INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Ethics of Artificial Intelligence: A Bibliometric Review Analysis

Cândida Alexandra Caramba de Jesus

Master in Business Administration

Supervisor:
Ph.D. Carlos Miguel Correia Hernandez Jerónimo, Invited Assistant Professor,
ISCTE Business School

October, 2023

BUSINESS
SCHOOL

Department of Marketing, Strategy and Operations

Ethics of Artificial Intelligence: A Bibliometric Review Analysis

Cândida Alexandra Caramba de Jesus

Master in Business Administration

Supervisor:
Ph.D. Carlos Miguel Correia Hernandez Jerónimo, Invited Assistant Professor,
ISCTE Business School

October, 2023

To my grandfather

"Computers will overtake humans with AI at some point within the next 100 years. When that happens, we need to make sure the computers have goals aligned with ours. Our future is a race between the growing power of technology and the wisdom with which we use it."

Stephen Hawking at the Zeitgeist 2015 conference in London

Acknowledgements

I would like to express my heartfelt gratitude to the ISCTE Business School for providing me with the invaluable opportunity to undertake this research. Their support and resources have been instrumental in the successful completion of this study.

I am deeply indebted to my supervisor, Carlos Jerónimo, whose guidance, expertise, and unwavering support have been pivotal throughout this research journey. His mentorship has been a source of inspiration and motivation, shaping the direction and quality of this work.

I extend my profound thanks to my husband for his unwavering encouragement, patience, and understanding during the long hours and late nights spent on this research. His support has been the cornerstone of my perseverance.

I also wish to acknowledge the countless individuals whose work and contributions have formed the foundation of this research, including the authors, scholars, and researchers in the field of AI ethics.

Lastly, my appreciation goes out to all my colleagues, friends, and family who have offered their encouragement and support, making this research possible. Your belief in me has been a driving force in my academic journey.

Resumo

A Inteligência Artificial (IA) tornou-se um componente integral da nossa sociedade global, transformando o presente e tendo o potencial de moldar o futuro, tornando, assim, o escrutínio ético desta tecnologia, fundamental. Esta revisão bibliométrica apresenta uma análise abrangente do panorama ético da Inteligência Artificial entre 2019 e 2023, enfatizando a importância das considerações éticas no domínio da IA. O principal objectivo foi compreender a natureza multifacetada dos princípios éticos que guiam o desenvolvimento da IA, identificando tendências-chave e avaliando as estruturas intelectuais, conceptuais e sociais em evolução no âmbito ético da IA. A metodologia envolveu uma análise bibliométrica de artigos de pesquisa das bases de dados WoS e Scopus, usando técnicas de co-citação e co-palavras para explorar ligações e temas. Entre as conclusões principais, destacam-se o reconhecimento de oito de doze princípios éticos centrais na IA, conforme sugerido e identificado por Jobin et. Al., (2019): transparência, justiça e equidade, responsabilidade, privacidade, confiança, sustentabilidade e beneficência. Além disso, é admitido o aumento das contribuições académicas e a natureza dinâmica deste campo. A pesquisa também destaca revistas, autores e países influentes, sublinhando a natureza interdisciplinar da ética da IA. É crucial um compromisso interdisciplinar contínuo para enfrentar os desafios éticos emergentes e desenvolver implicações práticas, diretrizes éticas e novos enquadramentos neste campo em evolução. Estas noções são essenciais para orientar o desenvolvimento ético da IA num cenário em constante mutação.

PALAVRAS-CHAVE: ética em IA, princípios éticos, colaboração interdisciplinar, análise bibliométrica, estrutura intelectual, análise de co-citação, análise de co-palavras.

CLASSIFICAÇÃO JEL: M10, O33, K23
M10 – Business Administration: General
O33 – Technological Change: Choices and Consequences, Diffusion Processes
K23 – Law and Economics: Regulated Industries and Administrative Law

Abstract

Artificial intelligence (AI) has become an integral part of our global society, transforming the present and holding the potential to shape the future, making ethical scrutiny crucial. This bibliographic analysis presents a comprehensive analysis of the landscape of Artificial Intelligence ethics from 2019 to 2023, emphasising the relevance of ethical considerations within artificial intelligence. The primary objective was to understand the multifaceted nature of ethical principles guiding AI development, identify key trends, and assess the evolving intellectual, conceptual, and social structures in AI ethics. The methodology involved a bibliometric analysis of research articles from the WoS and Scopus databases, employing co-citation and co-word techniques to explore connections and themes. Key findings include the recognition of eight of twelve core AI ethics principles as suggested and identified by Jobin et al. (2019): transparency, justice and fairness, responsibility, privacy, trust, sustainability, and beneficence. Moreover, the increasing scholarly contributions and the dynamic nature of the field are acknowledged. The research also highlights influential journals, authors, and countries, underscoring the interdisciplinary nature of AI ethics. Continued interdisciplinary engagement to address emerging ethical challenges and develop practical implications, ethical guidelines, and novel frameworks for this evolving field is crucial. These insights are essential for guiding ethical AI development in a rapidly changing landscape.

KEYWORDS: AI ethics, ethical principles, interdisciplinary collaboration, bibliometric analysis, intellectual structure, co-citation, co-word analysis.

JEL CLASSIFICATION: M10, O33, K23

M10 – Business Administration: General

O33 – Technological Change: Choices and Consequences, Diffusion Processes

K23 – Law and Economics: Regulated Industries and Administrative Law

Table of Contents

Acknowledgements iii

Resumo v

Abstract vii

List of Figures ix

List of Tables xi

CHAPTER 1. Introduction 13

1.1. Background 13

1.2. Objectives and Research Issues 14

1.3. Structure of the Dissertation 14

CHAPTER 2. Methodology 15

2.1. Bibliometric Analysis 15

2.1.1. Applied Techniques 16

2.1.1.1. Citation analysis 16

2.1.1.2. Co-citation analysis 16

2.1.1.3. Keyword co-occurrence analysis 16

2.1.1.4. Clustering 17

2.1.1.5. Visualization 17

2.1.2. Chosen databases 17

2.1.2.1. Web of Science 17

2.1.2.2. Scopus 18

CHAPTER 3. Research Results 19

3.1. Descriptive Analysis: Main Findings about the Collection 21

3.1.1. Most influential Journals 24

3.1.2. Most influential authors 26

3.1.3. Most relevant contributions by Country 28

3.2. Intellectual structure of the field: Co-citation Analysis 30

3.2.1. Article Co-citation analysis 30

3.2.2. Journal Co-citation analysis 31

3.3. Conceptual structure of the field 32

3.3.1. Most Relevant Keywords: Keyword analysis 32

3.3.2. Co-word analysis through keyword co-occurrence 34

3.4. The social structure of the field – Collaboration analysis 36
3.4.1. Author collaboration network 36
3.4.2. Country collaboration network 37

CHAPTER 4. Discussion and Conclusion 39
4.1. Overview of Ethical Considerations in AI from 2019 to 2023 39
4.2. Most Influential Journals, Authors, and Countries in AI Ethics 40
4.3. Intellectual and Conceptual structure in AI Ethics 40
4.4. Social structure and Collaborative Networks in AI Ethics 40
4.5. Limitations 41
4.6. Implications and Future Research 41
4.7. Conclusion 42

CHAPTER 5. References 43

List of Figures

Figure 1. Annual Scientific Production 22
Figure 2. Core sources by Bradford's Law 24
Figure 3. Corresponding author's countries 28
Figure 4. Countries scientific production 29
Figure 5. Countries distribution by total citation count 29
Figure 6. Article co-citation network 31
Figure 7. Source co-citation network 32
Figure 8. Refined word cloud 34
Figure 9. Keyword co-occurrence network map 36
Figure 10. Author collaboration network map 37
Figure 11. Country Collaboration Map 38

List of Tables

Table 1. Data Retrieval Criteria 20
Table 2. Author Keywords 21
Table 3. Annual scientific production 22
Table 4. Documents with more than one hundred citations 23
Table 5. Source distribution by Bradford's Law 25
Table 6. Most relevant sources 25
Table 7. Distribution of authors per number of publications 26
Table 8. Most prolific authors within core sources 27
Table 9. Author keywords within core sources 33

CHAPTER 1

INTRODUCTION

1.1. Background

Artificial Intelligence (AI) has firmly established itself as one of the most transformative technologies (Jobin et al., 2019; Morley et al., 2020). In an Era defined by rapid technological advancements, the prominence of AI is unmistakable. AI has permeated every aspect of day-to-day lives, from the way we work and communicate to the way we make decisions (Floridi et al., 2020). The ubiquitous presence of AI-driven systems, from virtual assistants on smartphones to the automation of critical industrial processes, underscores the profound impact of AI on modern society.

The AI discussions are no longer confined to the technical and scientific community; they have split into the public domain (Tsamados et al., 2022). Conversations about AI's potential and ethical implications are prevalent in news headlines, boardrooms, legislative chambers, and living rooms (Hagendorff, 2020). As society grapples with the implications of this powerful technology, it has become abundantly clear that a thorough and critical examination of AI's ethical dimensions is imperative.

The pervasive presence of AI technologies, coupled with their potential to shape the future, demands a rigorous exploration of the ethical dimensions accompanying their use (Floridi et al., 2020). The ethical considerations surrounding AI are inextricably linked to real-world consequences (Ryan, 2020). The decisions made about AI, the implemented policies, and the established standards will have a profound impact on issues such as privacy, security, fairness, accountability, and the very nature of work itself.

Furthermore, AI's influence extends beyond the boundaries of a single discipline. It intersects with fields as diverse as philosophy, law, sociology, psychology, medicine, and economics (Coeckelbergh, 2020). This interdisciplinary nature underscores the need for a comprehensive review of AI ethics, informed by a broad spectrum of knowledge and expertise.

As discussions about AI ethics proliferate and evolve (Milano et al., 2020), it becomes increasingly apparent that a re-evaluation of the existing body of scientific literature is warranted (Robbins, 2019). AI ethics is a rapidly evolving field, and the dynamic interplay of technology and ethics necessitates a thorough and up-to-date understanding of the subject matter.

1.2. Objectives and Research Issues

This research aims to analyse the ethical aspects of artificial intelligence thoroughly. The main goal is to understand the ethical implications and considerations associated with artificial intelligence technologies by assessing the current state-of-the-art. Moreover, the research intends to identify gaps in existing research and highlight emerging trends. Aiming to answer the following question: What insights does bibliographic literature offer regarding ethical considerations in artificial intelligence from 2019 to 2023?

1.3. Structure of the Dissertation

This research is divided into four distinct sections or chapters, structured as follows:

Chapter 1, the current chapter, serves as the introduction and provides an overview of this research's background, context, and objectives, setting the foundation for the entire research.

Chapter 2 - Research Methodology: This chapter provides an in-depth exploration of the research methodology employed, encompassing bibliometric analysis. It offers insights into the techniques used, databases accessed, and the rationale underpinning their selection.

Chapter 3 - Research Findings: This section presents the findings, commencing with a descriptive analysis of the research field's landscape. It then delves into examining the intellectual and conceptual structures within the field, alongside exploring collaboration analysis within the same domain.

Discussions and Conclusion: Chapter 4 engages in detailed discussions and thorough analysis of the research results presented in Chapter 3. It offers valuable insights into the significance of the findings and their broader implications and, finally, provides a summary of the key research findings, acknowledges any limitations, draws overall conclusions, and offers recommendations for future research endeavours or practical applications.

CHAPTER 2

METHODOLOGY

A systematic approach is essential to achieve a comprehensive understanding of the critical research areas, primary concerns, trends, and the field evolution. Due to the extensive scope of the review and the large collection, manual review could be more practical. Consequently, bibliometric mapping techniques were chosen to provide an objective and comprehensive overview of this field's structure.

2.1. Bibliometric Analysis

Bibliometric analysis is a robust and widely employed method for dissecting vast scientific data. It allows one to uncover the developmental subtleties within a particular field and highlight emerging areas within it (Donthu et al., 2021).

Bibliometric analysis is a valuable approach for gauging, monitoring, and exploring scientific outputs. It aids in visualizing and broadening knowledge within a specific research domain, revealing relationships among key publications, authors, institutions, themes, and other attributes in the area under investigation (José de Oliveira et al., 2019). This method unravels and visualizes the collective scientific knowledge and evolutionary intricacies within established domains, providing a solid foundation for propelling the field forward in innovative and meaningful ways. Scholars benefit by gaining a comprehensive overview, identifying knowledge gaps, generating new research paths, and positioning their intended contributions within the field (Donthu et al., 2021).

The application of bibliometric analysis aims to map the current state of a scientific theme by characterizing bibliometric parameters. These parameters, readily available in scientific research platforms, are utilized to provide a comprehensive understanding of the field. Importantly, this method possesses a universal applicability across diverse fields of knowledge. Researchers are encouraged to tailor and adapt the method according to their specific requirements and objectives (José de Oliveira et al., 2019).

Donthu et al. (2021) divide bibliometric analysis into two main types: Performance analysis and science mapping. Performance analysis examines how researchers contribute to a field, while science mapping examines how these contributions relate. In addition to these fundamental techniques, Donthu et al. (2021) introduce complementary techniques to enhance the outcomes of bibliometric studies. They propose three enrichment pathways centred around network analysis: network metrics, clustering, and visualization.

This research primarily focuses on science mapping and leverages the suggested enrichment techniques. However, performance analysis techniques will also be employed. Specifically, co-citation analysis, keyword occurrence analysis and citation analysis will be utilized, and these techniques will be further elaborated in the following sections.

2.1.1. Applied Techniques

For this research, the chosen methodology involves applying citation analysis, co-citation analysis, and co-word analysis. This approach helps illustrate the structure of the field in terms of conceptual elements, recent influential contributions, and the intellectual structure of the discipline (Donthu et al., 2021). Employing this combination will offer a comprehensive view of the current ethical landscape within artificial intelligence research.

2.1.1.1. Citation analysis

Citation analysis involves tallying the number of citations an article has accumulated from the publication date to the date of bibliographic data retrieval. This analysis gauges the impact of a publication based on its citation count. By employing citations, one can scrutinize the highly influential publications in a research field, offering insights into the intellectual dynamics of that domain (Donthu et al., 2021).

2.1.1.2. Co-citation analysis

In a co-citation analysis, two publications are linked when they appear together in the reference list of another publication. The advantage of employing co-citation analysis is that, alongside identifying the most influential publications, researchers can unveil thematic clusters based on these cited works. This method focuses specifically on highly cited publications, making it suitable for researchers to pinpoint seminal publications and foundational knowledge within a field (Donthu et al., 2021).

2.1.1.3. Keyword co-occurrence analysis

Like co-citation analysis, co-word analyses operate under the assumption that words frequently used together share a thematic relationship. This analytical method enriches the understanding of past and present analyses. It provides a glimpse into the future of the research field, identifying two keywords as co-occurring if they are used together in describing a single document. Moreover, it can predict future research trends by considering significant keywords from a publication's implications and future research directions (Donthu et al., 2021; Eck & Waltman, 2009).

2.1.1.4. Clustering

Clustering groups related themes or social entities based on the analysis type. It helps to grasp how a research field unfolds and evolves. For instance, thematic clusters from co-citation analysis or bibliographic coupling reveal vital themes and their evolution in the field (Donthu et al., 2021; Linnenluecke et al., 2020).

2.1.1.5. Visualization

Bibliometric analysis often involves network visualization software, which varies from user-friendly graphical tools like VOSviewer to command-driven options such as Bibliometrix. When choosing bibliometric software or a combination for analysis and visualization, it is essential to weigh the software features and the flexibility of the resulting network (Moral-Muñoz et al., 2020). For this research, the Bibliometrix package in R, supplemented with Biblioshiny, was specifically chosen.

2.1.2. Chosen databases

Selecting a suitable bibliographic database is a crucial step with a substantial impact on bibliometric analysis. Careful planning is essential to obtain accurate results and avoid unnecessary revisions (Zhu & Liu, 2020). Combining multiple platforms for extracting scientific data can lead to a more comprehensive bibliometric analysis despite the challenges of integrating data from different databases (José de Oliveira et al., 2019). Once this research aims to provide a broad and comprehensive portrayal of the ethical aspects of Artificial Intelligence, two databases were chosen: Web of Science (WoS) and Scopus, the two primary and most comprehensive sources of publication metadata and impact indicators (Pranckutė, 2021).

2.1.2.1. Web of Science

WoS, initially named Web of Knowledge, blazed the trail as the first bibliographic database. Its origins are Eugene Garfield's pioneering work in the 1960s at the Institute for Scientific Information (ISI). After being acquired by Thompson Reuters in 1992, ISI assumed its current moniker - WoS and, in 2016, came under the ownership of Clarivate Analytics (Pranckutė, 2021).

WoS represents a broad-based and judiciously curated database featuring specialized indexes organized by content type or thematic focus. At its core lies the WoS Core Collection (WoS CC), encompassing six principal citation indexes: (1) Science Citation Index Expanded (SCIE); (2) Social Sciences Citation Index (SSCI); (3) Arts & Humanities Citation Index (A&HCI); (4) Conference Proceedings Citation Index (CPCI); (5) Books Citation Index (BKCI); and (6) the Emerging Sources Citation Index (ESCI) (Pranckutė, 2021).

2.1.2.2. Scopus

For over four decades, WoS stood as the exclusive provider of bibliographic data until 2004, when Elsevier launched Scopus, introducing a significant alternative. As time progressed, Scopus emerged as a robust bibliographic data source, earning its position alongside WoS and occasionally surpassing it in reliability (Pranckutė, 2021). It has been consistently demonstrated that Scopus offers broader coverage compared to WoS CC. This holds for both early and recent content coverage analysis. Scopus stands out for indexing a more extensive set of distinctive sources that are not encompassed by WoS (Pranckutė, 2021).

CHAPTER 3

RESEARCH RESULTS

Following the explanation of the research's methodology, this section outlines how research results were determined and bibliometric maps were created. Identifying relevant articles published in the last five years involved conducting advanced keyword searches in WoS and Scopus.

In this research, the search terms "artificial," "intelligence," and "ethics" were deliberately chosen to establish a precise and focused search strategy. These terms were selected based on their core relevance to artificial intelligence ethics.

The decision not to include synonyms or synonymous expressions was made based on a quick review of existing literature and a preliminary search. For instance, when conducting an initial search using the term "Artificial Intelligence ethics," it was observed that many articles using synonymous expressions were retrieved. However, upon closer examination, it was observed that these articles often covered similar topics and themes, making it redundant to include all synonymous expressions.

For example, an article titled "Ethical Considerations in Artificial Intelligence" was found in the initial search. Subsequently, a search with the term "ethical technology" returned articles discussing similar ethical issues in the context of AI. In such cases, including both "Artificial intelligence ethics" and "ethical technology" would result in duplicate content and potentially obscure the central focus of the research. Moreover, while synonymous expressions such as "ethical technology" and "AI ethic" are relevant, it is essential to note that including every possible synonym in the search query could lead to an extensive and potentially unwieldy dataset. The decision to use "ethics" as the primary keyword is justified by the fact that it is a widely recognized and comprehensive term within AI research.

The words "ethics", "artificial", and "intelligence" were combined through the Boolean operator AND and applied to the titles, abstracts, and keywords of the documents.

The primary sources considered to ensure the reliability of the literature review are articles and peer-reviewed reviews published in journals, given their rigorous peer-review process. In line with this criterion, other document types, such as conference papers, notes, letters, books, book chapters, editorials, doctoral theses, master's dissertations, and non-scientific publications, were excluded from the research findings, as advised by José de Oliveira et al., (2019). Additionally, the scope was narrowed to include only documents published in English within the knowledge domains of computer science, business economics, business management and accounting, government law, and science and technology, aligning with the research's specific focus.

The data retrieval criteria are in Table 1, including the specific queries used to obtain the dataset from both WoS and Scopus.

Table 1. Data Retrieval Criteria

Parameters	Inclusion criteria	Exclusion criteria
Database	Scopus and WoS	Other databases
Publication period	Between 2019 and 2023	Older publications
Document type	Article and review article	Other document types
Research area (WoS)	Computer Science, Business economics, Science Technology Other Topics, Government Law	Other research areas
Subject area (Scopus)	Computer Science, Business Management and accounting	Other subject areas
Language	English	Other languages
Query (WoS)	TS=(ethics AND artificial AND intelligence) AND PY=(2019-2023) AND DT=(Article OR Review) AND LA=(English) AND SU=(Computer Science OR Business Economics OR Government Law OR Science Technology Other Topics) [1]	-
Query (Scopus)	TITLE-ABS-KEY (ethics AND artificial AND intelligence) AND PUBYEAR > 2018 AND PUBYEAR < 2024 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "BUSI")) AND (LIMIT-TO (LANGUAGE , "English"))	-

Source: Self-elaborated

A total of 697 articles from WoS and 845 from Scopus were collected. Following removing of 443 duplicate articles, a dataset of 1099 unique articles was obtained. This dataset was then consolidated and analysed using Microsoft Excel and Bibliometrix to create comprehensive bibliometric maps.

To evaluate the potential for augmenting the initial search with additional keywords, an analysis of author-provided keywords was conducted on the retrieved 1099 articles. As a foundational criterion, any keyword present in at least 10% of the entire dataset was considered. Table 2 presents the findings, indicating that 'machine learning' met the relevance threshold. However, given that 'machine learning' is a subset of artificial intelligence, encompassing various other branches, it was not included in the initial keyword search.

[1] TS (topic), PY (Publication year), DT (document type) SU (research area)

Table 2. Author Keywords

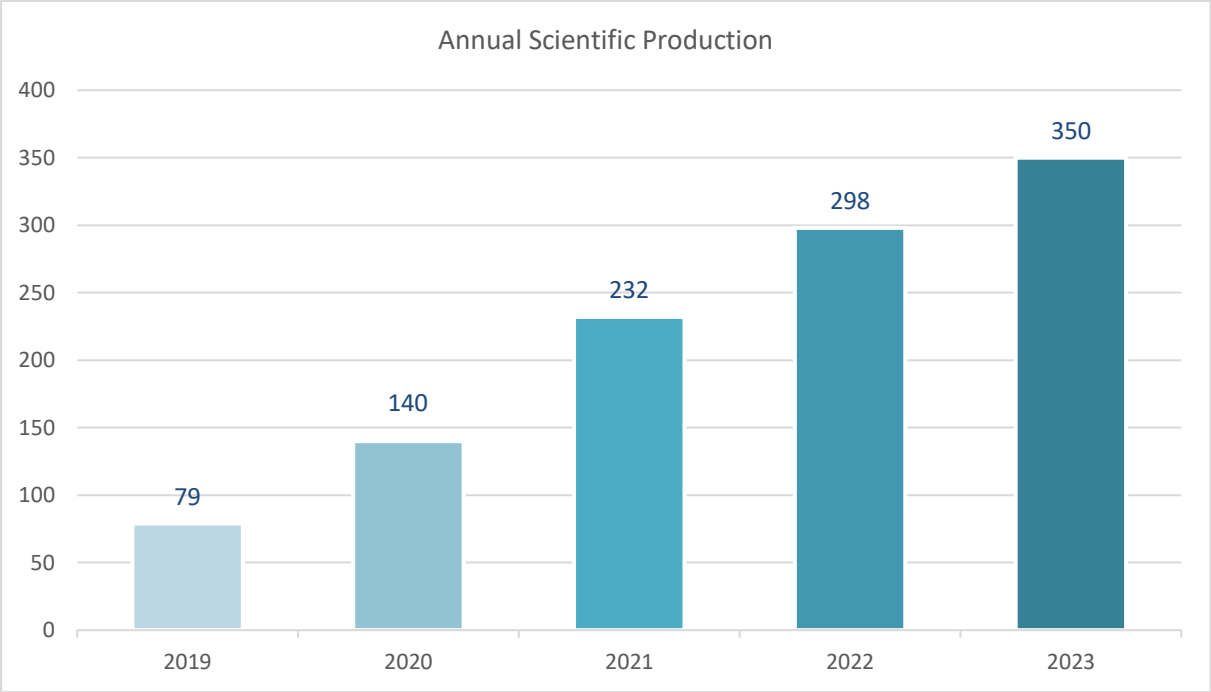
Words	Occurrences	Percentage
artificial intelligence	599	55
ethics	382	35
ai ethics	105	10
machine learning	105	10
ai	62	6
artificial intelligence (ai)	55	5
privacy	41	4
governance	36	3
machine ethics	35	3
fairness	34	3

Source: Self-elaborated

3.1. Descriptive Analysis: Main Findings about the Collection

With the proliferation of artificial intelligence technology and the popularity of generative AI tools like ChatGPT, an upward trend is expected in the annual scientific production of papers addressing concerns, particularly ethical considerations related to this technology, over the past five years. As anticipated, the data substantiates this expectation, revealing a clear and consistent rise. Since 2019, the extant scientific documentation in this research collection has grown from 79 to 1099 documents, signifying an impressive 1391% increase (with an annual growth rate of 45%). This trend is visually depicted in Figure 1, illustrating the number of documents over the past five years.

Figure 1. Annual Scientific Production



Source: Self-elaborated

Moreover, table 3 illustrates the distribution of documents by year, effectively highlighting the significant increase as the years advance. Notably, in 2019, a mere 7% of the entire dataset is accounted for, while 2023, encompassing data up to October, represents 32% of the total. This data reinforces the continuing growth in scientific output, thereby substantiating the pertinence of this research in comprehending the current state of the art in the field of ethics in artificial intelligence.

Table 3. Annual scientific production

Years	Documents	Percentage of 1099
2019	79	7%
2020	140	13%
2021	232	21%
2022	298	27%
2023	350	32%

Source: Self-elaborated

Regarding the documents, the average number of citations per document stands at 11.5. Additionally, 2899 author-assigned keywords have been identified, and these documents originate from 435 distinct sources. Table 4 presents the top articles with 100 or more citations.

Table 4. Documents with more than one hundred citations

Document	DOI	TC	TCPY	NTC
JOBIN A, 2019, NAT MACH INTELL The global landscape of AI ethics guidelines	10.1038/s42256-019-0088-2	797	159.4	21.79
DAVENPORT T, 2020, J ACAD MARKET SCI How artificial intelligence will change the future of marketing	10.1007/s11747-019-00696-0	449	112.25	15.28
HAGENDORFF T, 2020, MIND MACH Ethics of AI Ethics: Na Evaluation of Guidelines	10.1007/s11023-020-09517-8	332	83	11.3
AMANN J, 2020, BMC MED INFORMATICS DECIS MAK Explainability for artificial intelligence in healthcare	10.1186/s12911-020-01332-6	308	77	10.48
MITTELSTADT B, 2019, NAT MACH INTELL Principles alone cannot guarantee ethical AI	10.1038/s42256-019-0114-4	242	48.4	6.62
TAMBE P, 2019, CALIF MANAGE REV Artificial Intelligence in Human Resources Management	10.1177/0008125619867910	214	42.8	5.85
DWIVEDI YK, 2023, INT J INF MANAGE “So what if ChatGPT wrote it?”	10.1016/j.ijinfomgt.2023.102642	169	169	52.67
MORLEY J, 2020, SCI ENG ETHICS From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices	10.1007/s11948-019-00165-5	160	40	5.44
MARTIN K, 2019, J BUS ETHICS Ethical Implications and Accountability of Algorithms	10.1007/s10551-018-3921-3	154	30.8	4.21
O'SULLIVAN S, 2019, INT J MED ROB COMPUT ASSISTED SURG Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery	10.1002/rcs.1968	146	29.2	3.99
REDDY S, 2020, J AM MED INFORM ASSN A governance model for the application of AI in health care	10.1093/jamia/ocz192	142	35.5	4.83
STEPHANIDIS C, 2019, INT J HUM-COMPUT INT Seven HCI Grand Challenges	10.1080/10447318.2019.1619259	142	28.4	3.88
KAPLAN A, 2020, BUS HORIZONS Rulers of the world, unite! The challenges and opportunities of artificial intelligence	10.1016/j.bushor.2019.09.003	128	32	4.35
FELZMANN H, 2019, BIG DATA SOC Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and textual concerns	10.1177/2053951719860542	111	22.2	3.04
HANCOCK JT, 2020, J COMPUTER-MEDIATED COMMUN AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations	10.1093/jcmc/zmz022	111	27.75	3.78
WINFIELD AF, 2019, PROC IEEE Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems	10.1109/JPROC.2019.2900622	105	21	2.87
THIEBES S, 2021, ELECTRON MARK Trustworthy artificial intelligence	10.1007/s12525-020-00441-4	104	34.67	8.58

Source: Self-elaborated

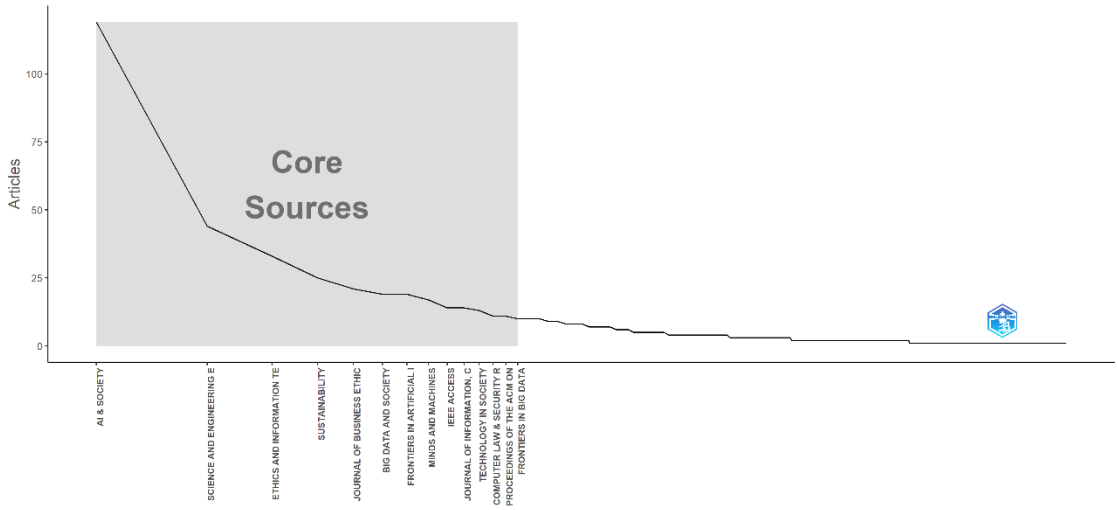
Regarding the authors, there are 3218 unique authors. Among them, 280 have authored single-authored documents. On average, each document is co-authored by 3.42 individuals, and the international co-authorship rate stands at 18.2%.

These dimensions will be explored in greater detail in the subsequent sections.

3.1.1. Most influential Journals

After acquiring the dataset of 1099 documents for analysis, a pivotal step involved identifying the most pertinent journals within the field. This was accomplished using Bradford's law, a method which posits that when scientific journals are organized by decreasing productivity of articles on a given subject, they can be classified into a core nucleus of periodicals that are mainly dedicated to the subject and several outer zones containing an equal number of articles as the nucleus (Onyancha & Ocholla, 2022). This pattern follows a 1:n:n² distribution and can be observed in Figure 2.

Figure 2. Core sources by Bradford's Law



Source: Bibliometric Biblioshiny application

Bradford's law, first described by Samuel C. Bradford in 1934, is a helpful model for estimating the diminishing returns associated with searching for references in scientific journals. In this context, when journals in a specific field are sorted by the number of articles, they can be grouped into three zones, each comprising roughly one-third of all articles (Onyancha & Ocholla, 2022). As detailed in Table 5, Zone 1 encompasses 14 journals, representing only 3% of the total but housing a substantial 34% of the dataset.

Table 5. Source distribution by Bradford's Law

Zone	# Journals	% Journals	# Articles	% Articles
1	14	3%	370	34%
2	104	24%	368	33%
3	317	73%	361	33%
Total	**435**	**100%**	**1099**	**100%**

Source: Self-elaborated

These journals were selected based on their production of the highest number of documents pertinent to this subject area. Notably, the most influential journal in this domain is 'AI & SOCIETY,' which has published 119 documents. It is followed by 'SCIENCE AND ENGINEERING ETHICS,' with 44 documents, and 'ETHICS AND INFORMATION TECHNOLOGY,' with 33 documents.

A comprehensive list of these influential journals can be found in Table 6, highlighting their essential role in the research methodology and underscoring their significance in establishing the foundation for subsequent analysis.

Table 6. Most relevant sources

Sources	Documents	% of 1099
AI & SOCIETY	119	11%
SCIENCE AND ENGINEERING ETHICS	44	4%
ETHICS AND INFORMATION TECHNOLOGY	33	3%
SUSTAINABILITY	25	2%
JOURNAL OF BUSINESS ETHICS	21	2%
BIG DATA AND SOCIETY	19	2%
FRONTIERS IN ARTIFICIAL INTELLIGENCE	19	2%
MINDS AND MACHINES	17	2%
IEEE ACCESS	14	1%
JOURNAL OF INFO, COMMUNICATION AND ETHICS IN SOCIETY	14	1%
TECHNOLOGY IN SOCIETY	13	1%
COMPUTER LAW & SECURITY REVIEW	11	1%
PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION	11	1%
FRONTIERS IN BIG DATA	10	1%
TOTAL	**370**	**34%**

Source: Self-elaborated

After identifying the core journals, it is necessary to extract all documents from the initial dataset published in any selected journals. This process resulted in the creating of a new dataset consisting of 370 documents. This collection now serves as the primary dataset for this research going forward.

3.1.2. Most influential authors

Upon narrowing the analysis to the 370 documents encompassed by the core sources identified through Bradford's law, the dataset comprises 840 unique authors. Among them, 110 authors are single-authored, and international co-authorship is present in 13.78% of cases. On average, there are 2.65 co-authors per document.

It is noteworthy that most authors contributed only one article, accounting for 761 authors, which constitutes 90.6% of the total dataset. Furthermore, 98.7% of authors, totalling 829 individuals, have authored fewer than four articles. Table 7 shows a detailed breakdown of author publication frequency.

Table 7. Distribution of authors per number of publications

# documents	# Authors	% of 840 authors
1	761	90.6%
2	54	6.4%
3	14	1.7%
4	6	0.7%
5	1	0.1%
>5	4	0.5%

Source: Self-elaborated

After conducting a comprehensive quantitative analysis of the authorship landscape, the next crucial step is to identify the most significant authors within the research area central to this research. Luciano Floridi emerges as the most prolific author, having contributed 16 publications to the most pertinent sources. Following is Mariarosaria Taddeo, with 11 contributions, and Mark Ryan, with 8.

For a detailed breakdown of authors who have published more than three documents in the core 14 journals, please refer to Table 8.

Table 8. Most prolific authors within core sources

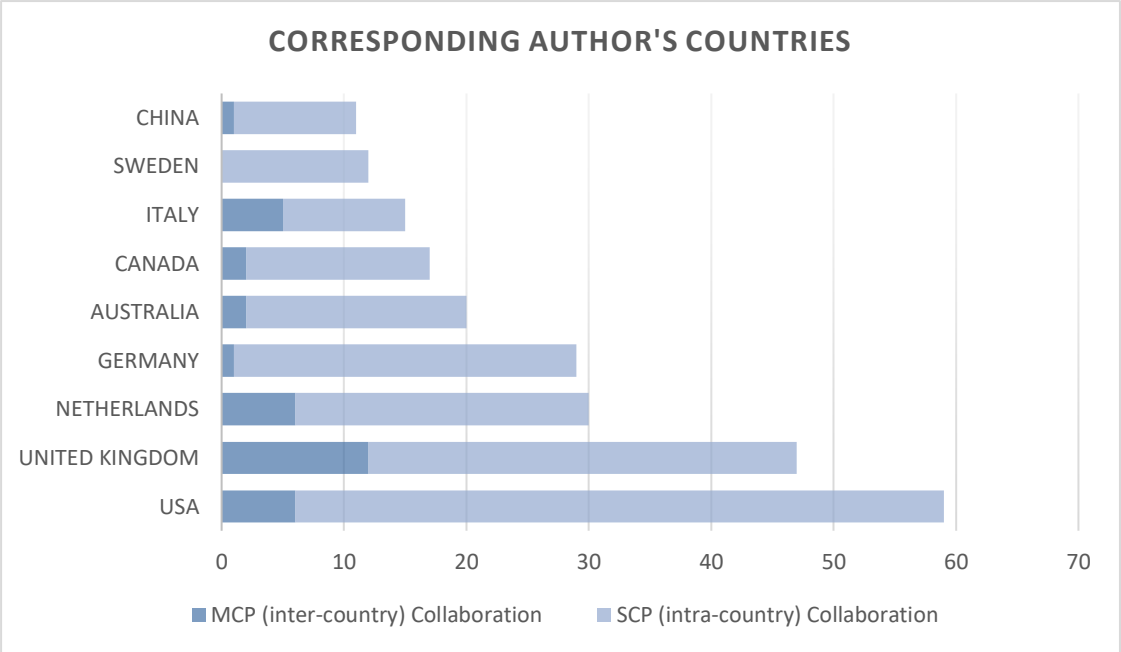
Authors	# Articles	Fractionalized [2]	Affiliation
Luciano FLORIDI	16	3.94	University of Oxford
Mariarosaria TADDEO	11	2.61	The Alan Turing Institute
Mark RYAN	8	4.62	Wageningen University
Jessica MORLEY	6	1.14	University of Oxford
Josh COWLS	5	0.91	University of Oxford
Nazanin ANDALIBI	4	1.5	University of Michigan
Mark COECKELBERGH	4	2.5	Universität Wien
Jacob MÖKANDER	4	1.17	Princeton University
Scott ROBBINS	4	3	Delft University of Technology
Bernd STAHL	4	1.12	University of Nottingham

Source: Self-elaborated

The results reveal that the authors are primarily associated with the United States of America (USA), making it the most significant contributor with 59 published documents. Following is the United Kingdom (UK) with 47 documents, the Netherlands with 30, and Germany with 29 documents. This geographical distribution closely aligns with the concentration of the most prolific authors. Notably, the UK has five of the most relevant authors, the USA and the Netherlands have two each, and Germany has 1 (please refer to Figure 3).

[2] Fractional authorship: Indicate the relative contribution of each author to the published works. A value of 1 represents the standard contribution of a sole author, while values greater than 1 signify that the author has made a more substantial contribution than a typical co-author. This system helps quantify the depth of each author's involvement in collaborative research or publication projects.

Figure 3. Corresponding author's countries



Source: Self-elaborated, based on Bibliometrix Biblioshiny applications output

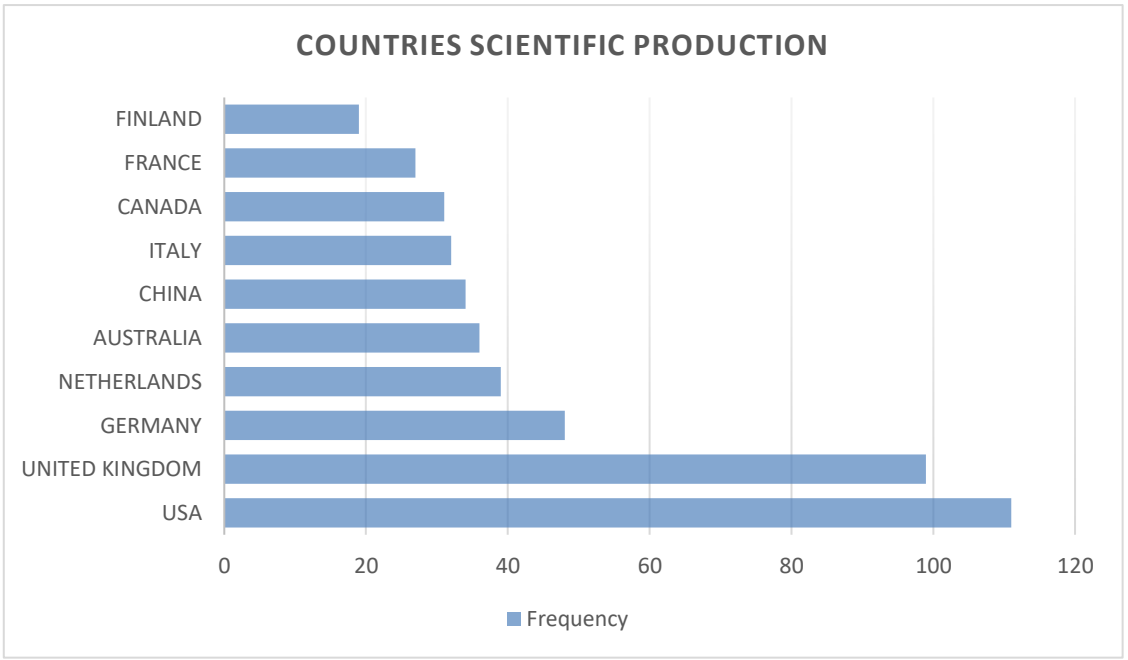
3.1.3. Most relevant contributions by Country

Subsequently, an analysis of scientific production by Country was conducted to discern which countries are the foremost contributors to the advancement of the field. As depicted in Figure 4, the USA is the leading contributor with 111 publications, followed by the UK with 99 documents, Germany with 48, and the Netherlands with 39.

In a citation-focused analysis on a per-country basis, visible in Figure 5, the UK emerges as the most cited Country despite not being the most significant contributor in the number of publications. The UK garners a remarkable 1026 citations, averaging 21.8 citations per article, followed by the USA, which accumulates 572 citations and boasts an average of 9.7 citations per article. Germany, with 572 citations and an average of 19.2 citations per article, and the Netherlands, with 513 citations and 17.1 citations per article, also feature prominently.

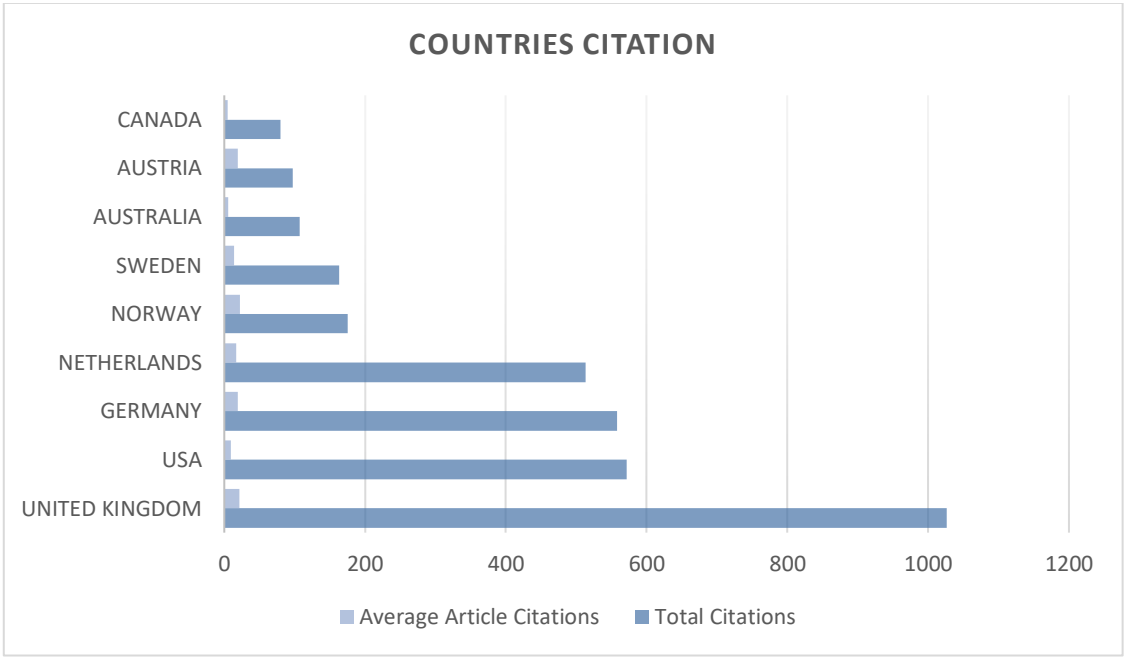
It is worth noting that while the top four most productive countries also rank among the most cited, the order of productivity does not mirror the order of citation impact, in which the UK is notably the most impactful.

Figure 4. Countries scientific production



Source: Self-elaborated, based on Bibliometrix Biblioshiny applications output

Figure 5. Countries distribution by total citation count



Source: Self-elaborated, based on Bibliometrix Biblioshiny applications output

With a comprehensive descriptive analysis of the research field, encompassing key findings that offer a clear and well-defined characterization of the dataset, the next phase involves constructing the field's intellectual structure through co-citation analysis and the conceptual structure through co-word analysis. Following these steps, the social structure will be analysed, including author and Country collaboration networks.

3.2. Intellectual structure of the field: Co-citation Analysis

3.2.1. Article Co-citation analysis

The co-citation network that shows relations between cited-reference works is divided into two clusters and shown in Figure 6.

The red cluster is centred on earlier works and authors in the field, possibly as the foundational groundwork for discussions and research within this domain. Notable articles and authors in this cluster include (Floridi & Sanders, 2004), (Moor, 2006), and (Coeckelbergh, 2010), who hold the highest betweenness, closeness, and PageRank values, indicating their substantial influence and frequent co-citation. Additionally, (SPARROW, 2007) demonstrates a significant impact regarding closeness and PageRank, signifying its pivotal role within the co-citation network.

In contrast, the blue cluster encompasses a more contemporary and diverse array of authors and articles, likely representing recent developments and discussions in the field.

Jobin et al. (2019) emerge as the central node in this cluster, boasting the highest betweenness, closeness, and PageRank values, underscoring its critical reference status within the co-citation network. Furthermore, Floridi et al. (2018), Hagendorff (2020), Mittelstadt (2019), and Floridi & Cowls (2019) exhibit substantial influence with high PageRank values, highlighting their significance as references within this cluster. In contrast, Awad et al. (2018) demonstrate lower influence in PageRank, indicating less frequent citations than more influential articles in this cluster.

Figure 6. Article co-citation network

Source: Bibliometrix Biblioshiny application

3.2.2. Journal Co-citation analysis

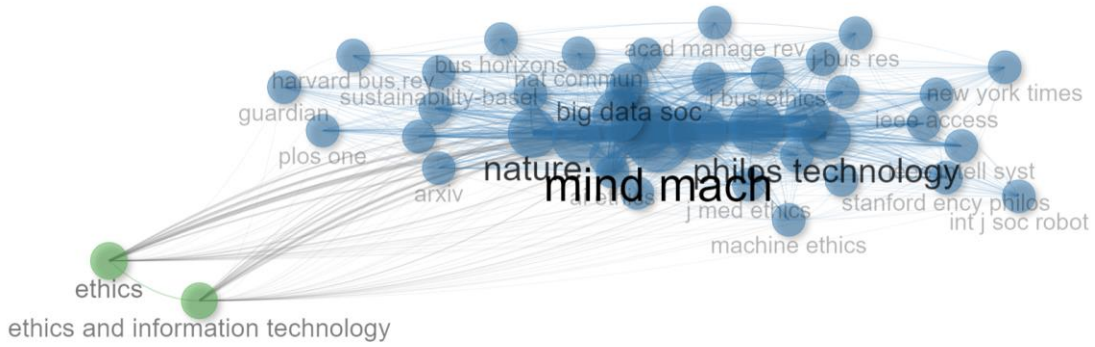
Journal co-citation analysis, displayed in Figure 7, is a valuable tool for uncovering key themes, trends, and influential works within a specific research domain. It aids in understanding the intellectual structure and connections within the field.

The data covers the period from 2011 to 2021. It is worth noting that older publications, specifically those from 2011 to 2013, exhibit lower influence in terms of betweenness, closeness, and PageRank. In contrast, newer publications, such as those from 2020 and 2021, have lower betweenness values, potentially indicating emerging research areas.

Journals such as "Science and Engineering Ethics," "Ethics and Information Technology," "Minds and Machines," and "AI & Society" hold significant influence within this field, as evidenced by their substantial betweenness, closeness, and PageRank values. This indicates that research published in these journals is frequently cited and interconnected.

The data is categorized into two clusters, identified with the colours blue and green. It is evident that the blue cluster, which encompasses sources like "Science and Engineering Ethics" and "Minds and Machines," boasts higher betweenness, closeness, and PageRank values. This suggests that journals within the blue cluster wield more significant influence within the co-citation network.

Figure 7. Source co-citation network



Source: Bibliometrix Biblioshiny application

3.3. Conceptual structure of the field

The objective of co-word analysis is to chart the conceptual structure of a research field by examining word co-occurrences within a bibliographic collection. The initial approach involved analysing author-assigned keywords, allowing a deeper understanding of the significant research topics, trends, and themes from the last five years, and constructing a subsequent keyword co-occurrence network map.

3.3.1. Most Relevant Keywords: Keyword analysis

To achieve more pertinent results and mitigate possible repeated patterns, Bibliometrix was directed to consider the following sets of words as synonyms:

a) 'Artificial intelligence' 'artificial-intelligence' 'AI' 'artificial intelligence technologies' 'artificial intelligence (AI)' 'artificial intelligence systems' 'AI systems'
b) 'Machine-learning' 'machine learning'
c) 'Algorithm' 'algorithms' 'algorithmics'
d) 'decision-making' 'decision making'
e) 'Ethics' 'ethic' 'AI ethic' 'ethical technology' 'machine ethics' and 'artificial intelligence ethic'.

While the results appear to align with the patterns observed in the initial output derived from the broader dataset, the refined analysis underscores and adds that the literature predominantly emphasizes topics related to governance, fairness, privacy, responsibility, transparency, and sustainability in the context of artificial intelligence as reflected in table 9.

To further complete the analysis, a subsequent refinement was performed, instructing the software to exclude the terms 'artificial intelligence,' 'ethics,' and 'machine learning.' This refinement generated a word cloud, depicted in Figure 8, which further underscores that the central research areas encompass those mentioned earlier, alongside topics of bias (9 occurrences) and accountability (8 occurrences) in artificial intelligence.

Table 9. Author keywords within core sources

Author Keywords	# Occurrences
artificial intelligence	275
ethics	207
machine-learning	44
governance	17
fairness	15
privacy	14
responsibility	14
transparency	13
algorithm	10
sustainability	10

Source: Self-elaborated

Figure 8. Refined word cloud

Source: Bibliometrix Biblioshiny application

3.3.2. Co-word analysis through keyword co-occurrence

The keyword co-occurrence network map depicted in Figure 9 considers only author-assigned keywords and the previously established synonym rule. This network map comprises five clusters, each represented by a distinct colour. Node size reflects keyword frequency, with the most significant nodes represented by circles denoting the most impactful terms. Distances between nodes signify the strength of their relationship, and thicker lines indicate stronger links between items.

The network map presents five clusters, each identified by a unique colour:

a) The red cluster includes keywords such as artificial intelligence, algorithmic decision-making, emotion recognition, and automation, indicating solid associations between these terms.

b) The purple cluster encompasses words related to ethics, responsible AI, sustainability, and AI governance, implying thematic connections among these concepts.

c) The green cluster includes machine learning, digital ethics, responsibility, accountability, transparency, fairness, privacy, and bias.

d) The yellow cluster emphasizes governance and regulation, while the blue cluster centres on applied and business ethics.

The red cluster, highlighting artificial intelligence and technology-related terms, exhibits close ties to the purple cluster, which focuses on ethics and responsible AI, suggesting a robust link between AI and ethical considerations.

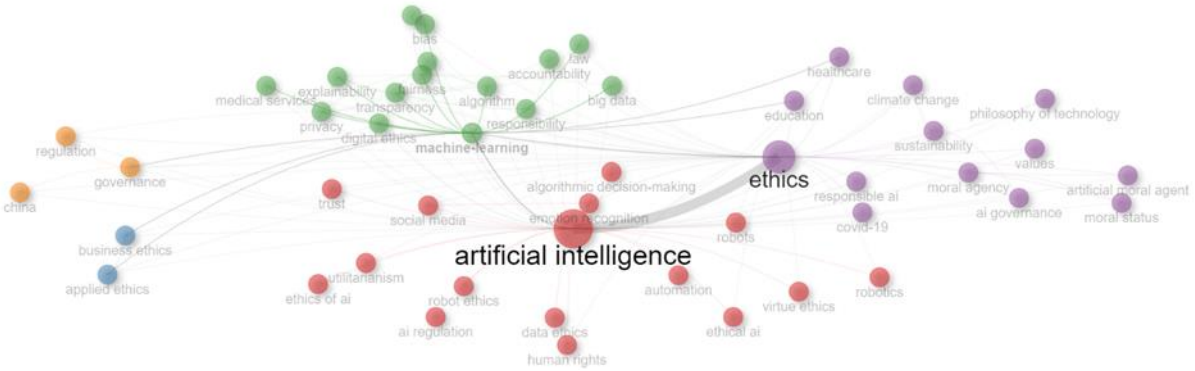
The green cluster, comprising machine learning and fairness-related terms, also connects with the red and purple clusters, indicating the convergence of machine learning and ethics. In contrast, the yellow cluster, with keywords such as "governance" and "regulation," appears to have weaker connections with the other clusters, suggesting a potential separation of governance concerns from ethical and technological domains.

The blue cluster, featuring applied and business ethics, links to the green cluster (machine learning) and the yellow cluster (governance), hinting at the relevance of ethics in business and practical contexts. The network map underscores the strong relationship between AI/technology and ethics, with "ethics" as a central node, signifying the crucial role of ethical considerations in AI development and deployment.

Inter-cluster connections suggest opportunities for interdisciplinary discussions and collaborations. For example, discussions on "sustainability" and "climate change" within the purple cluster may involve experts from both technological and ethical domains. The weaker links of the yellow and blue clusters with other clusters might indicate a degree of separation between governance and businesses and the broader ethical or technological considerations in AI. This separation could be due to the swift growth of the AI field, which may not have allowed sufficient time for these areas to establish stronger connections.

Morley et al. (2020) emphasize that the discourse on the ethical implications of AI has primarily centred on principles – the 'what' of AI ethics, encompassing concepts like beneficence, non-maleficence, autonomy, justice, and explicability. However, there has been less emphasis on the 'how' of implementing these ethical principles in practice. This dynamic reflects a situation where awareness of potential issues is rapidly increasing. However, the ability of the AI community to take effective action to mitigate associated risks is still in its early stages.

Figure 9. Keyword co-occurrence network map



Source: Bibliometrix Biblioshiny application

In summary, the co-occurrence network map illuminates the intricate interplay between technology, ethics, governance, and regulation in artificial intelligence. It highlights prospects for further exploration and collaboration in these domains.

3.4. The social structure of the field – Collaboration analysis

3.4.1. Author collaboration network

The author collaboration network map comprises multiple clusters or groups of authors engaged in collaborative research. Each cluster has a unique number and colour, as visible in Figure 10.

Notable contributors include:

a) Floridi L. is part of Cluster 1 and exhibits a substantial "Betweenness" value, implying a central role in connecting various authors or groups within the network.

b) Taddeo M., situated in Cluster 2, demonstrates a high "PageRank," signifying significant influence within that cluster.

c) Ryan M., a member of Cluster 3, possesses a notable "Closeness" value, indicating efficient information flow within that cluster.

d) Within Cluster 5, authors like Andalibi N., Roemmich K., and others showcase a "Closeness" value of 1, underscoring their strong interconnections within the cluster.

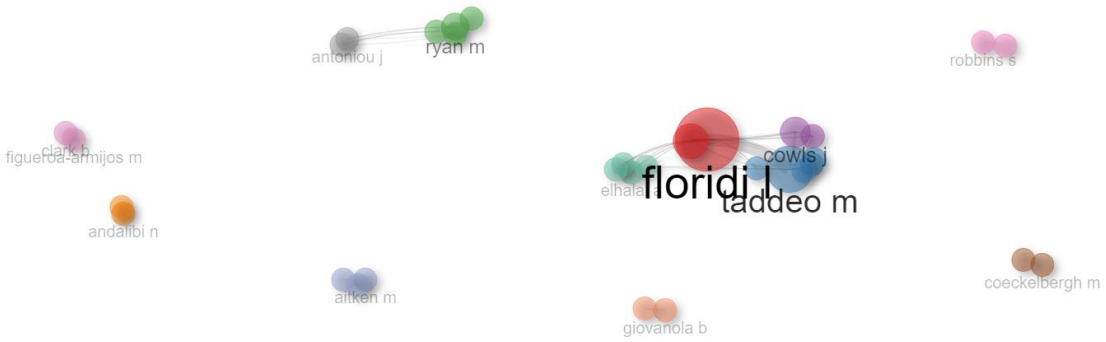
Some authors exhibit low values in all three metrics, such as Formosa P. and Garcia F., possibly indicating their involvement in smaller, closely-knit groups within their respective clusters.

On the other hand, some authors, like Jiya T. in Cluster 3, have a "Betweenness" of 0, suggesting a relatively isolated position within their cluster.

In Cluster 9, Ziosi M. holds a high "PageRank" but not the highest "Betweenness.", which suggests their significant influence within the cluster without necessarily serving as a bridge to other clusters.

The collaboration map reflects robust connections among authors from the same geographical regions. For instance, authors affiliated with the University of Oxford (UK) appear to collaborate closely, aligning with the substantial representation of UK authors within the dataset.

Figure 10. Author collaboration network map



Source: Bibliometrix Biblioshiny application

3.4.2. Country collaboration network

The Country collaboration network depicted in Figure 11 illustrates various international collaborations, highlighting joint research efforts in the field. For instance, the USA collaborates with multiple countries, with the United Kingdom serving as its most frequent collaborator, followed by Canada and Italy. This extensive network reflects the USA's strong global presence, with collaborations spanning regions such as Europe, Asia, and Australia.

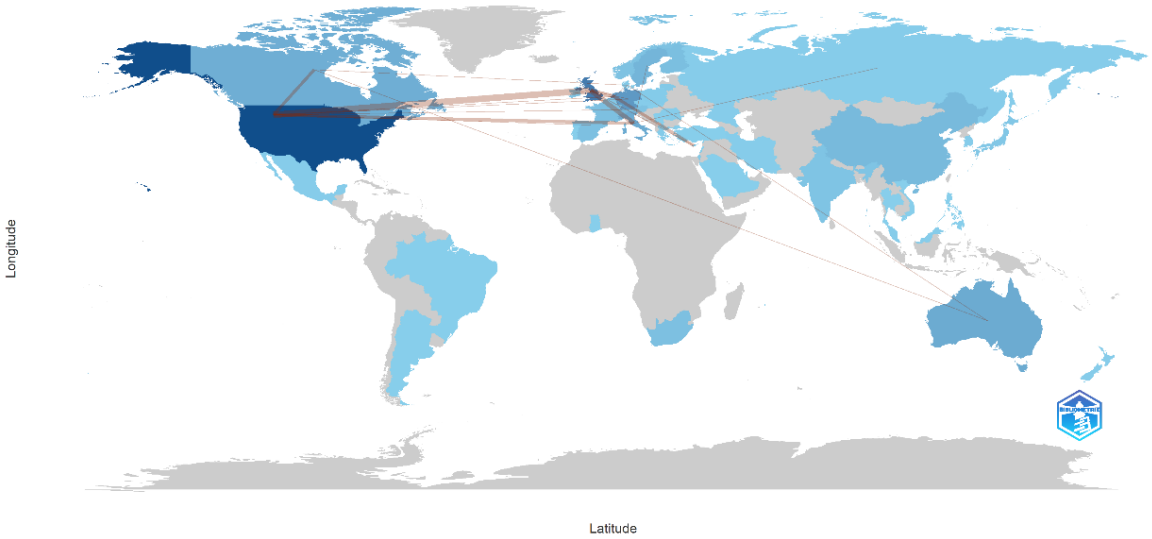
The United Kingdom also boasts extensive collaborations with several countries, including Italy, the United States, and the Netherlands, reinforcing its central role in international research partnerships.

Notably, collaborations are not confined to specific regions. Australia, for example, collaborates with countries like Canada, Denmark, Lebanon, and Sweden, showcasing a diverse network of research partnerships. The Netherlands, on the other hand, participates in collaborations with multiple countries, including Cyprus, Denmark, Ireland, and Switzerland, underscoring its broad international network.

Some collaborations, such as those between South Africa and Mexico or Spain and Turkey, appear to have regional characteristics, potentially focusing on shared research interests or regional challenges.

Conversely, collaborations between countries like China and Finland or Macedonia and North Macedonia suggest emerging research connections, possibly driven by common research interests.

Figure 11. Country Collaboration Map



Source: Bibliometrix Biblioshiny application

CHAPTER 4

DISCUSSION AND CONCLUSION

This research primarily focused on the ethical aspects of artificial intelligence and aimed to comprehend the current state-of-the-art in this field, identifying key trends and research gaps.

4.1. Overview of Ethical Considerations in AI from 2019 to 2023

Understanding AI ethics is a complex task due to the rapidly growing field and the need for a single, consolidated definition. (Morley et al., 2020) emphasize the importance of ethical principles like beneficence, non-maleficence, autonomy, justice, and explicability to guide AI development.
(Floridi et al., 2020) highlight the challenge of balancing various factors in AI ethics, including privacy, fairness, and consent.

Jobin et al. (2019) point out significant differences in the interpretation of ethical principles in AI. While transparency, justice, fairness, non-maleficence, responsibility, and privacy are common principles, there is no single ethical principle shared by all the documents they analysed. From their research, eleven overarching ethical values and principles emerged: Transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity.

This research aligns with at least eight principles, evident in the word cloud: Transparency (including explainability and explicability), justice and fairness (encompassing non-bias and non-discrimination), responsibility, privacy, trust, and sustainability. Although not visible in the word cloud, beneficence is also a central topic in some articles in the bibliographic collection. For instance, (Floridi et al., 2020) highlight the principle of beneficence, emphasising that AI should benefit people and the natural world, particularly in the context of AI for Social Good (AI4SG). They stress that AI4SG projects should not merely adhere to this principle but actively promote it, ensuring that the benefits of AI4SG are preferable and sustainable. Additionally, (Burr et al., 2020) identify critical social domains and broader themes central to ongoing discussions and research on the ethics of digital well-being.

Governance and fairness are emerging as central themes, signifying the field's depth and breadth beyond traditional ethical considerations.

Annual Scientific Production trends: The distribution of documents by year demonstrates a marked upsurge in scholarly contributions, with 2023 accounting for 32% of the total dataset, indicative of an escalating trend. This growth is pivotal in comprehending the current state-of-the-art and emphasising the pressing relevance of ethics in AI technologies, as well as the need to address its implications.

This trend reflects the dynamic nature of AI technologies, particularly the expanding influence of generative AI tools like ChatGPT. As IA becomes more integrated into various aspects of society, research on ethical considerations surrounding AI has understandably surged.

4.2. Most Influential Journals, Authors, and Countries in AI Ethics

Core Journals: Through applying Bradford's law, it became evident that a limited number of journals significantly contribute to the field. Notably, 'AI & Society', 'Science and Engineering Ethics' and 'Ethics and Information Technology' were identified as the most influential journals, collectively accounting for 34% of the bibliographic collection. These key sources shape the academic landscape in AI ethics, and their contributions reflect the diverse and interdisciplinary nature of the field.

Prolific Authors: Luciano Floridi emerged as the most prolific author, followed by Mariarosaria Taddeo and Mark Ryan, reflecting their extensive contributions to core journals.

Geographical Distribution: The USA, the UK, Germany, and the Netherlands significantly contributed to published documents. Despite being second in the number of documents, the UK demonstrated the highest average citations per article, highlighting the impact and quality of research from this region. Regarding academic significance, it is advisable to closely consider studies from Oxford University (UK), which has emerged as a prominent and influential contributor in this field.

4.3. Intellectual and Conceptual structure in AI Ethics

Co-citation Analysis: The co-citation network revealed two distinct clusters. The red cluster reflects foundational works and authors, while the blue cluster signifies contemporary and diverse articles. Both clusters present significant influencers in their respective periods, underscoring the evolution of discussions in AI Ethics.

Co-word analysis: The keyword co-occurrence network demonstrated five thematic clusters. It showed the interplay between AI-related technology and ethical considerations, suggesting opportunities for interdisciplinary collaboration. Additionally, it highlighted potential separations between governance concerns and ethical/technological domains.

4.4. Social structure and Collaborative Networks in AI Ethics

Author Collaboration: The author collaboration network indicated strong connections among authors from similar geographical regions. It portrayed collaborative research efforts and highlighted influential authors within their respective clusters, emphasising group dynamics and collaborative tendencies.

Country Collaboration: International collaborations showcased extensive networks of research partnerships across various regions, reflecting the global nature of AI ethics research. Collaborations were not restricted to specific regions and indicated diverse and emerging research connections.

4.5. Limitations

This research has several limitations. These limitations arise from the bibliometric techniques applied, as well as the selection of journals, articles, and initial search parameters.

Firstly, due to the multidisciplinary nature of the subject, it is challenging to cover the entire discipline comprehensively.

Secondly, a significant limitation is that the analysis relies on data from specific databases (WoS and Scopus) and the selection of specific knowledge domains, which may lead to excluding contributions from sources not included in this analysis.

By confining the search to specific research areas, dimensions of the field were inevitably not examined, particularly in research areas connected to social sciences or medical research, where ethical considerations are highly relevant and may have been overlooked.

Additionally, applying synonym rules in keyword analysis, while executed thoughtfully, introduces subjectivity and may impact the precision of keyword selection.

4.6. Implications and Future Research

The rapid growth in AI ethics research highlights the increasing awareness of ethical issues associated with AI technologies. Understanding the intellectual, conceptual, and social structures within AI ethics presents ample opportunities for interdisciplinary collaboration and future research directions. The identified trends and influential factors underscore the significance of ethics in the development and deployment of AI technologies, urging for continued exploration and multidisciplinary engagement in this field.

The identified research themes and topics reflect the evolving nature of AI ethics. As AI advances, ongoing research is vital to address emerging ethical challenges.

Future research in AI ethics should address the practical implications of ethical considerations in AI development, which may involve the development of ethical guidelines and frameworks, as well as the exploration of novel ethical challenges posed by AI technologies.

4.7. Conclusion

The research delves into the landscape of AI ethics from 2019 to 2023, exploring the complex and dynamic realm of ethical considerations within artificial intelligence. While it does not create new principles, it acknowledges and emphasizes the multifaceted nature of ethical principles guiding AI development, such as beneficence, autonomy, justice, and transparency. Highlighting the increasing significance of ethics in AI technologies, the research recognizes eight AI ethics fundamental principles evident in the research corpus: transparency, justice and fairness, responsibility, privacy, trust, sustainability, and beneficence. It underscores the growing scholarly contributions and the surge in awareness regarding ethical implications, particularly with the expanding influence of generative AI tools like ChatGPT. The research also maps the most influential journals, authors, and countries in the domain of AI ethics, shedding light on their substantial contributions and the interdisciplinary nature of this field.

Furthermore, the research employs co-citation and co-word analyses, revealing the intellectual, conceptual, and social structures in AI ethics. It delineates two distinct clusters in co-citation, emphasising the evolution of discussions over time, while the co-word analysis demonstrates thematic clusters and potential interdisciplinary collaboration opportunities.

In conclusion, the research calls for continued interdisciplinary engagement and future research directions to address the evolving ethical challenges in AI technologies, emphasising the need for practical implications, ethical guidelines, and novel frameworks to navigate the ethical landscape in this evolving field.

CHAPTER 5

REFERENCES

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 310. https://doi.org/10.1186/s12911-020-01332-6

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Burr, C., Taddeo, M., & Floridi, L. (2020). The Ethics of Digital Well-Being: A Thematic Review. *Science and Engineering Ethics*, 26(4), 2313–2343. https://doi.org/10.1007/s11948-020-00175-8

Coeckelbergh, M. (2010). Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, 12(3), 235–241. https://doi.org/10.1007/s10676-010-9221-y

Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 26(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research,* pp. 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eck, N. J. van, & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635–1651. https://doi.org/10.1002/asi.21075

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1), 205395171986054. https://doi.org/10.1177/2053951719860542

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

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(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020a). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, 26(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020b). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, 26(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Floridi, L., & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, 14(3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Floridi, L., & Strait, A. (2020). Ethical Foresight Analysis: What it is and Why it is Needed? *Minds and Machines*, 30(1), 77–97. https://doi.org/10.1007/s11023-020-09521-y

Hagendorff, T. (2020a). 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

Hagendorff, T. (2020b). 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

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

José de Oliveira, O., Francisco da Silva, F., Juliani, F., César Ferreira Motta Barbosa, L., & Vieira Nunhes, T. (2019). Bibliometric Method for Mapping the State-of-the-Art and Identifying Research Gaps and Trends in Literature: An Essential Instrument to Support the Development of Scientific Projects. In Scientometrics Recent Advances. IntechOpen. https://doi.org/10.5772/intechopen.85856

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 26(1), 89–120. https://doi.org/10.1007/s11948-018-00081-0

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, 45(2), 175–194. https://doi.org/10.1177/0312896219877678

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, 160(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & SOCIETY*, 35(4), 957–967. https://doi.org/10.1007/s00146-020-00950-y

Milano, S., Taddeo, M., & Floridi, L. (2021). Ethical aspects of multi-stakeholder recommendation systems. *The Information Society*, 37(1), 35–45. https://doi.org/10.1080/01972243.2020.1832636

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Moor, J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), 18–21. https://doi.org/10.1109/MIS.2006.80

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *El Profesional de La Información*, 29(1). https://doi.org/10.3145/epi.2020.ene.03

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines*, 31(2), 239–256. https://doi.org/10.1007/s11023-021-09563-w

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020a). 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., Floridi, L., Kinsey, L., & Elhalal, A. (2020b). 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*, 38(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8

OECD. (2019, May 22). Forty-two countries adopt new OECD principles on Artificial Intelligence. OECD. https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm

OECD. (2023). Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research. OECD Publishing, Paris. https://doi.org/10.1787/a8d820bd-en

Onyancha, O. B., & Ocholla, D. N. (2022). Dispersion of ICT-related subject terms in information and knowledge management publications: A Bradford analysis. *Humanities and Social Sciences Communications*, 9(1), 176. https://doi.org/10.1057/s41599-022-01189-2

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

Robbins, S. (2019). A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines*, 29(4), 495–514. https://doi.org/10.1007/s11023-019-09509-3

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & SOCIETY*, 36(1), 59–77. https://doi.org/10.1007/s00146-020-00992-2

Ryan, M. (2020a). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, 26(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Ryan, M. (2020b). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, 26(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Ryan, M., & Stahl, B. C. (2021a). 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, M., & Stahl, B. C. (2021b). 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

SPARROW, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhé Shaelou, S., Patel, A., Ryan, M., & 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

Stahl, B. C., Antoniou, J., Ryan, M., Macnish, K., & Jiya, T. (2022). Organisational responses to the ethical issues of artificial intelligence. *AI & SOCIETY*, 37(1), 23–37. https://doi.org/10.1007/s00146-021-01148-6

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022a). The ethics of algorithms: key problems and solutions. *AI & SOCIETY*, 37(1), 215–230. https://doi.org/10.1007/s00146-021-01154-8

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022b). The ethics of algorithms: key problems and solutions. *AI & SOCIETY*, 37(1), 215–230. https://doi.org/10.1007/s00146-021-01154-8

van Wynsberghe, A., & Robbins, S. (2019). Critiquing the Reasons for Making Artificial Moral Agents. *Science and Engineering Ethics*, 25(3), 719–735. https://doi.org/10.1007/s11948-018-0030-8

Zhu, J., & Liu, W. (2020). A tale of two databases: the use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321–335. https://doi.org/10.1007/s11192-020-03387-8