JOURNAL OF COMMUNICATION, LANGUAGE AND CULTURE

Anglophone Narratives in Artificial Intelligence: Mapping Research on Cultural Bias and Ethical Concerns in Data-Driven Communication

Valentine Kirimi Muriira[1*], Venoth Nallisamy[1], Jane Gikonyo[2], Hussein Barabwd[3]

[1]Faculty of Education and Liberal Arts, INTI International University, Nilai, Malaysia
[2]Faculty of Education, Tangaza University, Nairobi, Kenya
[3]Faculty of Education, Hadhramaut University, Seiyun, Yemen

*Corresponding author: kirimivalentine@gmail.com; ORCID iD: 0009-0002-6466-5637

ABSTRACT

This study presents a bibliometric analysis of English-language, Scopus-indexed scholarship on cultural bias and ethical concerns in artificial intelligence (AI)-driven communication, covering 1,919 documents published between 2015 and 2025. Using co-citation and co-word analyses conducted in VOSviewer, the study maps the intellectual structure and thematic development of this interdisciplinary field. The findings indicate exponential growth in scholarly output, particularly from 2023 onward, coinciding with the widespread adoption of generative AI tools and large language models. Co-citation analysis identified five thematic clusters: fairness toolkits and justice frameworks; algorithmic bias and word embeddings; explainable AI and algorithmic accountability; critical studies of race and inequality; and philosophical and filtering-based sources of bias. Co-word analysis reveals that while terms such as "artificial intelligence", "algorithmic bias", and "machine learning" dominate the literature, the cultural dimensions of bias are frequently addressed in isolation from technical and ethical frameworks, suggesting a fragmented research landscape in which technical optimisation tends to receive more attention than deeper socio-cultural analysis. This study is limited to English-language publications indexed in Scopus and does not include grey literature, non-English scholarship, or other databases; findings should therefore be interpreted as reflecting English-language, Scopus-indexed research rather than the full body of global scholarship on the topic. The study contributes a structured, data-driven map of a rapidly expanding field and identifies gaps that future interdisciplinary research may address, including the need for integrated frameworks that bridge technical, cultural, and ethical dimensions of AI in communication.

Keywords: artificial intelligence, cultural bias, communication, AI-driven communication, AI ethics, bibliometric analysis

Received: 14 December 2025, **Accepted:** 4 April 2026, **Published:** 30 July 2026

1.0 Introduction

The advent of artificial intelligence (AI) and data-driven technologies has changed how we create, disseminate, and interact with content. AI-generated content, algorithmic media, and large language models (LLMs) are at the forefront of this transformation and are affecting everything from e-commerce social media to e-commerce and more. Large datasets, frequently curated and processed using algorithms that can perpetuate biases (Ntoutsi et al., 2020; Samuel-Okon, 2024), heavily underpin these systems. While AI provides us with unprecedented opportunities for personalisation and efficiency, it also raises severe ethical issues, particularly regarding respect for cultural biases. These biases can be displayed in many ways, such as stereotypes, exclusions, and misrepresentations, all of which have enormous consequences in Anglophone communication (Barnes & Hutson, 2024; Vindigni, 2025).

The central problem is that, although artificial intelligence systems can be objective and impartial, they often are not and instead reflect and perpetuate the societal biases embedded in their training data. For example, AI-powered tools, such as chatbots and content recommendation algorithms, could potentially have a disproportionate effect on one set of demographic groups in favour of others, resulting in an unequal distribution of information and opportunities. This situation not only exacerbates the existing inequalities but also challenges the fundamental principles of fairness and inclusivity that underpin effective Anglophone communication (Baird & Schuller, 2020; Thakur et al., 2024). When not designed properly, AI can create favourable narratives across cultural boundaries (Mohanty, 2025).

To address these issues, it is important to systematically and on a large scale ‘map’ the academic landscape of AI-driven communication and cultural biases. Traditional narrative reviews, although very insightful, are often not as objective and comprehensive as needed to fully appreciate the breadth and depth of research in this interdisciplinary field. A bibliometric approach, on the other hand, enables a data-driven method for discovering trends, key contributors, and thematic clusters within a vast body of literature (Chinta et al., 2024; Ibrahim et al., 2023).

Through an analysis of citation patterns, co-word associations, and co-citation networks, bibliometric analysis helps researchers understand the intellectual structure of a field, meaning that it not only identifies the works that are the most influential to the field but also emerging research themes and gaps in knowledge. This approach provides a more objective, quantifiable assessment of the academic discourse on AI and cultural ethics and produces a clearer picture of how these issues are changing over time and with respect to region or location (Chinta et al., 2024; Robledo-Giraldo et al., 2023).

The major objectives of this study are as follows:

a. To plot the growth curve and publication trends in the field of research on cultural bias and ethics in AI-driven communication.

b. To identify the intellectual foundations and influential works via co-citation analysis to highlight important papers that have shaped the discourse around AI biases.

c. To identify and visualise the dominant thematic clusters and their conceptual relationships using co-word analysis.

d. To identify patterns of cultural bias addressed in AI-related scholarship and outline key areas where further interdisciplinary research is needed to bridge existing gaps in this field.

These objectives will guide the enquiry into the cultural implications and ethical concerns associated with AI to provide a sound framework for addressing future research needs within this important field of study. This paper aims to offer a comprehensive overview of the field, including both a macro-level understanding and a detailed analysis of its intellectual structure and evolution.

2.0 Literature Review

2.1 Conceptual Foundations: Bias and Ethics in AI Systems

Notions such as "cultural bias", "algorithmic bias", and "ethical concerns" are critical for explaining the challenges and implications of AI systems, especially in the domain of communication technologies. Cultural bias is also the unintentional perpetuation of cultural stereotypes and exclusions in AI systems, usually a result of biassed data fed into algorithms or the design of algorithms (Shuford, 2024). Algorithmic bias occurs when AI systems perpetuate and reproduce inequalities and unfair outcomes because of discriminatory data or faulty algorithms; for example, facial recognition systems have been found to correctly identify individuals belonging to marginalised groups (Polo & Ailodion, 2025). AI's ethical issues are complex and can focus on various principles, including accountability, transparency, and justice. The goal of fairness in AI is to create systems that do not treat people differently and are fair to all users. Accountability defines responsibility for AI systems, and transparency is the requirement that AI decision-making processes must be understandable and open to scrutiny. The principle of justice sharpens AI technologies to increase inequalities in our society rather than increasing equal treatment and protection for all groups (Agu et al., 2024; Weiner et al., 2025).

In the case of communication technologies, such ethical issues are very relevant. AI-based media storytelling and social media are playing an ever-growing role in determining how we think, what is considered representative of our culture, and society's dynamics. As AI tools of all sorts (sometimes called "chatbots", "recommendation algorithms", or "automated content generation systems") spread, the need arises to attempt to gauge the impact that manifestations of the bias embedded in such technologies may have on the diversity and inclusivity of the type of narratives that are represented (AbuSahyon et al., 2023).

2.2 The Intersection of AI and Studies of Communication

The impact of AI on communication has been an important area of academic study. Scholars have examined how AI systems frame public narratives, affect representation in culture, and dictate the accessibility of information. AI is even being asked to generate news, moderate content on social media, and create ads (Ibrahim et al., 2023). These systems are powerful in determining conversations in society, and they often negate and reinforce the current biases in society (Huriye, 2023). In addition, there is a rising demand for Anglophone or decolonial approaches to AI. These approaches face the challenge of Western-centric viewpoints taking over and the need to integrate a range of cultural points of view in developing AI (Vindigni, 2025). Decolonial AI frameworks advocate for AI systems to go beyond the biases embodied within algorithms designed by Western care providers and to act and interact with local knowledge, values, and social realities. This shift is to develop AI technologies that are more inclusive and respect cultural diversity and Anglophone views. Such approaches also entail AI being designed and governed with a strong focus on social justice to ensure that the benefits of AI are equitably distributed across all populations (Inuwa-Dutse, 2023; Pasipamire & Muroyiwa, 2024).

This discourse is gaining momentum as the appreciation of AI technologies grows, as they not only shape the future of communication but also the structures within society. By including the Anglophone perspective, researchers are hopeful of creating AI systems that are more in line with the needs of diverse communities so that all voices are heard and represented and the risks and entrenchment of systemic inequalities are addressed (Huriye, 2023).

2.3 Emerging Trends in AI and Cultural Bias Research: A Shift Towards Ethical and Sociotechnical Considerations

The academic discourse on artificial intelligence (AI) and cultural bias has undergone a significant change from the initial technical optimisation to a more interdisciplinary character, with a growing emphasis on ethical issues, especially the aspect of algorithmic fairness and its implications for society. Initially, AI research focused on improving the efficiency and accuracy of algorithms (Bura et al., 2025). However, current research also reveals the need to examine how a particular algorithm can reinforce societal inequalities, especially among marginalised groups. This literature review highlights the importance of fairness, accountability, and transparency in AI systems, and frameworks such as the AI

Fairness 360 toolkit play a central role in managing algorithmic bias (Bellamy et al., 2019). The cultural implications of AI have become a priority among scholars, who encourage a more comprehensive and holistic approach to technical progress through a cultural lens (Shuford, 2024; Sreerama & Krishnamoorthy, 2022).

With this change, there is an emerging acknowledgement of the international repercussions of AI technologies and a demand to identify more varied and inclusive research agendas that consider diverse cultural opinions. An emerging trend is the emergence of decolonial AI models that question the prevailing Western perspective and seek to introduce local knowledge and social backgrounds into the development of AI (Inuwa-Dutse, 2023; Vindigni, 2025). Emerging technologies such as generative AI tools (including GPT-3 and ChatGPT) only add to the increasing importance of addressing cultural biases in the design of data and algorithms because the impact of these tools is widespread across the world (Ezeaka & Umennebuaku, 2024; Gurkan & Suchow, 2024). This changing trend reflects a broader understanding that AI systems must be not only technically competent but also culturally sensitive to make the world a more open and fairer place.

3.0 Methods

This research employs a quantitative descriptive research approach to bibliometric study, which is a systematic mapping of the intellectual structure and thematic development of research on cultural bias and ethics in AI-driven communication. The bibliometric approach makes it possible to conduct a structured and data-driven exploration of the current state of research with a particular focus on identifying trends, key contributors, and nascent themes in this interdisciplinary field. The methodology involves several steps of gathering information from database websites such as Scopus and then cleaning the collected data and doing the next step of data analysis using VOSviewer software (Peng et al., 2024; Saiz-Alvarez, 2024). The following is an elaborate explanation of the research design, data collection strategy, data cleaning, and analysis framework used in this study.

3.1 Research Design and Data Source

The research design employs a quantitative approach of bibliometric analysis to study the academic literature related to AI, cultural bias, and ethics in communication technology. Bibliometric analysis is a well-known method for measuring trends, co-citation networks, and intellectual structures of academic disciplines, especially if the subject matter is spread across different disciplines, such as AI and communication (Bura et al., 2025; Zhang, 2024). The Scopus database was selected as the primary source of data for this research because of its broad and complete database of peer-reviewed publications referring to the computer science, social science, and interdisciplinary study domains. Scopus has excellent indexing of scholarly articles, conference proceedings, and patents, making it an ideal source for bibliometric research, especially in the field of AI and ethics (Chinta et al., 2024).

The information gathered from the metadata is important for publication characteristics such as titles, abstracts, authors, keywords, and citation counts, which are important for performing citation and co-citation analyses (Saiz-Alvarez, 2024). Owing to the nature of the analysis, Scopus, which is known to be one of the leading databases for bibliometric studies, was selected. A wide-ranging set of relevant and, most importantly, high-impact publications will be included in the analysis; therefore, this study provides an overall view of the research landscape (Ibrahim et al., 2023).

3.2 Data Collection Strategy

The strategy of data collection is a systematic and structured approach to collecting all related documents. A thorough search query was devised to reflect the literature at the intersection of AI, cultural bias, communication, and ethics. This query was performed on the Scopus database to identify relevant papers published between 2015 and 2025 which corresponds with the rise of deep learning technologies and increasing Anglophone concerns about AI ethics and biases (Ejjami, 2024; Shuford, 2024).

3.2.1 Search Query Formulation

The search query was designed to identify studies examining cultural bias, algorithmic bias, and ethical issues in artificial intelligence, with particular attention to communication, media, and representational contexts. The search strategy combined four keyword clusters: (1) bias and fairness terms, (2) artificial intelligence and computational technology terms, (3) communication and media-related terms, and (4) ethics and accountability terms.

The following Boolean search string was used:

TITLE-ABS-KEY("cultural bias" OR "algorithmic bias" OR bias OR discrimination OR fairness OR inequality OR representation OR inclusivity OR exclusion OR marginalisation OR stereotype OR prejudice OR disparity OR "digital divide" OR "digital inequality") AND ("artificial intelligence" OR AI OR "machine learning" OR ML OR "deep learning" OR "neural network" OR algorithm OR "large language model" OR LLM OR LLMs OR "natural language processing" OR NLP OR "computer vision" OR "algorithmic system" OR "automated decision" OR "AI system") AND (communication OR media OR narrative OR discourse OR representation OR framing OR storytelling OR content OR information OR "digital media" OR "social media" OR news OR journalism OR broadcasting) AND TITLE-ABS-KEY(ethic* OR "responsible AI" OR "AI ethics" OR "algorithmic accountability" OR “transparency” OR “explainability” OR “fairness” OR “justice”).

This combination was intended to capture studies addressing both the technical and sociocultural dimensions of AI systems. The inclusion of terms such as “cultural bias” and “algorithmic bias” ensured that the search remained focused on AI-related inequalities, while terms related to communication and media aligned the search with the study’s central focus. To strengthen reproducibility, the final review should report the databases searched, search date, publication date range, language restrictions, document types, and any inclusion or exclusion criteria applied during screening.

3.3 Screening and Eligibility Criteria

The initial search results were narrowed using Scopus's built-in metadata filters. Records were restricted by source type and source title to peer-reviewed journals relevant to communication, media studies, artificial intelligence, computer science, ethics, and interdisciplinary social science; by document type to "Article" only, excluding conference papers, book chapters, editorials, and reviews; by publication stage to "Final" publications, excluding preprints and in-press records; by publication year to the period 2015–2025, capturing the rise of deep learning through the current period of generative AI adoption; and by language to English-language publications only. These filters narrowed the initial results to a corpus of documents indexed under source titles and categories relevant to the study's scope, but they operate at the level of publication metadata rather than document content. No manual title-and-abstract screening was performed to independently verify that each of the 1,919 retrieved documents is substantively about cultural bias in AI-driven communication; this is acknowledged as a limitation of the current study.

3.4 Data Screening and Final Dataset

The use of the above filters resulted in a first list of 2,132 documents. After applying the filters (publication years between 2015 and 2025, final stage publications, and English language), the final dataset contained 1,919 documents. The metadata of these 1,919 publications was exported in CSV format to clean and preprocess them further. No further manual screening was performed beyond these metadata filters; the resulting corpus reflects documents matching the search string and bibliographic filters described above rather than a set independently verified for topical relevance to cultural bias and AI-driven communication, as addressed in the Limitations of this study. This dataset was used as the basis for the bibliometric analysis and visualisation (Shuford, 2024).

3.5 Cleaning and Preprocessing of the Data

To ensure precision and reliability in the bibliometric analysis, the data were cleaned and pre-processed. This step was necessary to ensure that the co-word and co-citation networks generated were meaningful and representative of real trends in the literature. Thesaurus File Creation: The thesaurus file was created for the standardisation of keywords by combining synonymous terms/variants. For example, terms such as "AI" and "artificial intelligence", "LLM" and "large language model", and "machine learning" and "ML" were put together to avoid fragmentation in the analysis. This process prevents the issue of analysis underestimating the importance of some concepts due to variations in terminology (Bura et al., 2025). Application in VOSviewer: The thesaurus file was applied inside VOS, which allowed the consolidation of the terms and the creation of a unified dataset to be used in the network analysis. This ensures that phrases with similar concepts are represented as a single entity, resulting in more precise thematic network representation (Saiz-Alvarez, 2024).

3.6 Analytical Framework and Tools

The cleaned data were then imported into VOSviewer version 1.6.20 for network construction and visualisation. Three main bibliometric techniques were used for data analysis:

a. *Trend Analysis:* This technique involved investigating the trend in the number of publications over time, identifying the leading countries, institutions, and journals, as well as examining the development of research themes. This analysis was performed using simple analytics provided by Scopus, and the findings were presented in the form of tables and figures to observe the important trends and changes in the research landscape (Bura et al., 2025).

b. *Co-citation Analysis:* Co-citation analysis was used to explore the intellectual foundations of the field by identifying pairs of documents that are frequently cited together. This technique is effective in mapping scholarly networks and identifying the most influential papers in the field of AI ethics and communication research (Zhang, 2024).

c. *Co-word analysis*: Co-word analysis was used to determine the conceptual structure of the field. The presence of thematic clusters in the literature is identified using the co-occurrence of author keywords across the dataset analyses. VOSviewer was used to create a network map to visualise the strength of associations between keywords to help identify dominant research themes, such as algorithmic bias, fairness, and cultural representation in AI (Shuford, 2024).

The systematic process of document identification, screening, and formation of the final dataset is summarised in Table 1, which includes information about the initial search and the use of progressive filters (publication year, stage, and language) and analytical thresholds used for the Scopus-derived corpus.

Table 1

Document Identification and Screening Process

Step	Process / Filter	Criteria / Parameters	Result (Documents)
Initial Search	Search query executed in Scopus	The search string is the following: TITLE-ABS-KEY("cultural bias" OR "algorithmic bias" OR bias OR discrimination OR fairness OR inequality OR representation OR inclusivity OR exclusion OR marginalisation OR stereotype OR prejudice OR disparity OR "digital divide" OR "digital inequality") AND ("artificial intelligence" OR AI OR "machine learning" OR ML OR "deep learning" OR "neural network" OR algorithm OR "large language model" OR LLM OR LLMs OR "natural language processing" OR NLP OR "computer vision" OR "algorithmic system" OR "automated decision" OR "AI system") AND (communication OR media OR narrative OR discourse OR representation OR framing OR storytelling OR content OR information OR "digital media" OR "social media" OR news OR journalism OR broadcasting) AND (ethic* OR "responsible AI" OR "AI ethics" OR "algorithmic accountability" OR “transparency” OR “explainability” OR “fairness” OR “justice”)	n = 2,132
Primary Screening	Application of Scopus refinement filters	• Publication Year: 2015 to 2025 • Publication Stage: "Final" • Language: English Excluded 113 Document	n = 1,919
Final Dataset	Data export for analysis	CSV export of complete metadata for the 1,919 eligible documents.	Final Corpus: n = 1,919
Analytical Scope	VOSviewer network thresholds	• Co-word Analysis: Minimum keyword occurrence = 10 • Co-citation Analysis: Minimum citation count = 5	Focus on significant nodes within the corpus

4.0 Results

4.1 Trend Analysis

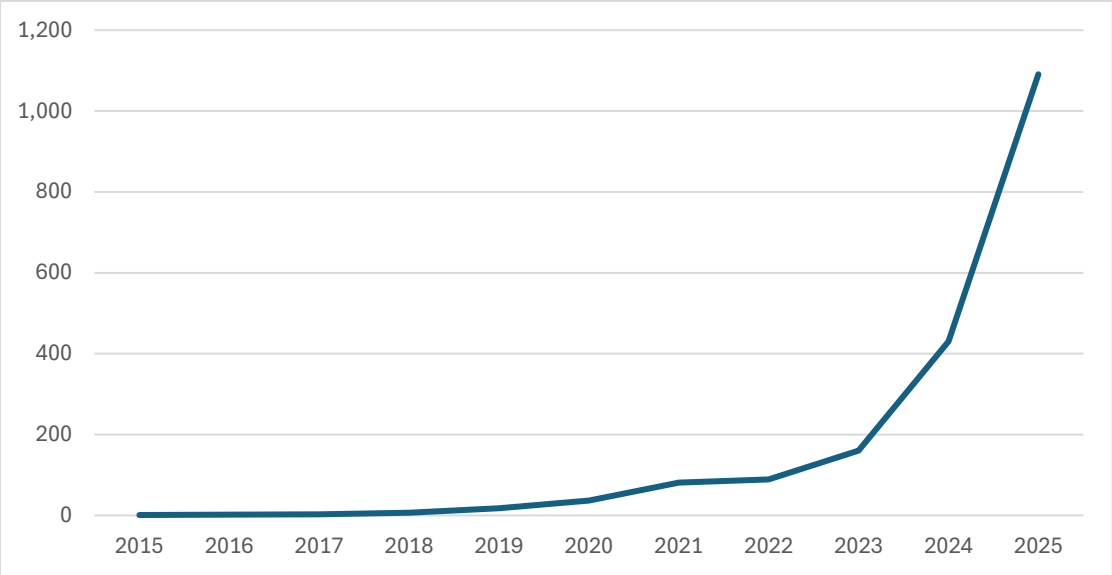
4.1.1 RO1. To plot the growth curve and publication trends in the field of research on cultural bias and ethics in AI-driven communication.

This transformation of academic interest in cultural bias and ethical issues in data-driven communication and AI can be well visualised based on the publication trends from 2015 to 2025. This analysis is based on data from Scopus and shows a dramatic and increasing growth trend in scholarly data, indicating the rapid development and increasing importance of the field (Shahzad et al., 2024; Tao

et al., 2025). Figure 1 displays the number of documents published in a year and within the scope of this study.

Figure 1

Trend Analysis



Note. Data retrieved from the Scopus database.

This trend can be divided into three distinct phases:

Nascent Phase (2015-2019): The output of scholarly work was minimal but showed early signs of growth in this phase. The field was in the beginning stages, with publications growing progressively from single digits to 18 documents in 2019. This phase is likely to be the earliest stage of the academic realisation of the societal consequences of AI, which came about around the time of the early debates about algorithmic bias following high-profile cases and the expansion of machine learning model capacities (Barnes & Hutson, 2024).

Accelerating Phase (2020-2022): A very pronounced upward trajectory started in 2020, and publications more than doubled compared to the previous year, reaching 37. This acceleration is still rather sharp for 2021 (81 documents) and 2022 (89 documents). This stage spans the events that occurred around the globe and technological leaps, all of which brought ethics to the forefront of public discourse and academia. The heavy reliance on digital communication due to the covid-19 pandemic, along with the widespread use of social media and content moderation AI, and the growing interest from the public and regulatory bodies in large language models (such as the release of GPT-3 in 2020), has led to a surge in research interest in the cultural and ethical aspects of data-induced systems (Arjanto et al., 2025).

Exponential Growth Phase (2023-2025): The trend enters a period of explosive growth starting in 2023 with 160 documents, increasing to an unprecedented 430 documents in 2024, and projected to reach 1091 documents by 2025. This is an increase of almost seven times compared with 2022 to 2025. This exponential increase is the major result of the trend analysis. This suggests that research into cultural bias and artificial intelligence ethics within communication is moving out of a specialised niche and becoming a central, urgent, and mainstream research area. This phase is likely a manifestation of the availability and use of generative artificial intelligence (AI) tools (for example, ChatGPT and Midjourney) by the Anglophone public in late 2022 and 2023, causing a flurry of interdisciplinary

research into the deep implications of generative AI in issues of cultural representation, misinformation, privacy, and ethical governance (Ezeaka & Umennebuaku, 2024; Gurkan & Suchow, 2024).

4.2 Co-*citation Analysis*

4.2.1 RO2. To identify the intellectual foundations and influential works via co-citation analysis to highlight important papers that have shaped the discourse around AI biases.

Co-citation analysis was carried out using the minimum criteria of 10 co-citations from the 6050 references cited, with 55 meeting these criteria and showing their significance in the respective field. The analysis showed five different clusters representing different thematic areas: Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue), Cluster 4 (yellow), and Cluster 5 (purple). The top 10 most cited references in this analysis refer to influential works that have shaped the current discourse on AI's cultural bias and ethics. Leading the list is (Buolamwini & Gebru, 2018), Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. With 51 citations, it is among the top 10 papers that have been instrumental in raising awareness of gender and racial biases in AI systems. Close behind it is (Dwork et al., 2012) (46 citations), which provides a starting point for the legal framework to tackle the issue of algorithmic fairness. Other important contributions include Benjamin's (2019) Race After Technology: Abolitionist Tools for the New Jim Code (27 citations), in which the author critiques the racial inequalities that are coded into occupational AI technologies, and Chouldechova and Roth's (2018) Fairness and Machine Learning (26 citations) which sets out a framework to ensure fairness in AI.

Table 2 presents the ten most frequently co-cited references identified in the co-citation analysis, ranked by citation count and total link strength, reflecting the seminal works that have most significantly shaped scholarly discourse on cultural bias and ethics in AI-driven communication within the Scopus-indexed corpus. Co-citation network map of the 55 most frequently co-cited references (minimum co-citation threshold = 5) visualised using VOS Viewer. Node size reflects citation frequency; proximity and colour indicate cluster membership across five identified thematic groups.

Table 2

Top 10 Most-Cited References, Sorted by Citation Count and Total Link Strength

Rank	Cited Reference	Citations	Total Link Strength
1	(Buolamwini & Gebru, 2018)	51	70
2	(Dwork et al., 2012)	46	57
3	(Benjamin, 2019)	27	47
4	(Chouldechova & Roth, 2018)	26	41
5	(Barredo Arrieta et al., 2020)	17	17
6	(Bender et al., 2021)	17	17
7	(Hoffmann et al., 2018)	16	28
8	(Caliskan et al., 2017)	14	19
9	(Bellamy et al., 2019)	14	17
10	(Burrell, 2016)	13	29

Cluster 1 (Red) – Fairness Toolkits, Definitions, and Justice Perceptions

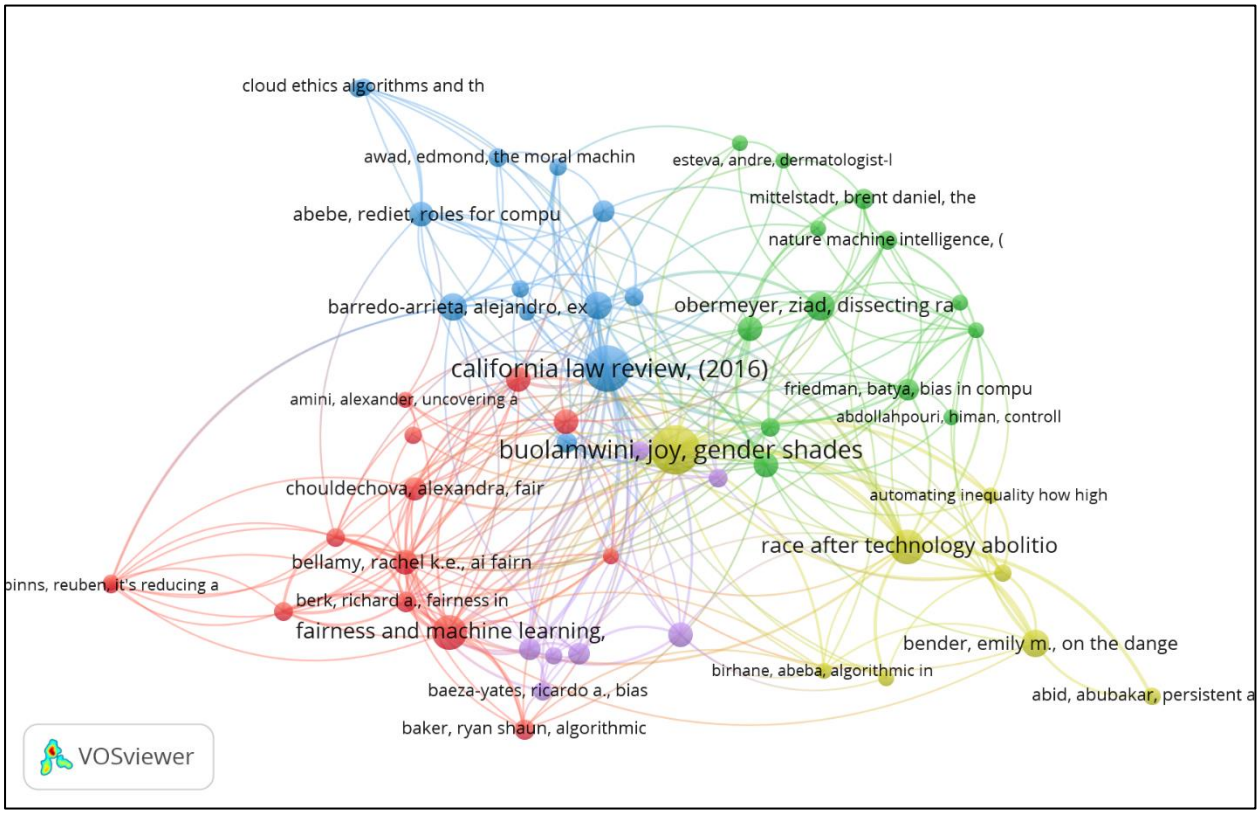
Cluster 1 (Red) is mostly concerned with fairness toolkits and definitions of fairness and justice perceptions when applied to AI systems. At the centre of this cluster are important pieces that focus on how to mitigate algorithmic bias through the use of software tools and frameworks. One of the most important contributions is called AI Fairness 360 (Bellamy et al., 2019), which addresses issues of fairness in AI systems. Similarly, Bird et al. (2020) proposed another major toolkit for measuring and rendering machine learning models. These toolkits have played a significant role in defining the meaning of fairness in AI, and their impact is likely to increase as AI systems are increasingly used in sensitive areas such as education, healthcare, and criminal justice.

Cluster 2 (Green) - Algorithmic Bias, Oppression, and Basic Works of Word Embeddings

Cluster 2 (Green) is a cluster of foundation works which address the question of the origin of algorithmic bias, the role of AI in perpetuating oppression, and the role of word embeddings through the perpetuation of social biases. One of the more influential papers published from this family of research is Caliskan et al. (2017), which demonstrates the numerous human-like biases that machine learning models trained on large language corpora are able to develop, which is ground-breaking for the field of AI ethics. This work paves the way for other studies on how AI systems can perpetuate societal inequalities. Bolukbasi et al. (2016) also added to this discussion by showing how biases against gender and race are programmed into natural language processing word embeddings, which in turn leads to the reinforcement of harmful stereotypes in AI applications.

Figure 2

Co-Citation Analysis Using VOSviewer Mapping



Note. Generated using VOSviewer.

Cluster 3 (Blue) Explainable Artificial Intelligence (XAI), Algorithmic Accountability & Societal Role of Computing

Cluster 3 (Blue) is around the nascent field of Explainable AI (XAI); as well as around algorithmic accountability in general. (Adadi & Berrada, 2018) survey on XAI provides an overview of the field, such as how to make machine learning models more interpretable not only to its developers, but to its end-users. This is important in situations in which AI systems are used to impact important decision-making processes such as in healthcare, finance and criminal justice. The focus on explainability can also be realised in (Burrell, 2016) work, dealing with the fact of opacity in the machine learning algorithms with its consequences for accountability and transparency.

Cluster 4 (Yellow) - A.I., Language Models and Race/Inequality in Critical Studies

Cluster 4 (Yellow) is critically focused on the societal implications of AI and racial and inequality strife in particular. A seminal work in this cluster is (Buolamwini & Gebru, 2018) Gender Shades that sheds light on the intersectional disparities of accuracy in the commercial gender classification systems and how they amplify painful gender and racial disparities that have long existed in society. This study has become a cornerstone in the conversation of how AI technologies are required to be made for different populations to take into account multiple people of different backgrounds, but focusing on making AI systems not only fair, but also inclusive (Bender et al., 2021).

Cluster 5 (Purple) - Sources of Algorithmic Bias and Sources of fairness Philosophy Bias in Web/Filtering

Cluster 5 (Purple) in particular, the philosophical basis to how fairness applies to AI, and where algorithmic bias is coming from particularly in the area of web filtering and personalization. (Akter et al., 2021) research into algorithmic bias in data-driven innovation shows the impact algorithmic bias can have on such disparate industries as marketing and healthcare. This cluster also deals with the role play and relevance of AI in web filtering, and personalization, (Baeza-Yates, 2018) work on bias on the web will be a key reference Baeza-Yates discusses how search engines and systems that make recommendations are able to perpetuate bias by favouring one type of content over others, putting an iron on the societal cracks that already exist (Bozdag, 2013).

4.3 Co-Word Analysis

4.3.1 RO3. To identify and visualise the dominant thematic clusters and their conceptual relationships using co-word analysis

A co-word analysis was conducted to identify the key themes in the literature on artificial intelligence (AI), fairness, algorithmic bias, ethics, and decision-making. A minimum co-occurrence threshold of 10 was applied. Of the 4,667 keywords identified, 54 met this threshold. These keywords were grouped into four clusters: Cluster 1 (red), Cluster 2 (green), Cluster 3 (blue), and Cluster 4 (yellow). Table 3 presents the 10 most frequently occurring keywords, representing central topics in the field. These were “Artificial Intelligence” (297 occurrences), “Algorithmic Bias” (261), “Algorithmics” (224), “Machine Learning” (197), “Decision-Making” (95), “Human” (92), “Machine-Learning” (90), “Ethical Technology” (80), “Ethics” (77), and “Fairness” (75). Figure 3 presents the co-word network visualisation, illustrating the relationships among these keywords and their thematic clusters.

Table 3 lists the ten most frequently occurring author keywords in the co-word analysis, ranked by occurrence count and total link strength, and provides an indication of the dominant conceptual themes within Anglophone, Scopus-indexed scholarship on AI bias and ethics in communication. Figure 3 presents the co-word network map generated in VOSviewer, visualising the thematic structure of the corpus through the co-occurrence patterns of the 54 keywords that met the minimum occurrence threshold, organised into four distinct clusters.

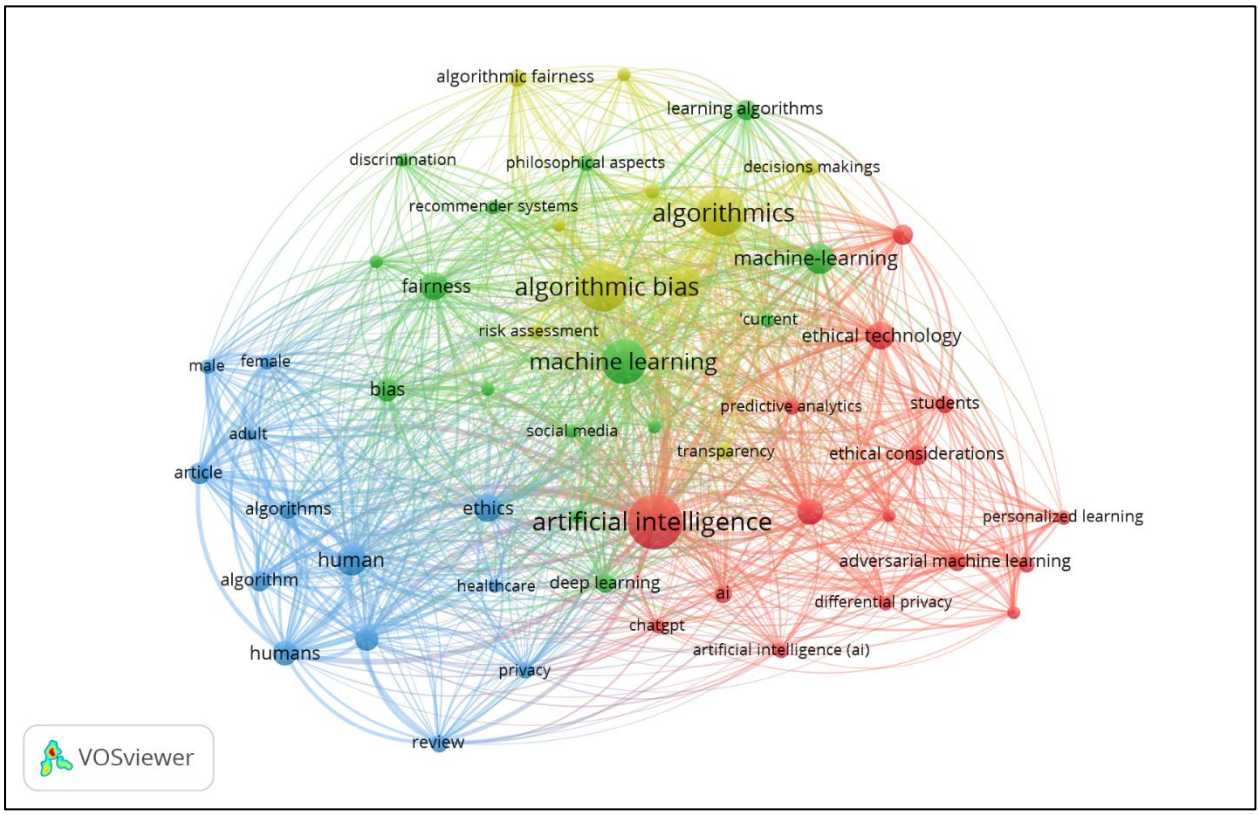
Table 3

Top 10 Most Frequently Occurring Keywords, Sorted by Occurrence and Total Link Strength

Rank	Keyword	Occurrences	Total Link Strength
1	artificial intelligence	297	1142
2	algorithmic bias	261	887
3	algorithmics	224	858
4	machine learning	197	864
5	decision making	95	454
6	human	92	600
7	machine-learning	90	429
8	ethical technology	80	368
9	ethics	77	368
10	fairness	75	349

Figure 3

Co-Word Mapping



Note. Generated using VOSviewer

Cluster 1 (Red) - Artificial intelligence technologies Learning Methods Privacy

Cluster 1 (Red) the intersection of AI technologies and learning methods and privacy reflects the growing importance of the development of AI systems that are not just advanced but ethical systems. Key words such as Artificial Intelligence (297 occurrences), Adversarial Machine Learning (37 occurrences), and ChatGPT (21 occurrences) are indicative of the fast-evolving nature of AI technologies, particularly in terms of Natural Language Processing and Machine Learning. This cluster reflects the emergence of the data privacy topic (67 occurrences) and scholars discussing the tension in AI systems that they need to strike between performance and preserving sensitive user privacy information. Studies such as Sreerama & Krishnamoorthy (2022) discuss the ethical implications of artificial intelligence systems in ensuring fairness and preventing bias, especially in fields such as predictive analytics (25 occurrences), which are increasingly applied in fields such as healthcare, finance, and education (Sreerama & Krishnamoorthy, 2022). Similarly, Konidena et al. (2024) discussed the integration of federated learning (20 occurrences) as a measure of maintaining privacy by decentralising data processing, where personal information can be exposed during activity training processes (Konidena et al., 2024).

Cluster 2 (Green) - Bias, Fairness and Machine learning Applications

Cluster 2 (Green) takes into consideration the important elements of bias, fairness and the application of machine learning (197 occurrences) in different sectors. This cluster focuses on ethical issues related to ensuring that machine learning models do not further societal inequalities through algorithmic bias (261 occurrences), cultural bias (19 occurrences), and gender bias (20 occurrences). The inclusion of deep learning (38 occurrences) and discrimination (18 occurrences) support the fact that these artificial intelligence models have the probability of disproportionately reproducing bias if left unmonitored and uncorrected. Researchers such as Bozdag (2013) and Sreerama and Krishnamoorthy (2022) provide insight into the mechanisms of detecting and mitigating the impact of bias in machine learning models by emphasising the importance of fairness frameworks in AI to mitigate the risk of discriminatory outcomes (Awad et al., 2018). Similarly, Osasona et al. (2024) delve into the concept of making AI systems more equitable through fairness-aware algorithms, especially in areas where biassed decisions can have long-spanning social consequences, such as healthcare and criminal justice.

Cluster 3 (blue) - Human centric Studies, Healthcare and General Ethics

Cluster 3 (Blue) is about human-centric studies, healthcare, and general ethics in AI. Key terms such as human (92 occurrences), ethics (77 occurrences), and algorithm bias (54 occurrences) signal societal problems with the use of AI in key areas such as healthcare and law enforcement. As AI is increasingly used in decision-making, particularly in healthcare, there is a need to ensure that it is compatible with ethical standards and does not perpetuate existing inequalities. (Bird et al. (2020) and Chinta et al. (2024) provide an extensive analysis of why a fairness-aware algorithm can be implemented for healthcare applications of AI so that patient care without biases to ensure healthcare outcomes is fair for all the various demographic groups (Chinta et al., 2024). Furthermore, Ratna et al. (2024) highlighted the importance of transparency in algorithmic decision-making processes and how the ability to explain the decision-making processes of AI algorithms can help build trust and accountability in industries such as healthcare, where decisions can potentially have a direct impact on human lives.

Cluster 4 (Yellow) - Algorithmic Bias, Decision and Societal Impact

Cluster 4 (Yellow) is on the impact of algorithmic bias and decision-making on the wider societal effects of AI technologies. Key keywords such as ’algorithmic bias’ (261 occurrences) and ’decision-making’ (95 occurrences) point to the important role that AI plays in high-stakes decision-making processes, especially within sectors such as criminal justice, finance, and healthcare. These sectors heavily rely on AI to make decisions that impact the lives of people, which is why it is important to ensure that these systems are fair and transparent. Bura et al., 2025) focused on the ethics of AI and the way this occurs through decision-making and a call for ethical frameworks to be strong in order to avoid AI being able to perpetuate systemic inequalities in industries like healthcare and law enforcement sectors (Bura et al., 2025).

5.0 Discussion and Implications

5.1 Discussion

Some of the important emerging trends identified are as follows: From technical to sociotechnical The huge number of publications reflects this very obvious change in the landscape of AI research, from a focus on publications dealing with either purely technical and optimising issues to a very strong and interdisciplinary interaction with sociotechnical problems. The focus has shifted from "how AI works" to "how AI impacts society and culture" (Baldin, 2024). The first growth is reactive to proactive Scholarship The appears reactive to technological milestone incidents in society. However, the projected volume for 2024 and 2025 predicts that the field is maturing, with scholars taking proactive measures to define the research agenda, anticipate risks, and propose frameworks for ethical AI development and deployment in Anglophone communication situations (Bura et al., 2025).

Consolidation of a Definitive Research Field The sheer scale of what has been or is on the horizon attests to the fact that "cultural bias and ethical concerns in data-driven communication" is no longer a sub-topic but is a well-defined and critically important field of study in its own right, attracting significant resources and scholarly attention globally[S1] (Peng et al., 2024). This trend of publication is a quantitative backbone for the study and clearly speaks about the timeliness and criticality of the research topic. The following co-citation and co-word analyses trace the intellectual form and theme development of this exponentially growing body of literature.

The co-word analysis also suggests important paradigm changes in AI ethics in that there are far more sociotechnical concerns besides merely the technical concerns of algorithmic accuracy, efficiency, and concern with human rights, privacy, and representation. As illustrated by Patel (2021), these shifts are connected to the rising popularity of generative AI systems such as GPT-3 and ChatGPT, which have caused the public to be aware of the perceived dangers of biased AI systems (Aninze, 2024). The analysis also revealed that most keywords and themes have evolved from technical optimisation to ethical considerations and are related to historical milestones such as the emergence of LLMs and their implications for societal representation. This is evidenced by the growing number of scholarly publications over the last five years from 2020 to 2025, marking a period of exponential growth in research relating to the societal impacts of AI (Wakunuma & Eke, 2024).

The results obtained in this study showcase major novel tendencies and unmet gaps in AI ethics. Although the findings are properly presented, it is vital to consider the bigger picture of these findings, especially in relation to the cultural and ethical aspects of AI. The research highlights trends such as the exponential increase in the number of research publications, the growing interdisciplinary orientation, and the emergence of decolonial AI frameworks, all of which demonstrate the criticality of considering ethical aspects within AI systems, particularly in the context of algorithmic bias and fairness (Bura et al., 2025; Chinta et al., 2024).

New trends indicate a shift in AI research towards a more sophisticated approach, which considers the influence of such technologies on society. In addition to the technical optimisation of AI systems, researchers have raised issues regarding their ethical and cultural implications. This change emphasises the increased acceptance of AI as a sociotechnical system, which must be subject to governance frameworks that consider the cultural diversity, Anglophone views, and moral accountability of developers (Chouldechova & Roth, 2018; Shuford, 2024). Nevertheless, despite this development, the question of how these problems are dealt with is still highly fragmented, with technical issues of fairness and bias being frequently discussed outside the cultural aspect. This is one of the gaps in the research field which should be bridged.

To further develop this analysis, it is necessary to discuss how these new trends in AI ethics influence the future of this field. The article postulates that although the emphasis on fairness models such as AI Fairness 360 has increased, cultural and ethical issues have become separated in discussions instead of being considered as part of AI design (Bura et al., 2025; Shuford, 2024). The findings indicate that there is an acute lack in the field: the necessity to have a more holistic, integrated approach that integrates both the technical and cultural aspects into the governance of AI technologies.

In the future, AI studies should fill such gaps by creating more comprehensive frameworks that can close the gap between technical optimisation and cultural ethics. One of the main future directions must be to comprehend the role of AI in world communication and its effect on marginalised groups, especially the ability of AI to support already existing inequalities (Vindigni, 2025). The need to include decolonial views and frameworks is vital to ensure that AI technologies do not undermine different cultural settings and values (Huriye, 2023; Inuwa-Dutse, 2023).

Finally, this study promotes the idea of being proactive in the field of AI research, where ethical considerations are not only addressed in response to problems but are proactively incorporated into the research process. This proactive strategy is vital in addressing possible harms and making AI systems technically competent and ethically sound, inclusive, and fair (Baird & Schuller, 2020; Gurkan & Suchow, 2024).

5.2 Practical and Societal Implications

From a practical perspective, the results of this study provide important lessons on the responsibility of AI engineers, product managers, and policymakers in shaping the future of AI systems. For AI developers, the existence of clusters pertinent to bias and fairness, especially in the context of data privacy and machine learning models, means that there will need to be proactive steps that are taken in the design of systems to avoid such amplification of societal biases. Konidena et al. (2024) emphasise the importance of interdisciplinary approaches, considering algorithmic transparency and bias mitigation techniques to be embedded in the development process to prevent unethical outcomes in high-risk applications such as recruitment algorithms and predictive policing. Furthermore, the attention to diverse datasets and fairness to AI models reflects the fact that it has now become increasingly recognised that considerations of ethical implications of AI cannot be investigated in isolation of each but as part of an integrated design process, which considers not just the technical performance but also the social responsibility of AI (Barnes & Hutson, 2024).

For policymakers, this study is a call for urgent action to regulate artificial intelligence systems with far-reaching consequences for public welfare. The development of widespread and rigors regulatory frameworks to create standards for AI transparency and accountability is essential for the proliferation of ethical discourse on algorithmic fairness and bias auditing. (Singh Chadha, 2024) the need to have clear standards to assess and manage bias in AI systems, especially in areas where algorithmic decisions have direct impacts on people's lives in sector such as credit scoring and healthcare decision making (Singh Chadha, 2024). Additionally, Gurkan and Suchow (2024) call for a more inclusive approach to AI governance, in which regulators seek to actively engage with diverse cultural perspectives to ensure that Anglophone ethics are considered when making AI policy decisions.

The geographical distribution of ethics research in AI also affects how Anglophone discussions define bias and ethics in AI. Research has shown that Western countries, mainly the U.S., Germany, and the U.K., are in the driver's seat in terms of communicating about the ethics of AI, while that of the Anglophone South is marginalised. This uneven distribution poses important questions about whose cultural perspectives are most important in the direction of ethical governance of AI. (Curto et al., 2024) in their study titled argue that this disparity reinforces current power structures, where Western ideologies are the sole design and governance in terms of AI design and governance, potentially sidelining the needs of Western society and non-Western values. The findings of this research highlight the importance of cooperation among different parts of the world to ensure that the development of AI is representative of the world's population, considering the diversity of cultures. This inclusion worldwide will be one of the keys to creating a fairer future for AI technology.

6.0 Conclusion, Limitations, and Future Avenues

6.1 Conclusion

This study provides the first general bibliometric picture of the interdisciplinary research landscape at the intersection of artificial intelligence (AI), cultural bias, and communication ethics. Using a quantitative bibliometric method, trends and intellectual structures of new and emerging themes from publications between the years 2015-2025 are analysed. The analysis revealed several important facts.

First, the existence of three 'big clusters' (albeit poorly connected) centred around the technologies of AI, bias, and fairness shows how fragmented the field of AI can be. Second, there is emerging literature linking AI ethics and algorithmic bias, but these are often dealt with separately which hints at conceptual gaps that need to be filled. Finally, the rise of technologies such as Generative AI and LLMs (for example, ChatGPT) has been correlated with a dramatic rise in scholarly interest in ethics and bias prevention (Birhane, 2021; Grincevičienė et al., 2019).

The main contribution of this study is to provide an empirical basis for navigating this rapidly evolving field. The bibliometric analysis provides insight into how the field develops, including quantitative analysis of the growth trajectory, finding thematic clusters, and finding the intellectual network, to understand the growth of the field of research on AI-powered communication and the increasing importance of the ethical and cultural factors of AI design and implementation (Wakunuma & Eke, 2024).

6.2 Limitations of the Study

While this study provides some interesting findings, it is important to recognise some of the weaknesses unusual to the study: First, the decision to use Scopus as a primary database, although a comprehensive database, has certain inherent biases. For example, the type of books not included in the database could be conference proceedings, which are common for computer science research, and those published in books or book chapters (common in the humanities). Studies such as Gurkan and Suchow (2024) address this difficulty and suggest that more datasets should be used to cover the entire variety of research contributions.

Second, the scope of the study is limited by the search string and may not identify all relevant studies. For example, terms such as "digital inequity" or "algorithmic accountability" would have missed important studies that used different terms. As stated by (Ramamoorthy, 2025) keyword dependency is a frequent weakness of the bibliometric approach which could be subject to the risk of closing down potentially relevant research because of the difference in language or focus.

A temporal lag also impacts the findings, as publications until a certain point are considered in the analysis. Emerging research from 2024 and 2025, which may show the most breaking research, is unlikely to be cited in full and not in the integrated networks. As Molla and Ahsan (2025) pointed out, the introduced research will require some time to gain momentum and manifest in bibliometric analysis. Lastly, bibliometric mapping attempts to measure associations and trends, but does not focus on the quality and validity of the research itself. This means that in whilst the analysis does provide some information on the prevalence of certain topics does not address the robustness and rigour of the studies that are involved this is identified by (Singh et al., 2024)) in their study on AI in customer retention

Furthermore, possible regional or lingual biases in the dataset should be critically reflected. Similar to many other academic databases, Scopus is over-represented in studies of Western countries, which can impact the general outlook of the dataset. More diverse databases or studies in non-Western areas would be beneficial in reducing such bias and offering a more Anglophone perspective on the field. In addition, non-English publications might not have been included in the research, which might lead to the exclusion of research in areas where English is not the dominant language. This restriction underscores the need to consider linguistic diversity and geographical variations when conducting bibliometric research.

6.3 Future Research Avenues

Given the gaps and findings identified, several directions for future research can be taken as actionable. Substantive gaps in the literature, namely, the decontextualization between cultural bias and ethical governance, point out the need for new integrated frameworks to address the technical and cultural issues of AI systems (Curto et al., 2024). Future studies should attempt to connect such clusters using bias mitigation methodologies from technical clusters and critical cultural frameworks from ethical governance clusters, as in the study by Shahzad et al. (2024) on the ethics of AI in academic research.

Methodological extensions could take the form of qualitative reviews or systematic research of the best-cited papers in Generative AI research, especially those which include discussions of ethical solutions and bias mitigation strategies. Ramamoorthy 2025) suggests a deeper qualitative analysis that can be used to create a better perception of evaluation techniques for generative AI systems. In light of the database bias, replication of this analysis with the Web of Science or Google Scholar could be performed in future research to obtain a broader range of research. Integrating grey literature in seminal AI ethics institutes such as Gurkan and Suchow (2024) to ensure an Anglophone and comprehensive perspective on the ethics of Artificial Intelligence.

Finally, one possible area for future research is the new relationship between multimodality and cultural bias. The increasing application of AI in creating videos and audio requires greater exploration of how these technologies can reinforce or debunk cultural stereotypes. Studies such as Bahroun et al. (2023) can be seen as an indicator of the need for research on the effect of generative AI on cultural representation in the media as well as in education. This is an area with many opportunities for exploration as multimodal AI systems become increasingly sophisticated and integral to media production. In conclusion, this study identified important trends, challenges, and avenues for future research on AI's cultural bias and ethical governance (Shukla, 2024; Wakunuma & Eke, 2024). Through careful consideration of these future research directions, scholars and practitioners can further refine the integration of AI technologies into society to ensure their effectiveness and equity.

Acknowledgement

The authors would like to extend their heartfelt gratitude to INTI International University University for their contribution.

Conflict of Interest

The authors have declared that no competing interests exist.

Author Contribution Statement

VKM: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation. VN: Project Administration, Writing – Review & Editing. JG: Project Administration, Supervision, Writing – Review & Editing. HB: Project Administration, Supervision, Writing – Review & Editing.

Funding

This research received no external funding.

Ethics Statement

This research did not require IRB approval because it used publicly available secondary data with no identifying personal information.

Data Access Statement:

Research data supporting this publication are available upon request to the corresponding author.

Author Biography

Valentine Kirimi Muriira is a postgraduate student at INTI International University, Malaysia, and a researcher and researcher trainer at Lasalle Technical Training Institute, Karen, Nairobi. Her research interests include intercultural communication, corpus linguistics, language identity, and Instructional design.

Venoth Nallisamy, PhD, is a Senior Lecturer in the School of Education at INTI International University, Nilai, Malaysia. His research interests include educational technology, instructional design, curriculum development, and online learning environments. He has published in international journals and continues to support ongoing research in the field of education.

Jane Gikonyo, PhD, is Associate Dean of the School of Education and an education consultant at Tangaza University, Nairobi, Kenya, and a researcher trainer at Lasalle Technical Training Institute, Karen, Nairobi.

Hussein Barabwd, PhD, is a Senior Lecturer in the School of Education at Hadhramaut University, Mukalla, Yemen.

References

AbuSahyon, A. S. A. E., Alzyoud, A., Alshorman, O., & Al-Absi, B. (2023). AI-driven technology and chatbots as tools for enhancing English language learning in the context of second language acquisition: A review study. *International Journal of Membrane Science and Technology*, *10*(1), 1209-1223. https://doi.org/10.15379/ijmst.v10i1.2829

Adadi, A., & Berrada, M. (2018). Peeking inside the black box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Agu, E. E., Abhulimen, A. O., Obiki-Osafiele, A. N., Osundare, O. S., Adeniran, I. A., & Efunniyi, C. P. (2024). Discussing ethical considerations and solutions for ensuring fairness in AI-driven financial services. *International Journal of Frontline Research in Multidisciplinary Studies, 3*(2), 1–9. https://doi.org/10.56355/ijfrms.2024.3.2.0024

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management, 60*, Article 102387. https://doi.org/10.1016/j.ijinfomgt.2021.102387

Aninze, A. (2024). Artificial intelligence life cycle: The detection and mitigation of bias. *International Conference on AI Research, 4*(1), 40–49. https://doi.org/10.34190/icair.4.1.3131

Arjanto, P., Makulua, I. J., Sampe, P. D., Huliselan, N., & Ellis, R. (2025). Augmenting human connection: A systematic review of artificial intelligence in counseling practices, ethics, and cultural adaptation. *Jurnal Bimbingan Dan Konseling Pandohop, 6*(1), 1–15. https://doi.org/10.37304/pandohop.v6i1.20310

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

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM, 61*(6), 54–61. https://doi.org/10.1145/3209581

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability, 15*(17), Article 12983. https://doi.org/10.3390/su151712983

Baird, A., & Schuller, B. (2020). Considerations for a more ethical approach to data in AI: On data representation and infrastructure. *Frontiers in Big Data, 3*, Article 25. https://doi.org/10.3389/fdata.2020.00025

Baldin, S. A. (2024). The impact of digital transformations in the economy and society on migration and public administration in the migration sphere. *Ekonomika i Upravlenie: Problemy, Resheniya, 5/9*(146), 6–13. https://doi.org/10.36871/ek.up.p.r.2024.05.09.001

Barnes, E., & Hutson, J. (2024). Navigating the ethical terrain of AI in higher education: Strategies for mitigating bias and promoting fairness. *Forum for Education Studies, 2*(2), Article 1229. https://doi.org/10.59400/fes.v2i2.1229

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development, 63*(4/5). https://doi.org/10.1147/JRD.2019.2942287

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., & Walker, K. (2020). *Fairlearn: A toolkit for assessing and improving fairness in AI* (MSR-TR-2020-32). Microsoft Research.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), Article 100205. https://doi.org/10.1016/j.patter.2021.100205

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems, 29*, 4349–4357

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology, 15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the Conference on Fairness, Accountability and Transparency* (pp. 77–91). https://proceedings.mlr.press/v81/buolamwini18a.html

Bura, C., Kamatala, S., & Myakala, P. K. (2025). Ethical challenges in data science: Navigating the complex landscape of responsibility and fairness. *International Journal of Current Science Research and Review, 8*(3). https://doi.org/10.47191/ijcsrr/V8-i3-09

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), Article 2053951715622512. https://doi.org/10.1177/2053951715622512

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Chadha, K. S. (2024). Bias and fairness in artificial intelligence: Methods and mitigation strategies. *International Journal for Research Publication and Seminar, 15*(3), 36–49. https://doi.org/10.36676/jrps.v15.i3.1425

Chinta, S. V., Wang, Z., Palikhe, A., Zhang, X., Kashif, A., Smith, M. A., Liu, J., & Zhang, W. (2024). *AI-driven healthcare: A review on ensuring fairness and mitigating bias* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2407.19655

Chouldechova, A., & Roth, A. (2018). *The frontiers of fairness in machine learning* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1810.08810

Curto, G., Jojoa Acosta, M. F., Comim, F., & Garcia-Zapirain, B. (2024). Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. *AI & Society, 39*(2), 617–632. https://doi.org/10.1007/s00146-022-01494-z

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (pp. 214–226). Association for Computing Machinery. https://doi.org/10.1145/2090236.2090255

Ejjami, R. (2024). The adaptive personalization theory of learning: Revolutionizing education with AI. *Journal of Next-Generation Research 5.0*. https://doi.org/10.70792/jngr5.0.v1i1.8

Ezeaka, N. B., & Umennebuaku, V. A. (2024). Development communication in the artificial intelligence (AI) era: Navigating cultural complexity and technological advancements. *African Journal of Culture, History, Religion and Traditions, 7*(3), 98–107. https://doi.org/10.52589/AJCHRT-HQCS2XJ7

Grincevičienė, V., Barevičiūtė, J., Asakavičiūtė, V., & Targamadzė, V. (2019). Equal opportunities and dignity as values in the perspective of I. Kant's deontological ethics: The case of inclusive education. *Filosofija, Sociologija, 30*(1), 80–88. https://doi.org/10.6001/fil-soc.v30i1.3919

Gurkan, N., & Suchow, J. W. (2024). *Exploring public opinion on responsible AI through the lens of cultural consensus theory*. arXiv. https://doi.org/10.48550/ARXIV.2402.00029

Hoffmann, A. L., Roberts, S. T., Wolf, C. T., & Wood, S. (2018). Beyond fairness, accountability, and transparency in the ethics of algorithms: Contributions and perspectives from LIS. *Proceedings of the Association for Information Science and Technology, 55*(1), 694–696. https://doi.org/10.1002/pra2.2018.14505501084

Huriye, A. Z. (2023). The ethics of artificial intelligence: Examining the ethical considerations surrounding the development and use of AI. *American Journal of Technology, 2*(1), 37–45. https://doi.org/10.58425/ajt.v2i1.142

Ibrahim, Z., Mohamed, J., Kassim, N., & Bahrudin, I. A. (2023). The assistive technology for teaching and learning of social skills for autism spectrum disorder children: Multimedia interactive social skills module application. *European Journal of Educational Research, 12*(3), 1465–1477. https://doi.org/10.12973/eu-jer.12.3.1465

Inuwa-Dutse, I. (2023). FATE in AI: Towards algorithmic inclusivity and accessibility. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3617694.3623233

Konidena, B. K., Malaiyappan, J. N. A., & Tadimarri, A. (2024). Ethical considerations in the development and deployment of AI systems. *European Journal of Technology, 8*(2), 41–53. https://doi.org/10.47672/ejt.1890

Mohanty, S. (2025). *Fine-grained bias detection in LLM: Enhancing detection mechanisms for nuanced biases* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2503.06054

Molla, M. A. M., & Ahsan, M. M. (2025). *Artificial intelligence and journalism: A systematic bibliometric and thematic analysis of global research* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2507.10891

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). *Bias in data-driven AI systems: An introductory survey* (arXiv:2001.09762). arXiv. https://doi.org/10.48550/arXiv.2001.09762

Osasona, F., Amoo, O. O., Atadoga, A., Abrahams, T. O., Farayola, O. A., & Ayinla, B. S. (2024). Reviewing the ethical implications of AI in decision-making processes. *International Journal of Management & Entrepreneurship Research, 6*(2), 322–335. https://doi.org/10.51594/ijmer.v6i2.773

Pasipamire, N., & Muroyiwa, A. (2024). Navigating algorithm bias in AI: Ensuring fairness and trust in Africa. *Frontiers in Research Metrics and Analytics, 9*, Article 1486600. https://doi.org/10.3389/frma.2024.1486600

Patel, D. B. (2021). Ethical AI: Addressing bias and fairness in machine learning models for decision-making. *Journal of Computer Science and Technology Studies, 3*(1), 13–17. https://doi.org/10.32996/jcsts.2021.3.1.3

Peng, Z., Fu, R. Z., Chen, H. P., Takahashi, K., Tanioka, Y., & Roy, D. (2024). AI applications in emotion recognition: A bibliometric analysis. *SHS Web of Conferences, 194*, Article 03005. https://doi.org/10.1051/shsconf/202419403005

Polo, E. P., & Ailodion, D. O. (2025). Tackling racial bias in AI systems: Applying the bioethical principle of justice and insights from Joy Buolamwini's "Coded Bias" and the "Algorithmic Justice League." *Bangladesh Journal of Bioethics, 16*(1), 8–14. https://doi.org/10.62865/bjbio.v16i1.129

Ramamoorthy, L. (2025). Evaluating generative AI: Challenges, methods, and future directions. *International Journal for Multidisciplinary Research, 7*(1), Article 37182. https://doi.org/10.36948/ijfmr.2025.v07i01.37182

Ratna, J., Kalnawat, A., Pawar, A. M., Jadhav, V. D., Srilatha, P., & Khetani, V. (2024). Transparency in algorithmic decision-making: Interpretable models for ethical accountability. *E3S Web of Conferences, 491*, Article 02041. https://doi.org/10.1051/e3sconf/202449102041

Robledo-Giraldo, S., Figueroa-Camargo, J. G., Zuluaga-Rojas, M. V., Vélez-Escobar, S. B., & Duque-Hurtado, P. L. (2023). Mapping, evolution, and application trends in co-citation analysis: A scientometric approach. *Revista de Investigación, Desarrollo e Innovación, 13*(1), 201–214. https://doi.org/10.19053/20278306.v13.n1.2023.16070

Saiz-Alvarez, J. (2024). Innovation management: A bibliometric analysis of 50 years of research using VOSviewer and Scopus. *World, 5*(4), Article 46. https://doi.org/10.3390/world5040046

Samuel-Okon, A. D. (2024). Smart media or biased media: The impacts and challenges of AI and big data on the media industry. *Asian Journal of Research in Computer Science, 17*(7), 128–144. https://doi.org/10.9734/ajrcos/2024/v17i7484

Shahzad, K., Khan, S., Iqbal, A., & Javeed, A. (2024). Identifying university librarians' readiness to adopt artificial intelligence (AI) for innovative learning experiences and smart library services: An empirical investigation. *Global Knowledge, Memory and Communication*. https://doi.org/10.1108/gkmc-12-2023-0496

Shuford, J. (2024). Exploring ethical dimensions in AI: Navigating bias and fairness in the field. *Journal of Artificial Intelligence General Science, 3*(1), 103–124. https://doi.org/10.60087/jaigs.vol03.issue01.p124

Shukla, S. (2024). Principles governing ethical development and deployment of AI. *International Journal of Engineering, Business and Management, 8*(2), 26–46. https://doi.org/10.22161/ijebm.8.2.5

Singh, C., Dash, M. K., Sahu, R., & Kumar, A. (2024). Artificial intelligence in customer retention: A bibliometric analysis and future research framework. *Kybernetes, 53*(11), 4863–4888. https://doi.org/10.1108/K-02-2023-0245

Tao, H., Wan, Y., Huang, L., & Zhao, Y. (2025). Applications and challenges of artificial intelligence in international trade and logistics: Current status, future developments, and policy recommendations. *International Journal of Computer Science and Information Technology, 5*(1), 35–47. https://doi.org/10.62051/ijcsit.v5n1.04

Thakur, N., & Sharma, A. (2024). Ethical considerations in AI-driven financial decision making. *Journal of Management & Public Policy, 15*(3), 41–57. https://doi.org/10.47914/jmpp.2024.v15i3.003

Venkatasubbu, S., & Krishnamoorthy, G. (2022). Ethical considerations in AI addressing bias and fairness in machine learning models. *Journal of Knowledge Learning and Science Technology, 1*(1), 130–138. https://doi.org/10.60087/jklst.vol1.n1.p138

Vindigni, G. (2025). Gender bias and cultural misrepresentation in AI: A critical inquiry into cross-cultural communication and algorithmic design. *European Journal of Applied Science, Engineering and Technology, 3*(3), 51–72. https://doi.org/10.59324/ejaset.2025.3(3).06

Wakunuma, K., & Eke, D. (2024). Africa, ChatGPT, and generative AI systems: Ethical benefits, concerns, and the need for governance. *Philosophies, 9*(3), Article 80. https://doi.org/10.3390/philosophies9030080

Weiner, E. B., Dankwa-Mullan, I., Nelson, W. A., & Hassanpour, S. (2025). Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. *PLOS Digital Health, 4*(4), Article e0000810. https://doi.org/10.1371/journal.pdig.0000810

Zhang, C. (2024). AI in education: Opportunities, challenges, and pathways for equitable learning. *Journal of Education, Humanities and Social Sciences, 45*, 723–728. https://doi.org/10.54097/kfgp6j07