generate net benefits, (2) Ensure non-harmful AI deployment, (3) Uphold regulatory and legal compliance, (4) Protect privacy, (5) Promote fairness, (6) Encourage transparency and explainability, (7) Foster contestability, (8) Establish accountability.	2018