5-11-2023

Exploring Artificial Intelligence Bias, Fairness and Ethics in Organisation and Managerial Studies

Marco Smacchia
University of Chieti-Pescara, marco.smacchia@unich.it

Stefano Za
University of Chieti-Pescara, stefano.za@unich.it

EXPLORING ARTIFICIAL INTELLIGENCE BIAS, FAIRNESS AND ETHICS IN ORGANISATION AND MANAGERIAL STUDIES

Research Paper

Marco Smacchia, University of Chieti-Pescara, Italy, marco.smacchia@unich.it

Stefano Za, University of Chieti-Pescara, Italy, stefano.za@unich.it

Abstract

Due to the increasing adoption of AI technology in our society, this paper aims to develop a complete overview of the current debate on artificial intelligence bias fairness and ethics in organisation and managerial studies. To this end, we adopted the Computational Literature Review (CLR) method to conduct an impact and a topic modelling analysis of the relevant literature, using the Latent Dirichlet Allocation (LDA) technique. As a result, we identified and analysed 18 topics related to the selected domain. We further classified those topics into 5 categories creating a clear distinction between the social and the technical nature of a bias and its origins. Finally, focusing on the emerging topics, we proposed a set of guiding questions that might foster future research directions. This paper provides insights to scholars and managers interested in AI bias and ethical issues and could be used also as a guide to perform CLR.

Keywords: Algorithmic Bias, AI Fairness and Ethics, Artificial Intelligence, Computational Literature Review.

1 Introduction

In recent years, we have become accustomed to living in an increasingly automated world, where we can recognise an unfamiliar object or song in seconds, where cars start driving autonomously and where we can buy and receive an object in a few hours without getting up from our sofa. All these functions, and many more, are available thanks to the development of new technologies based on Artificial Intelligence (AI) (Haenlein and Kaplan, 2019). AI could be defined 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" (Haenlein and Kaplan, 2019). Since nowadays the data available is increasing above expectations and is generated from many sources that could be applied to a large variety of fields, we can use AI in almost every sector to accomplish a wide number of different tasks (Collins et al. 2021; Shobana and Kumar 2015; Tsai et al. 2015). AI is classified according to its cognitive capabilities compared to human intelligence (Zhu et al. 2021). The applications developed so far are labelled as artificial narrow intelligence (ANI) and are able to perform specific tasks autonomously using human-like capabilities. An example of ANI could be represented by machine learning (ML) algorithms. Other types of AI take a step forward in imitating human intelligence: the artificial general intelligence (AGI) that could learn, perceive, and understand like a human being and the artificial super intelligence (ASI) that could exceed human cognitive capabilities (ibid). On one side the development of AGI and ASI could have very large benefits to humankind, but on the other side opens to potential catastrophic risks for our society (Gill, 2016).

AI is one of the most relevant topics of the last decade and is having an impact on our society from government and companies to the single employee (Makarius et al. 2020a) representing often one of the

main components of a digital transformation process (Xu, Xu and Li, 2018). AI can be used in almost any organisational function supporting a great variety of decision or automation-based tasks, and it is likely to become an essential part of many jobs in the future (Huang and Rust, 2018). Therefore, it is of paramount importance that AI decisions do not reflect discriminatory behaviours (Mehrabi et al., 2021). Indeed, AI algorithms when applied to solve particular problems, could in some cases worsen the scenario by threatening rights, opportunities and wealth not only with the creation of new inequalities but also amplifying the existing ones (Hoffmann, 2019). Moreover, as many organisations and governments are beginning to use AI applications, the decisions of such systems could influence many people simultaneously, increasing the scope of potential problems related to ethics, fairness and algorithmic bias (Zuiderwijk, Chen and Salem, 2021).

The implementation of AI technologies along with their adoption often leads to some forms of resistance, frequently related to a lack of trust from employees and managers (Huang and Rust, 2018; Glikson and Woolley, 2020). Langer and König (2021) state that although trust in AI could be mainly related to its effectiveness and efficiency, it could be not confirmed when ethical issues are taken into consideration. Such issues request much attention to algorithmic bias and fairness and, to this end, AI-based systems are being reconsidered towards new approaches accordingly (Ntoutsi et al., 2020).

Due to the increasing adoption of AI technology in our society, the interest of academics has also increased, leading to new research, theories and questions (Makarius, Mukherjee, Fox and Fox, 2020b). In the last four years, studies on the societal impact of AI are increased exponentially, with a specific focus on ethical implications and fairness (Smacchia and Za, 2022). Trust, transparency and explainability of AI are becoming very important to many stakeholders and expertise in phenomena such as responsible AI will become essential to everyone working in this field (Meghan Rimol, 2021). In addition, ethical aspects are not always applied during the development of AI applications, revealing a disparity between technological and ethical advancement. As a result, further research is needed on methods and tools for implementing ethics into AI solutions design and development (Stahl, Timmermans and Mittelstadt, 2016; Vakkuri, Kemell and Abrahamsson, 2019).

In light of what is stated, our aim is to conduct a review, to map and evaluate (Tranfield, Denyer and Smart, 2003) the available literature on how the debate concerning AI bias, fairness and ethical implications has been developed inside organisational and managerial studies. Since the literature concerning the phenomena has increased in the last years (Smacchia and Za, 2022), among the several literature review methods, we decided to perform a computational literature review (CLR). This approach was preferred to others such as bibliometric analysis because through CLR, the articles in the selected domain could be qualitatively analysed, while the bibliometric approach is purely quantitative (Lamboglia, Lavorato, Scornavacca and Za, 2020). In particular, even though other literature review methods are supported by software, CLR goes further by using text mining and Machine Learning (ML) algorithms to examine the content of the article, allowing the machine to automatically perform time-consuming tasks.

The next section provides further details on the theoretical background followed by the research method. The presentation of the results with their implications and discussion closes the contribution.

2 Theoretical Background

In recent years, there has been a proliferation of new journals and conferences leading to an exponential increase in the existing literature (Mortenson and Vidgen, 2016). Furthermore, due to this growth in literature, scholars are relying heavily on literature reviews to inspect a particular research field (Badger, Nursten, Williams and Woodward, 2000). The increasing complexity and breadth of the scientific literature, along with the growing importance of unbiased reviews, has led to the need for systematic and easily replicable literature reviews (Antons, Breidbach, Joshi and Salge, 2021). Systematic Literature Review (SLR) is conducted using a systematic, rigorous and easily reproducible standard (Okoli and Schabram, 2010), and this method is preferred to analyse today's extensive literature (Rowe, 2014) and to avoid any bias in article selection, that is arbitrary in "Non-Systematic" reviews (Martin

Kunca 2018). However, the difficulty of conducting this type of research has increased significantly, with many researchers becoming discouraged by the time and effort required for such analyses and opting to focus on empirical studies (Mortenson and Vidgen, 2016). These problems have made it necessary to adopt new methods that allow the best practices of systematic literature review to be combined with computational methods (Antons et al., 2021) that would, on the one hand, speed up content analysis and, on the other hand, broaden the scope of reviews by identifying and extracting knowledge that would be precluded by manual analysis (Boyd-Graber, Hu and Mimno, 2017). One method that enables researchers to analyse large volumes of documents in a rigorous and timely manner is CLR. defined as:

"A structured process intended to augment human researchers' information processing capabilities through the use of machine learning algorithms that help analyse the content of a comprehensive text corpus in a specific knowledge domain (e.g., a research topic, academic journal, or scientific field) in a way that is scalable and real-time capable." (Antons et al. 2021)

CLR allows to automatically analyse a dataset by identifying themes through topic modelling. It is sometimes referred to as non-linear principal component analysis because it finds latent components (called topics) that can explain variance in the data (Hindle et al., 2020). Topic models are algorithms capable of discovering themes within a large number of unstructured documents by analysing the connections between words contained therein (Blei 2012). An example of a topic model is Latent Dirichlet Allocation (LDA), a probabilistic generative model for the collection and analysis of unstructured data. LDA is applied to find topics within a text on the basis of links between words and then classify texts according to the relevance of the topics found within them (Blei, Ng and Edu, 2003).

A CLR can be performed using the guidelines provided by Antons et al. (2021), where they describe a six-step process analysis:

1. *Define a conceptual goal* that motivates the review.
2. *Operationalise the CLR* by defining the boundaries that are going to be inspected.
3. *Choose a computational technique* that best suits the conceptual goal.
4. *Perform content analysis* by preparing the data and deploying the computational technique.
5. *Generate original insights* by observing the results provided by the computational analysis
6. *Present the findings* in a clear and accessible way.

In this paper, we use those guidelines as a baseline adapting the Smacchia and Za (2022) framework to run our CLR.

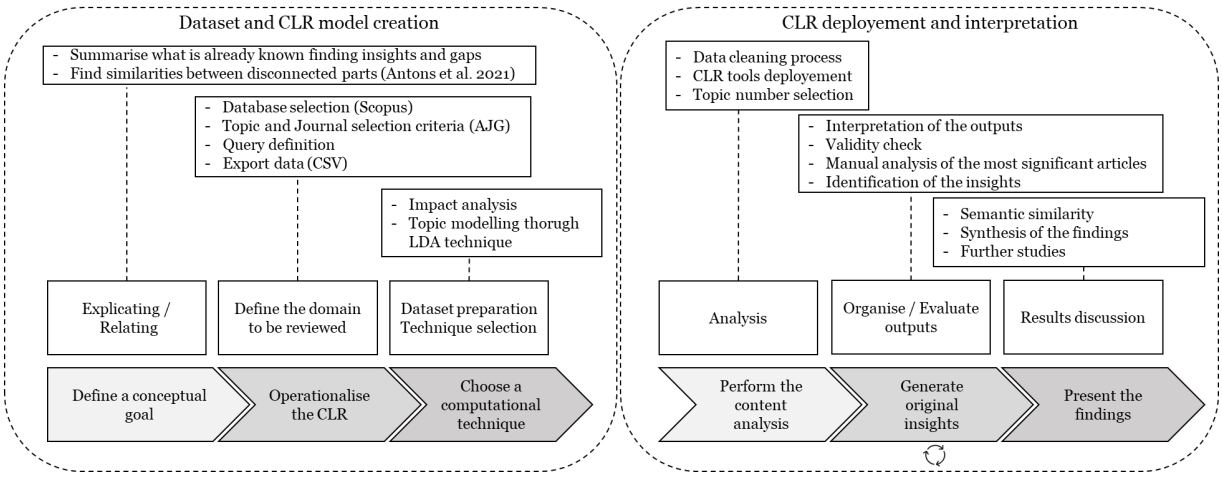

Figure 1. Research protocol adapted from Smacchia and Za, (2022).

Before starting the analysis, we examined the existing literature reviews on this research domain that we summarised in Table 1. We performed a query on Scopus, Web of Science and Google Scholar and

selected all the literature reviews in line with AI ethics, bias and fairness in organisational and managerial studies. Analysing the previous literature reviews, we found out that they are systematic literature reviews based on a limited set of papers. With a CLR we want to go further considering a larger dataset composed by more than a thousand articles, trying to investigate the topic using both a quantitative and a qualitative perspective.

Authors	Methodology	Focus	Findings
Meek et al., (2017)	Literature review of AI ethics contextualised into four categories using PEST tools.	Identify the central ethical issues related to AI and then isolate management recommendations by analysing those issues.	The authors provide a timeline of the most important milestones concerning Artificial Intelligence and a description of several ethical issues related to AI. In conclusion, they also make some recommendations to address AI ethical issues along with gaps for future studies.
Riazy et al., (2021)	Systematic lit. Review & case study 56 articles gathered from different databases published before 2021.	Fairness and explainability in learning analytics.	The authors identified methods to measure and mitigate discrimination in learning analytics. In the second part of the paper, they apply what discovered to mitigate discrimination problems in an open dataset (students with disability).
Akter et al., (2021)	Systematic literature review and thematic analysis on 40 papers published between 2016 and 2020.	Sources of algorithmic biases in data-driven information (DDI).	The authors provide a framework to describe three major algorithmic biases in DDI phases together with guidelines to address these biases focusing on data, method and managerial capabilities.
Khan et al., (2022)	Systematic lit. review of 27 primary studies published before February 2021.	AI ethics principles and factors that could negatively impact the adoption of AI.	The authors recognised 22 ethical principles and 15 challenging factors concerning artificial intelligence.
Kordzadeh and Ghasemaghaei, (2022)	Systematic lit. review and thematic analysis of 56 papers published between 2010 and 2019 (scholarly journal and conference proceedings).	Algorithmic bias.	The authors discovered eight theoretical concepts concerning algorithmic biases and study their relationships. They also pointed out that most studies have conceptually discussed the ethical, legal, and design implications of algorithmic bias, whereas only a limited number have empirically examined them.
van Giffen et al., (2022)	Systematic problem-centred literature review on 68 articles.	Machine learning biases in business, IS and marketing studies	The authors identified eight distinct ML biases and mapped them. They also propose twenty-four bias mitigation methods and a conceptual model to illustrate the application of ML algorithms in marketing that helped them to analyse the biases in a case study.
Ashok et al., (2022)	Systematic lit. review and qualitative synthesis on 59 papers published in 43 journals between 2018 and 2021	Ethical use of AI in digital technologies	The authors found fourteen digital ethics implications associated with digital technologies archetypes. After, they map every archetype based on the presence of every ethical implication found.

Table 1. Literature review comparison.

3 Research Method

The main purpose of this contribution, and thus our conceptual goal, is to explore and summarise the debate on AI bias, fairness and ethics in organisation and managerial studies. Hence, it was quite relevant to operationalise our literature review by identifying the appropriate list of journals on which to perform the query in order to select the papers to create our dataset. The Academic Journal Guide (AJG) of the Association of Business Schools (ABS) provides a list of journals classified according to the field of studies to which they belong and their ranking. We included all the journals in the ranking. The data were collected within Scopus since all the journals we decided to incorporate into the search were available in this academic database. Specifically, we performed a query in which we included all the chosen journals by entering their ISSN code and the following query:

TITLE-ABS ("deep learning" OR "artificial intelligence" OR "machine learning" OR "neural networks") OR AUTHKEY (“deep learning" OR "artificial intelligence" OR "machine learning" OR "neural networks")) AND (TITLE-ABS ("bias" OR "injustice" OR "ethics" OR "fairness" OR "trust") OR AUTHKEY ("bias" OR "injustice" OR "ethics" OR "fairness" OR "trust")) AND ISSN…

We choose not to include Scopus index keywords because they are often inaccurate and generate a lot of noise inside the dataset. We did not use time or other restrictions. The query returned 1217 articles published between 1981 and 2022; citations and abstracts data were exported in CSV format. The downloaded dataset was also revised and cleaned. In particular, the occurrences that didn’t have the author information (errata articles as well) were deleted and, where possible, the papers with missing abstract were integrated. In the end, the analysis was conducted on 1198 articles. We then analysed our dataset using the R tool Bibliometrix for a preliminary description.

Our first task was the choice of the computational technique to be used. We initially performed an impact analysis using the R programming language through the Bibliometrix package (Aria and Cuccurullo, 2017) to get quantitative data on the information contained in our dataset. Through this analysis, it is possible to estimate the impact of a given paper, author and journal. To measure these dimensions, the tool uses various metrics such as citation count, the impact factor (total citation count divided by the total number of papers) and the h-index which is commonly used to assess the impact of researchers (Hirsch, 2005). We then focused on the content analysis of the papers by performing topic modelling using the Latent Dirichlet Allocation analysis with the *lda* package. We adopted LDA because it is the most popular method used for topic modelling (Jelodar et al., 2019) and also this method seems to have a higher level of reliability and accuracy compared to the others (Shadikur Rahmane, 2020). LDA is an unsupervised generative probabilistic model of a corpus (Blei et al., 2003) that allows the identification of a set of topics among multiple documents. Each document is considered as a set of words that can be combined to form subsets of latent topics. The model assumes that the corpus includes k topics, and then distributes those topics across each document to see which one fits best. In this way, the analysis is more efficient because it avoids cross-checking every word with every document. LDA algorithm can be summarised as follow:

$$P(\boldsymbol{W}, \boldsymbol{Z}, \theta, \varphi, \alpha, \beta) = \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) P(W_{j,t}; \varphi_{z_{j,t}})$$

Where α represents the document-topic density (the percentage that a document is associated with a determined topic, if $\alpha < 1$ the documents tend to diverge to the single topics) β represents the topic-word density (the percentage that a word is associated with a determined topic) and θ with φ representing multinomial distributions of topics over documents and words over topics.

In order to perform the analysis, we had to clean our dataset. First, we deleted all the variables except for those related to abstracts and document id, then we removed punctuation, stop-words and tokenized the corpus to build the Document Term Matrix (DTM), which is a matrix that contains words and documents as dimensions. In the DTM the rows correspond to the documents and columns correspond to the terms, it is essential to inspect the frequency of terms inside a document collection. We also removed all the terms in our query, along with other recurring words that might yield incorrect results

such as “Elsevier”, “Springer”, “Research” and “Findings” (Mortenson and Vidgen, 2016). When creating the DTM, we decided to tokenise the text using one or two words to avoid ambiguity during topic analysis. In fact, some words have a different meaning when taken individually (i.e. Information and System instead of Information System). Since LDA is an unsupervised technique, that has to discover patterns from untagged data, the number of topics to be used was chosen a priori. To assess the value of K we used the cross-validation procedure based on the perplexity that measures how well a probabilistic method can predict a sample (lower levels of perplexity can better predict a sample). To this end, we used the R package *ldatuning* setting the algorithm for values from 1 to 100. To discover the value we used two metrics, CaoJuan2009 (Cao et al., 2009) and Deveaud2014 (Deveaud, SanJuan and Bellot, 2014), where the first one has to be minimised while the second maximised. After the cross-validation procedure, we further analysed the results to assess the correctness and the significance of each topic and, hence, to validate the number “K” selected. According to the theta parameter (distribution of topics over documents) that the algorithm gave to every document, we used topic modelling results to analyse the most representative contributions for each topic (the complete list is available at: https://bit.ly/3K1EbML). More specifically the theta parameter represents the probability that a document is contained in a certain topic, hence the documents selected were the most representative of each topic. Our goal was to discover how AI is implemented and deployed within different topics and also to detect the most popular fields of research in recent years. Figure 1 describes the research protocol we adopted to explore the literature using the CLR method. The part dealing with output interpretation and article in-depth analysis within the framework should follow an iterative cycle to have more precise and significant results. In fact, it is necessary to repeat the analysis several times to reach an output that is as relevant and reliable as possible. To find hidden patterns in the analysed literature we used the R package LDAvis (Sievert and Shirley, 2015). This interactive interface allows the visualisation of topics estimated by the LDA algorithm and it could be used to see the most associated keywords within each topic as well as a global distribution of them, including their similarities and differences.

4 Results

4.1 Impact Analysis

The impact analysis was conducted to have information about the impact of the areas in which papers related to the selected domain are published. Moreover, thanks to the citation count and H-Index we assessed the impact of the single articles and the journals, acquiring insights about the trend of publications including which field has acquired a relevant position in the debate. Concerning the article’s citations, only 230 (less than 20%) of them have zero citations (only 13 articles published before 2021) underlining the relevance of the papers in our dataset. The 20 most cited papers (the complete list is available at: https://bit.ly/3K1EbML) are distributed between 1998 and 2020 (10 articles published before 2016 and 10 articles published after). It can be noted that the oldest articles are represented mainly by technical studies with the aim of developing or enhancing AI algorithms. The majority of the newest studies adopt a social or organisational perspective investigating issues concerning the evolution of AI or the effect and/or the impact of AI adoption, suggesting a widening of the AI research stream integrating the technical issues with societal perspectives. We have also represented the top 20 journals in terms of number of citations with the H-Index, AJG Area and ranking (Table 2). Looking at the relevant AJG fields of the most cited journals we can observe a great variety of research domains, which is a sign of a wide breadth of topics discussed concerning AI-related ethical issues, fairness and biases. In particular, the most representative field in the ranking is Information Management followed by Operations Research and Management Science, Economics, Econometrics and Statistics and General Management, Ethics, Gender and Social Responsibility. By looking both at the AJG field and at the name of the source it can be made a distinction between the journals that are focused on societal and organisational studies and the journals that are focused more on technical and mathematical research. This confirms the same distinction recognised for the most representative articles.

Journal	H-Index	Citations	AJG Area	AJG Rank
Annals Of Statistics	5	3326	Econ. and Statistics	4*
Expert Sys. with Applications	25	2719	Information Management	1
Computers in Hum. Behavior	18	1125	Information Management	2
Int. Journal of Inf. Man.	12	1119	Information Management	2
Ethics And Inf. Technology	16	944	Information Management	1
Journal of Cleaner Production	12	687	Tourism and Sector Studies	2
Journal of Service Man.	4	567	Tourism and Sector Studies	2
Euro. J. of Operational Res.	7	478	Op. Res. and Manag. Sci.	4
Journal Of Business Research	11	471	Man., Ethics and Soc. Resp.	3
Decision Support Systems	11	427	Information Management	3
Business Horizons	7	422	Man., Ethics and Soc. Resp	2
Reliability Eng. & Sys. Safety	4	410	Op. Res. and Manag. Sci.	3
Evolutionary Computation	5	407	Op. Res. and Manag. Sci.	3
Management Science	4	397	Op. Res. and Manag. Sci.	4*
J. of the Acad. of Market. Sci.	1	366	Marketing	4*
Econometrics Journal	2	360	Econ. and Statistics	3
Int. J. of Human Comp. Stud.	6	343	Information Management	2
IEEE Trans. on Evo. Comp.	7	342	Op. Res. and Manag. Sci.	4
Info. Processing and Manag.	9	270	Information Management	2
Psychological Review	3	261	Psychology (General)	4

Table 2. Top 20 Journals ranked by citation count and AJG Area.

4.2 Content analysis

After the data cleaning process, we performed a topic model analysis on the articles' abstracts. Since the LDA technique is an unsupervised learning algorithm, we had to determine a priori the number of topics (K). Figure 2, graphically represents the distribution of topics (the y-axis indicates the level of perplexity and the x-axis the number of topics). To assess the correct number of topics to select we manually calculated the point at which the two distributions were closest. Based on the results of the metrics CaoJuan2009 and Deveaud2014, 18 topics were selected, that we manually inspected to determine the consistency and accuracy with respect to the documents contained inside them and also to see if two or more topics could be merged together.

The 18 topics are related to different aspects of AI bias, fairness and ethics inside organisational and managerial settings. They could be summarised as follows: 1) AI to predict customer intentions and the implications on privacy, 2) AI to predict system performances, 3) Trends, performances and biases of artificial neural networks (ANN), 4) Content analysis on AI tools, 5) Machine learning for prediction and classification, 6) Using digital technologies to overcome barriers inside enterprises, 7) AI to control network performances, 8) Learning analytics in education, 9) Reducing bias in ML algorithms, 10) Machine anthropomorphism, 11) Natural language processing (NLP) and image recognition training to reduce error, 12) Bright and dark side of AI in the public sector, 13) AI tools for monitoring social media, 14) AI & big data in finance, 15) ML to predict quality of service in healthcare, 16) AI fairness in decision-making, 17) AI becomes human, 18) Explainable AI to build trust between human and machine (see Table 3 for further details).

The number of documents contained within each topic varies between 40 and 80 ($m = 66.56$, $sd = 15.91$). The only two outliers are represented by topics 1 - AI to predict customer intentions and the implications on privacy and 2 - AI to predict system performances which have respectively 115 and 86 articles inside them. For each topic, we represented the number of publications between 1981 and 2022 to determine

the most recent trends in academic research. At the same time, we inspected the most important journals inside each topic to see which AJG research field is predominant. Every topic found had a relevant number of publications in the years 2021 and 2022, with the number of articles increasing by up to ten times. Such as topic 17 - AI becomes human and topic 6 - Using digital technologies to overcome barriers inside enterprises which went from an average of 1.1 and 2.4 articles respectively in 2019 to 14 and 34 articles in 2022. Particularly with regard to topic 17 - AI becomes human, the interest of academics dealing with organisational and social studies has significantly increased in recent years due to technological advances in the field of machine learning and the growing concerns about the invention of a sentient AI (Korteling et al., 2021; Tiku, 2022).

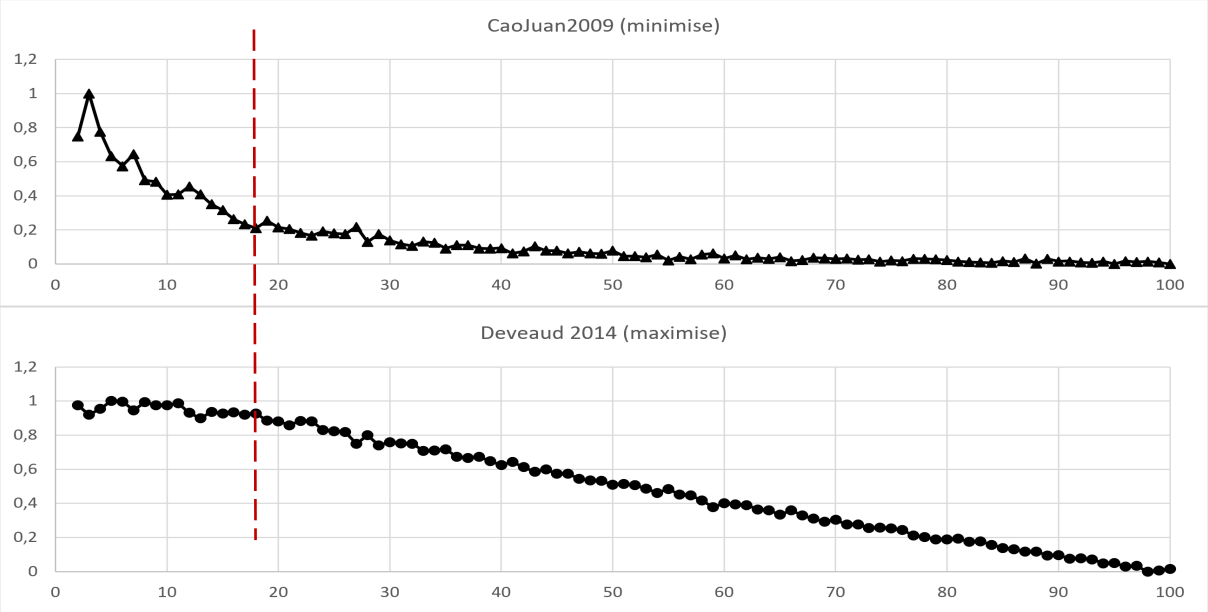

Figure 2. Selection of number of topics (K).

Within each topic, there are several journals (m = 42.78, sd = 8.61) and the most represented AJG field inside them is Information Management since it is the most related field within the selected research domain. The exceptions are represented by topic 15 - ML to predict the quality of service in healthcare in which the most representative fields (37%) are Social Sciences and Economics, Econometrics and Statistics. Topic 3 - Trends, performances and biases of artificial neural networks (ANN) with nearly 40% of the articles contained inside the research field of Operations Research and Management Science. Topic 14 - AI & big data in finance and topic 9 - Reducing bias in ML algorithms have respectively 31% and 56% of the articles belonging to the field Finance and Economics, Econometrics and Statistics. It can also be noted a convergence between the content of the topic and the most representative AJG fields. Overall, the three most important fields inside the dataset are Information Management, Operations Research and Management Science, Economics, Econometrics and Statistics and General Management, Ethics, Gender and Social Responsibility with respectively 434, 137, 105 and 67 articles. The situation in the AJG ranking areas (Figure 3) is similar to what was observed for the topics. The number of publications within them has increased significantly in recent years, a sign that AI is becoming an increasingly relevant phenomenon in all managerial and organisational fields, as well as the ethics and the issues of the prejudice related to that technology. Some areas such as Business and Economic History, Human Resource Management and Employment Studies, Organisational Studies, Public Sector and Health Care did not have any publications until the last 5 years. This aspect could be an indicator of the increasing interest in addressing ethical problems related to AI in both private and public organisations with attention to human resource management.

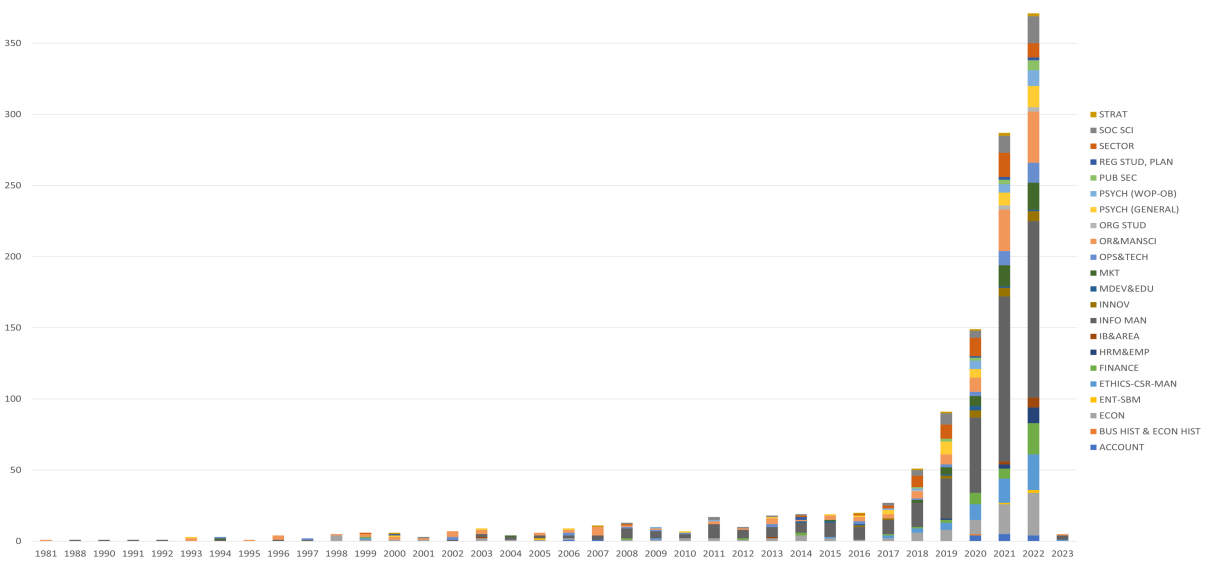

Figure 3. Number of publications per year since 1983 divided by AJG fields (see research method paragraph). Full legend available at: https://bit.ly/3K1EbML.

5 Discussion

After performing the impact and content analysis, we manually analysed the most relevant articles in each topic. A more in-depth analysis of the literature in each cluster helps to better understand the results of the topic modelling analysis and provides better insights. According to the theta value (distribution of topics over documents), we selected and analysed the most relevant papers inside every topic. As a result, we defined a more appropriate title and description for each topic, not only based on the list of keywords provided by the algorithm (Table 3).

Topic	Topic description
T1 - AI to predict customer intentions and implications on privacy	The topic focuses on the use of AI tools to analyse what factors are most impactful in predicting customer intentions and preferences, with a focus on privacy issues.
T2 - AI to predict system performances	The topic focuses on the use of ML and Artificial Neural Network (ANN) models to predict the performance of a specific system, especially those connected to renewable energy.
T3 - Trends, performances, and biases of ANN	The topic is focused on the study of the performances of neural network algorithms and on the description of the latest trend inside the field.
T4 - Content analysis on AI tools	The theme of this topic is related to studies that use primary and secondary data with the aim of describing the performances of AI and ML tools.
T5 - Machine learning for prediction and classification	The topic focuses on the development and study of new ML models for prediction and classification.
T6 - Using digital technologies to overcome barriers inside enterprises	The theme concerns the use of digital technology to overcome institutional bias, operational issues and communication difficulties inside companies, other studies focus on the barriers such as security, performances and standardisations emerging during a digital transformation process.
T7 - AI to control network performances	The theme is related to the use of AI tools such as ML algorithms to detect and mitigate faults within a network.
T8 - Learning analytics in education	This topic is mainly focused on the role of AI tools in the educational sector. More specifically, high emphasis is given to learning analytics tools and the role they play in supporting teachers in doing their job.

T9 - Reducing bias in ML algorithms	This topic is related to ML algorithms, trying to refine those algorithms by reducing biases with the introduction of different computational techniques
T10 - Machine anthropomorphism	This topic is related to the concept of automation anthropomorphism and its relative impact on humans.
T11 - NLP and image recognition training to reduce error	The topic focuses on the development and training of NLP and image recognition algorithms to avoid bias and improve accuracy. The main aim is to use those algorithms to find reliable patterns that reflect human judgement.
T12 - Bright and dark side of AI in the public sector	The main theme is the impact of AI applications in the public sector, empirically analysing the pros and cons of AI adoption and implications.
T13 - AI tools for monitoring social media	The papers in the topic apply AI techniques to explore the behaviour and intentions of companies and users on social media.
T14 – AI & big data in finance	The topic examines the application of AI and ML tools in the financial sector.
T15 - ML to predict quality of service in healthcare	This topic focuses on the use of ML algorithms to predict the quality of services in the healthcare sector, as well as to monitor parameters related to the improvement of human welfare.
T16 - AI fairness in decision-making	The topic is related to a delicate theme, the fairness of algorithmic decision-making and its impact on minorities.
T17 - AI becomes human	The papers in this topic investigate if AI could be seen, perceived and treated like a human being. Some relevant papers propose tests to evaluate AGI.
T18 - Explainable AI to build trust between human and machine	The human lack of trust in AI applications is the main theme of this topic. Particular attention is paid to the financial and healthcare sector in which the decisions made by an AI application could produce serious implications.

Table 3. Topics title and description.

Once clarified the content of the papers assigned to each topic, we tried to study their similarity to find hidden patterns inside the literature analysed. We used the R package LDAvis to explore the similarities between topics and have a complete overview of them. The intertopic distance map (figure 1) refers to the degree of difference or similarity between topics in a given text corpus, while the marginal topic distribution represents the overall prevalence of each topic in the corpus of documents being analysed. Moreover, we build a dendrogram using Hellinger's Distance, a metric used to quantify the difference between two probability distributions. The dendrogram was useful to determine the similarity of the topics. Based on the results shown in Figure 4 and on the previous in-depth topic analysis, we could group the topics into two main categories of discussion, such as: one where the bias emerges directly from the output of the algorithm (technical nature of the bias) while in the other one biases could be ascribed to socio-cultural components (socio-cultural nature of the bias).

Macro-topic 1 – Technical nature of the bias: this cluster is composed of topics 2) AI to predict system performances, 3) Trends, performances and biases of artificial neural networks (ANN), 5) Machine learning for prediction and classification, 7) AI to control network performances, 9) Reducing bias in ML algorithms, 11) Natural language processing (NLP) and image recognition training to reduce error. This set of topics focuses on technical and mathematical studies concerning AI algorithms. It concentrates on the enhancement of the mathematical and statistical techniques on which these algorithms are based. Thus, an attempt is made to reduce and address errors during the development and testing phase that may have adverse effects on the outcome of the systems in which they are used.

Macro-topic 2 – Socio-cultural nature of the bias: this cluster is composed of topics 1) AI to predict customer intentions and the implications on privacy, 4) Content analysis on AI tools, 6) Using digital technologies to overcome barriers inside enterprises, 8) Learning analytics in education, 10) Machine anthropomorphism, 12) Bright and dark side of AI in the public sector, 13) AI tools for monitoring social media, 14) AI & big data in finance, 15) ML to predict quality of service in healthcare, 16) AI fairness in decision-making, 17) AI becomes human, 18) Explainable AI to build trust between human and machine. In contrast to the first macro cluster, here the studies are mainly managerial and sociological in nature, specialised in the research on the implications of artificial intelligence algorithms

when applied in different contexts. The focus of this cluster is on empirical studies in different sectors in which various AI-based applications are implemented and adopted. The majority of those contributions consider different forms of organisational settings.

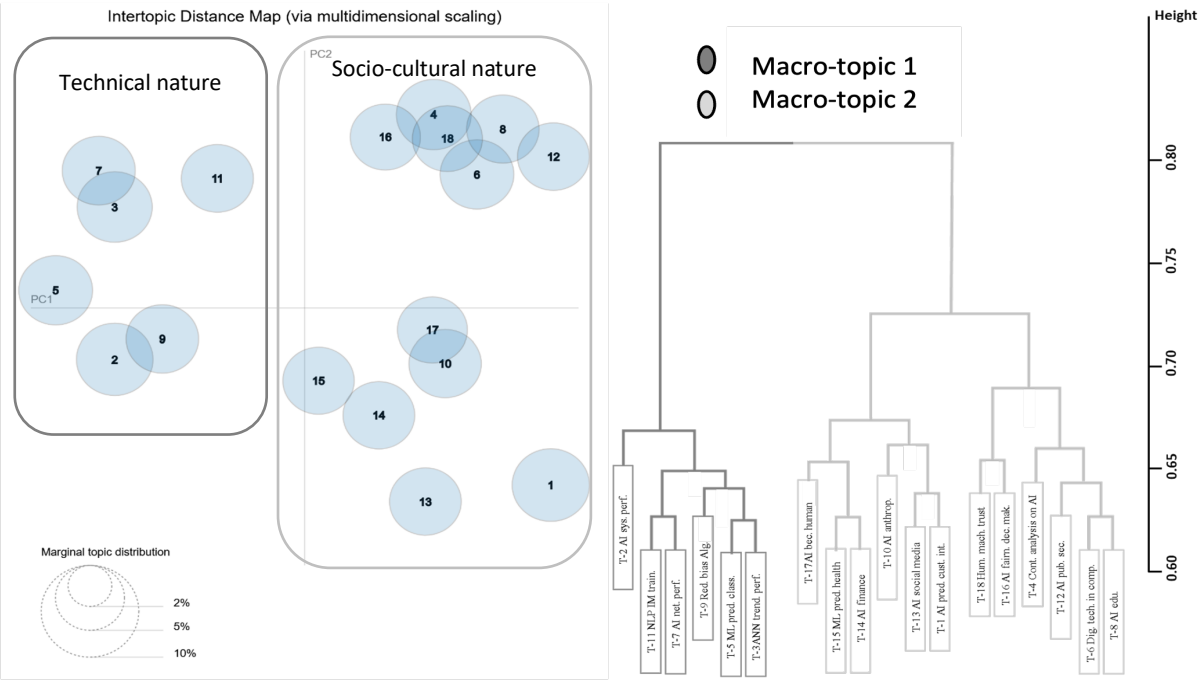

Figure 4. Intertopic distance representation. LDAvis (right) Hellinger's Distance (left).

Comparing the evidence produced by the cluster analysis with the results of the in-depth analysis performed manually, we were able to identify a further list of sub-categories, providing an intermediate classification of the papers, in between the two macro-topics and the 18 single topics, called meso-topics. We then identified two meso-topics belonging to Macro-topic 1 (*Technical nature of the bias*) and three meso-topics for Macro-topic 2 (*Socio-cultural nature of the bias*). Finally, looking at the meso-topics classification and description, we have had the chance to draft a tentative list of possible future research directions, summarized in Table 4. More details about each meso-topic are provided following.

Meso-topic 1 Algorithmic bias reduction: this cluster contains the topics 3) Trends, performances and biases of artificial neural networks (ANN), 7) AI to control network performances, 9) Reducing bias in ML algorithms, 11) Natural language processing (NLP) and image recognition training to reduce error. This set of topics belongs to the first macro-topic and is focused on the development, enhancement and implementation of various statistical techniques to reduce bias in AI algorithms. The boundaries of the subjects explored are very broad, from studies to reduce errors in speech and image recognition (Hagen, 2018; Wang, Liang, Xu and Lin, 2022) to research that has the objective to make sensor networks on unmanned airships as reliable as possible in order to avoid serious accidents (Yu et al., 2022).

Meso-topic 2 Predictive algorithms: this group includes topic 2) AI to predict system performances, and topic 5) Machine learning for prediction and classification. This meso-topic is dedicated to the study, development and enhancement of predictive algorithms. The reliability of this kind of algorithms is often crucial because they help technicians in the decision-making processes of important projects such as the construction of a renewable energy plant (Khatib, Mohamed, Mahmoud and Sopian, 2011; Abujazar et al., 2018) or new methods to dispose of waste in firms (Vu, Ng, Richter and An, 2022).

Meso-topic 3 AI for Individual behaviour: the topics in this cluster are 1) AI to predict customer intentions and the implications on privacy, 13) AI tools for monitoring social media, 14) AI & big data in finance, 15) ML to predict quality of service in healthcare. The contributions in this meso-topic are focused on the exploration of the factors (e.g., emotions, knowledge) affecting the adoption, the use and the relationship with AI technology (Chong, 2013; Vimalkumar, Sharma, Singh and Dwivedi, 2021).

The studies discuss a variety of subjects from ways to enhance algorithms for financial investment to surveys and interviews to improve healthcare services (Ali, Salehnejad and Mansur, 2018; Bhatia et al., 2021). Considerable importance is given to privacy concerns especially when the algorithms are dealing with sensible information (Ameen, Tarhini, Reppel and Anand, 2021; Ameen, Hosany and Paul, 2022).

Meso-topic 4 Trusting, understanding and exploiting AI: this group comprises the topics 4) Content analysis on AI tools, 6) Using digital technologies to overcome barriers inside enterprises, 8) Learning analytics in education, 12) Bright and dark side of AI in the public sector, 16) AI fairness in decision-making, 18) Explainable AI to build trust between human and machine. The studies included in this meso-topic try to address issues related to the lack of explainability and transparency of AI to increase trust in its users and to overcome many kinds of barriers. Great importance is given to the improvement of communication in many settings (Alam and Mueller, 2022; Vössing, Kühl, Lind and Satzger, 2022). More specifically, transparent AI could improve communications inside firms, participation of citizens in public administrations (e.g., less corruption) and also improve the quality of teaching in schools and universities thanks to AI-human augmentation (Anastasiadou, Santos and Montargil, 2021; Marshall, Pardo, Smith and Watson, 2022).

Meso-topic 5 Anthropomorphism of AI: in this cluster, we have topics 10) Machine anthropomorphism, 17) AI becomes human. The last meso-topic is dedicated to the evolution of AI, in particular to its anthropomorphisation (de Visser et al., 2016; Peng et al., 2022; Schelble et al., 2022). The papers in this meso-topic discuss how humans perceive AI, especially when it imitates their behaviour, and which could be the possible implication of crucial decisions taken by an AGI. Those papers usually propose tests and recommendations that try to address issues related to ethical concerns such as human replacement (Sparrow, 2004; Swanepoel, 2021).

Topics	Tentative future research questions
Meso-topic 1 Algorithmic bias reduction	Although the bias reduction from a technical point of view could make AI techniques more reliable, what are the implications of their use? For example, the case of NLP and image recognition tools used to make delicate decisions such as diagnosis in healthcare. Increasing the complexity of the AI system for making more accurate decisions, could affect bias reduction?
Meso-topic 2 Predictive algorithms	Some studies investigate the comparison between the predictive outcomes conducted only by humans with others that are performed by AI-human augmentation, in those cases which aspects could be relevant to assess beyond the technical perspective? What are the possible implications for individuals and organisations of regulators and policymakers interventions concerning the development of ML models used for prediction and classification?
Meso-topic 3 AI for Individual behaviour	What are the main factors that impact user privacy and trust concerns when using AI technology? What are the main issues related to the adoption of a technology based on AI algorithms from a customer (individual or organisation) point of view?
Meso-topic 4 Trusting, understanding, and exploiting AI	What are the consequences that algorithmic bias inside ML tools brings to the results of a firm's decision-making process? Is there a difference in reaction and reception of new AI-based technology between employees that have different roles? What will be the implication for employees to work side by side with AI? Are there factors that could cause a disparity between employees inside an organisation?
Meso-topic 5 Anthropomorphism of AI	What are the main attributes that influence the relationship between humans and machines when working together? In which case the machine could be perceived as a colleague and not a tool and what are the long-term implications of human-machine interaction? What are the legal implications of an AI that is comparable to a human being? How could policymakers address the regulatory challenges that arise from the development of AGI?

Table 4. Meso-topics titles and possible future research opportunities..

6 Conclusion

In this paper, we performed a CLR to explore the evolution of the debate concerning AI bias, ethics and fairness inside organisational and managerial studies. We performed firstly an impact analysis evaluating the impact of journals and articles, and afterwards, we conducted a content analysis using a topic modelling algorithm identifying 18 topics that are relevant and enough for clustering the papers of our dataset. We then manually analysed the most relevant papers within each topic in order to refine its title and description. Finally, combining the outcome of the automatic content analysis and the review performed by the authors, we identified the possibility to classify the 18 topics into two main categories called macro-topics (e.g., technical and socio-cultural nature of the bias) and afterwards looking in detail at the content of the papers belonging to each topic, we recognized five middle categories, called meso-topics. The results show coherence between the impact and the content analysis concerning the distinction between studies adopting a technical or a socio-cultural perspective in their discussion.

From a theoretical perspective, we provided a rigorous and easy replicable method to conduct a computational literature review study. More specifically we used the framework given by Antons et al. (2021) adding a further step concerning the in-depth analysis and description of each topic and thus making the interpretation of results an iterative process to improve the inspection of the topic model output. CLR allows us to qualitatively analyse a large amount of data. As mentioned, the benefits are numerous. First of all, a very large number of articles can be examined, in terms of both content and impact. CLR also allows scholars to conduct research in a time-saving manner, as it is quite difficult to analyse thousands of documents manually in a limited amount of time. Moreover, thanks to CLR is simpler to analyse multi-domain literature simultaneously to discover hidden patterns otherwise hard to be found. In this contribution, we also provided a classification of literature concerning AI bias, ethics and fairness with a focus on organisational and managerial studies that could help other researchers to have a better understanding of the domain and to have an outlook of the different research strands. While previous studies focus on a limited set of papers by exploring the phenomenon in a specific field, we explored a broader set of articles, investigating different aspects concerning the same phenomenon debating in different field of study. Furthermore, we presented a tentative set of research questions based on the content of each meso-topic that could provide opportunities for future research. Finally, we tried to describe in detail all the steps that allowed us to perform the CLR hoping that this article could be of support to scholars that are willing to carry out similar research.

From a managerial point of view, the article provides a clear distinction of the research strands that could be helpful to managers during the development, testing and adoption phases of an AI system. The different classifications help to identify which are the studies that address a particular issue: from the choice of a system used to predict the outcome of a specific project to the decision of implementing an AI application that works closely related to human beings. Moreover, focusing on AI ethical issues, this research provides some insights and references useful in promoting a responsible use of AI.

The contribution has some limitations. First of all, the analysis is only based on abstracts and not on the entire content of each paper. Even though the purpose of the abstract is "to facilitate quick and accurate identification of the *topic of published papers"* (Peter Luhn, 1958), we argue that a more in-depth analysis of the contents of the text corpus can improve the final results by providing further insights. Could be also interesting to explore more in-depth the theta distribution of the topic modelling output to study papers that are relevant to more than one topic. This analysis could allow a more precise description of hidden connections between topics and avoid biases related to the inspection of only the most relevant articles. Further studies could also concentrate on different kinds of sources such as conference proceedings, thus providing the last research trends inside the domain. Furthermore, new NLP techniques (e.g., neural networks) could be used to explore the same datasets and compare the results. Nowadays, the proliferation of scientific literature increases the need for systematic, replicable and rigorous literature reviews, as well as the resources needed to conduct them (Badger et al., 2000). In parallel, digital technologies are becoming more and more pervasive in our life, then should be fundamental to understand those tools and the implications connected to their uses from different perspectives.

References

Abujazar, M. S. S., S. Fatihah, I. A. Ibrahim, A. E. Kabeel and S. Sharil. (2018). "Productivity modelling of a developed inclined stepped solar still system based on actual performance and using a cascaded forward neural network model." *Journal of Cleaner Production*, *170*, 147–159.

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

Alam, L. and S. T. Mueller. (2022). "Examining Physicians' Explanatory Reasoning in Re-Diagnosis Scenarios for Improving AI Diagnostic Systems." *Journal of Cognitive Engineering and Decision Making*, *16*(2), 63–78.

Ali, M., R. Salehnejad and M. Mansur. (2018). "Hospital heterogeneity: what drives the quality of health care." *European Journal of Health Economics*, *19*(3), 385–408.

Ameen, N., S. Hosany and J. Paul. (2022). "The personalisation-privacy paradox: Consumer interaction with smart technologies and shopping mall loyalty." *Computers in Human Behavior*, *126*(October 2020), 106976.

Ameen, N., A. Tarhini, A. Reppel and A. Anand. (2021). "Customer experiences in the age of artificial intelligence." *Computers in Human Behavior*, *114*(June 2020), 106548.

Anastasiadou, M., V. Santos and F. Montargil. (2021). "Which technology to which challenge in democratic governance? An approach using design science research." *Transforming Government: People, Process and Policy*, *15*(4), 512–531.

Antons, D., C. F. Breidbach, A. M. Joshi and T. O. Salge. (2021). "Computational Literature Reviews: Method, Algorithms, and Roadmap." *Organizational Research Methods*, 1–32.

Aria, M. and C. Cuccurullo. (2017). "bibliometrix: An R-tool for comprehensive science mapping analysis." *Journal of Informetrics*, *11*(4), 959–975.

Ashok, M., R. Madan, A. Joha and U. Sivarajah. (2022). "Ethical framework for Artificial Intelligence and Digital technologies." *International Journal of Information Management*, *62*(November 2020), 102433.

Badger, D., J. Nursten, P. Williams and M. Woodward. (2000). "Should all literature reviews be systematic?" *Evaluation and Research in Education*, *14*(3–4), 220–230.

Bhatia, A., A. Chandani, R. Atiq, M. Mehta and R. Divekar. (2021). "Artificial intelligence in financial services: a qualitative research to discover robo-advisory services." *Qualitative Research in Financial Markets*, *13*(5), 632–654.

Blei, D. M. (2012). "Surveying a suite of algorithms that offer a solution to managing large document archives. Probabilistic topic models."

Blei, D. M., A. Y. Ng and J. B. Edu. (2003). "Latent Dirichlet Allocation Michael I. Jordan." *Journal of Machine Learning Research*, *3*, 993–1022.

Boyd-Graber, J., Y. Hu and D. Mimno. (2017). "Applications of topic models." *Foundations and Trends in Information Retrieval*, *11*(2–3), 143–296.

Cao, J., T. Xia, J. Li, Y. Zhang and S. Tang. (2009). "A density-based method for adaptive LDA model selection." *Neurocomputing*, *72*(7–9), 1775–1781.

Chong, A. Y. L. (2013). "A two-staged SEM-neural network approach for understanding and predicting the determinants of m-commerce adoption." *Expert Systems with Applications*, *40*(4), 1240–1247.

Collins, C., D. Dennehy, K. Conboy and P. Mikalef. (2021). "Artificial intelligence in information systems research: A systematic literature review and research agenda." *International Journal of Information Management*, *60*(June), 102383.

de Visser, E. J., S. S. Monfort, R. McKendrick, M. A. B. Smith, P. E. McKnight, F. Krueger and R. Parasuraman. (2016). "Almost human: Anthropomorphism increases trust resilience in cognitive agents." *Journal of Experimental Psychology: Applied*, *22*(3), 331–349.

Deveaud, R., E. SanJuan and P. Bellot. (2014). "Accurate and effective Latent Concept Modeling for ad hoc information retrieval." *Document Numerique*, *17*(1), 61–84.

Gill, K. S. (2016). "Artificial super intelligence: beyond rhetoric." *AI and Society*, *31*(2), 137–143.

Glikson, E. and A. W. Woolley. (2020). "Human trust in artificial intelligence: Review of empirical research." *Academy of Management Annals*, *14*(2), 627–660.

Haenlein, M. and A. Kaplan. (2019). "A brief history of artificial intelligence: On the past, present, and future of artificial intelligence." *California Management Review*, *61*(4), 5–14.

Hagen, L. (2018). "Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models?" *Information Processing and Management*, *54*(6), 1292–1307.

Hindle, G., M. Kunc, M. Mortensen, A. Oztekin and R. Vidgen. (2020). "Business analytics : Defining the field and identifying a research agenda." *European Journal of Operational Research*, *281*(3), 483–490.

Hirsch, J. E. (2005). *An index to quantify an individual's scientific research output*.

Hoffmann, A. L. (2019). "Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse." *Information Communication and Society*, *22*(7), 900–915.

Huang, M. H. and R. T. Rust. (2018). "Artificial Intelligence in Service." *Journal of Service Research*, *21*(2), 155–172.

Jelodar, H., Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li and L. Zhao. (2019). "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey." *Multimedia Tools and Applications*, *78*(11), 15169–15211.

Khan, A. A., S. Badshah, P. Liang, M. Waseem, B. Khan, A. Ahmad, … M. A. Akbar. (2022). "Ethics of AI: A Systematic Literature Review of Principles and Challenges." *ACM International Conference Proceeding Series*, 383–392.

Khatib, T., A. Mohamed, M. Mahmoud and K. Sopian. (2011). "Modeling of daily solar energy on a horizontal surface for five main sites in Malaysia." *International Journal of Green Energy*, *8*(8), 795–819.

Kordzadeh, N. and M. Ghasemaghaei. (2022). "Algorithmic bias: review, synthesis, and future research directions." *European Journal of Information Systems*, *31*(3), 388–409.

Korteling, J. E. (Hans), G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp and A. R. Eikelboom. (2021). "Human- versus Artificial Intelligence." *Frontiers in Artificial Intelligence*, *4*(March), 1–13.

Lamboglia, R., D. Lavorato, E. Scornavacca and S. Za. (2020). "Exploring the relationship between audit and technology. A bibliometric analysis." *Meditari Accountancy Research*.

Langer, M. and C. J. König. (2021). "Trust in Artificial Intelligence: Comparing trust processes between human and automated trustees in light of unfair bias." *Journal of Business and Psychology*, (January), 1–52.

Makarius, E. E., D. Mukherjee, J. D. Fox and A. K. Fox. (2020a). "Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization." *Journal of Business Research*, *120*(November 2019), 262–273.

Makarius, E. E., D. Mukherjee, J. D. Fox and A. K. Fox. (2020b). "Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization." *Journal of Business Research*, *120*, 262–273.

Marshall, R., A. Pardo, D. Smith and T. Watson. (2022). "Implementing next generation privacy and ethics research in education technology." *British Journal of Educational Technology*, *53*(4), 737–755.

Martin Kunca, M. J. M. and R. V. (2018). "A computational literature review of the field of System Dynamics from 1974 to 2017." *Journal of Simulation*, *12*(2), 115–117.

Meek, T., H. Barham, N. Beltaif, A. Kaadoor and T. Akhter. (2017). "Managing the ethical and risk implications of rapid advances in artificial intelligence: A literature review." *PICMET 2016 - Portland International Conference on Management of Engineering and Technology: Technology Management For Social Innovation, Proceedings*, 682–693.

Meghan Rimol. (2021). "Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation." *Gartner*, 1–3.

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman and A. Galstyan. (2021). "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys*, *54*(6).

Mortenson, M. J. and R. Vidgen. (2016). "International Journal of Information Management A computational literature review of the technology acceptance model." *International Journal of Information Management*, *36*(6), 1248–1259.

Ntoutsi, E., P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M. E. Vidal, … S. Staab. (2020). "Bias in data-driven artificial intelligence systems—An introductory survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), 1–14.

Okoli, C. and K. Schabram. (2010). "Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research." *Working Papers on Information Systems*, *10*(2010).

Peng, C., N. Merat, R. Romano, F. Hajiseyedjavadi, E. Paschalidis, C. Wei, … E. Boer. (2022). "Drivers' Evaluation of Different Automated Driving Styles: Is It Both Comfortable and Natural?" *Human Factors*, *44*(0), 1–38.

Peter Luhn. (1958). "The automatic creation of literature abstracts." *IBM Journal of Research*.

Riazy, S., K. Simbeck and V. Schreck. (2021). *Systematic Literature Review of Fairness in Learning Analytics and Application of Insights in a Case Study*. *Communications in Computer and Information Science* (Vol. 1473 CCIS). Springer International Publishing.

Rowe, F. (2014). "What literature review is not: Diversity, boundaries and recommendations." *European Journal of Information Systems*, *23*(3), 241–255.

Schelble, B. G., J. Lopez, C. Textor, R. Zhang, N. J. McNeese, R. Pak and G. Freeman. (2022). "Towards Ethical AI: Empirically Investigating Dimensions of AI Ethics, Trust Repair, and Performance in Human-AI Teaming." *Human Factors*.

Shadikur Rahman, Syeda Sumbul Hossain, Md. Shohel Arman, Lamisha Rawshan, Tapushe Rabaya Toma, Fatama Binta Rafiq, and K. B. Md. B. (2020). *Assessing the Effectiveness of Topic Modeling Algorithms in Discovering Generic Label with Description*. *Advances in Intelligent Systems and Computing* (Vol. 1130 AISC).

Shobana, V. and N. Kumar. (2015). "Big data - A review." *International Journal of Applied Engineering Research*, *10*(55), 1294–1298.

Sievert, C. and K. Shirley. (2015). "LDAvis: A method for visualizing and interpreting topics," 63–70.

Smacchia, M. and S. Za. (2022). "Artificial Intelligence in Organisation and Managerial Studies: A Computational Literature Review." *ICIS 2022 Proceedings. 6*, 0–17.

Sparrow, R. (2004). "The turing triage test." *Ethics and Information Technology*, *6*(4), 203–213.

Stahl, B. C., J. Timmermans and B. D. Mittelstadt. (2016). "The ethics of computing: A survey of the computing-oriented literature." *ACM Computing Surveys*, *48*(4).

Swanepoel, D. (2021). "The possibility of deliberate norm-adherence in AI." *Ethics and Information Technology*, *23*(2), 157–163.

Tiku, N. (2022). "The Google engineer who thinks the company's AI has come to life." *The Washington Post*, 1–12.

Tranfield, D., D. Denyer and P. Smart. (2003). "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review." *British Journal of Management*, *14*(3), 207–222.

Tsai, C. W., C. F. Lai, H. C. Chao and A. V. Vasilakos. (2015). "Big data analytics: a survey." *Journal of Big Data*, *2*(1), 1–32.

Vakkuri, V., K. K. Kemell and P. Abrahamsson. (2019). "Implementing Ethics in AI: Initial Results of an Industrial Multiple Case Study." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11915 LNCS*, 331–338.

van Giffen, B., D. Herhausen and T. Fahse. (2022). "Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods." *Journal of Business Research*, *144*(January), 93–106.

Vimalkumar, M., S. K. Sharma, J. B. Singh and Y. K. Dwivedi. (2021). ""Okay google, what about my privacy?": User's privacy perceptions and acceptance of voice based digital assistants." *Computers in Human Behavior*, *120*(February), 106763.

Vössing, M., N. Kühl, M. Lind and G. Satzger. (2022). "Designing Transparency for Effective Human-AI Collaboration." *Information Systems Frontiers*, (May), 877–895.

Vu, H. L., K. T. W. Ng, A. Richter and C. An. (2022). "Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation." *Journal of Environmental Management*, *311*(October 2021), 114869.

Wang, F., X. Liang, L. Xu and L. Lin. (2022). "Unifying Relational Sentence Generation and Retrieval for Medical Image Report Composition." *IEEE Transactions on Cybernetics*, *52*(6), 5015–5025.

Xu, L. da, E. L. Xu and L. Li. (2018). "Industry 4.0: State of the art and future trends." *International Journal of Production Research*, *56*(8), 2941–2962.

Yu, Z., Y. Zhang, B. Jiang, C. Y. Su, J. Fu, Y. Jin and T. Chai. (2022). "Distributed Fractional-Order Intelligent Adaptive Fault-Tolerant Formation-Containment Control of Two-Layer Networked Unmanned Airships for Safe Observation of a Smart City." *IEEE Transactions on Cybernetics*, *52*(9), 9132–9144.

Zhu, Y. Q., J. ueline Corbett and Y. Te Chiu. (2021). "Understanding employees' responses to artificial intelligence." *Organizational Dynamics*, *50*(2), 100786.

Zuiderwijk, A., Y. C. Chen and F. Salem. (2021). "Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda." *Government Information Quarterly*, *38*(3), 101577.