International Journal of Innovative Research and Scientific Studies, 8(5) 2025, pages: 1595-1606

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Structural bias and geographical imbalances in AI ethics: A review

Pujun Sun[1], Rajermani Thinakaran[2*], Hushalictmy Paliyanny[3], J. Somasekar[4], Jayalakshmi Murugan[5]

[1]*Faculty of Business and Communications, INTI International University, Seremban, Malaysia.*
[2,3]*Faculty of Data Science and Information Technology, INTI International University, Malaysia.*
[4]*Department of Computer Science and Engineering, Jain (Deemed–to-be University), Bangalore, Karnataka, India.*
[5]*Kalasalingam Academy of Research and Education, Virudhunagar, Tamil Nadu, India.*

Corresponding author: Rajermani Thinakaran (*Email: rajermani.thina@newinti.edu.my*)

Abstract

This study conducts a systematic literature review (SLR) and bibliometric analysis on global research in artificial intelligence (AI) ethics. The study systematically synthesizes the latest developments and focuses on the thematic area in the field by drawing on the PICOC protocol and the PRISMA framework. Above all, it underscores that algorithmic bias arises from embedded social structures, institutional power relations, and data limitations. Additionally, it demonstrates that academic production is geographically imbalanced, with the North overwhelmingly producing the academic output, and the South contributing at a very low level. To focus on key areas for improvement, this study utilizes literature screening and data visualization techniques to identify research gaps and recommends a paradigm shift from equity adjustment at the algorithmic level to a broader sociotechnical system reconstruction.

Keywords: AI Ethics, Algorithmic Bias, Global South, Sociotechnical Systems, Spatial Justice.

DOI: 10.53894/ijirss.v8i5.9202
Funding: This study received no specific financial support.
History: Received: 27 June 2025 / **Revised:** 28 July 2025 / **Accepted:** 30 July 2025 / **Published:** 8 August 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Artificial intelligence (AI) has rapidly penetrated areas such as healthcare, finance, education, and employment, profoundly reshaping human life. However, its widespread adoption has also raised a growing number of ethical and social issues, particularly in the form of algorithmic bias. Although most existing research has focused on identifying and correcting data bias and unfairness in the algorithms themselves [1, 2], this research paradigm tends to reduce the problem of bias to a technical one, while ignoring its deep social structural roots [3]. Algorithms are not value-neutral tools, and their design, development, and application are inevitably influenced by sociocultural, institutional norms, and power relations [4, 5]. The burden of such prejudice is often borne by already marginalized populations [4].

The United Nations' Sustainable Development Goals (SDG) 10 explicitly call for reducing inequalities within and between countries. As an emerging technology with disruptive power, AI technology is expected to have great potential to promote financial inclusion, improve health services, and enhance the quality of education, thereby helping to achieve financial inclusion goals [6-9]. However, more and more studies have pointed out that the application of AI technology also carries the risk of exacerbating social inequality, especially in the countries of the Global South [10-12]. Therefore, a comprehensive assessment of AI technologies and how they are shaped under the SDG10 framework should be cautious about their possible negative impacts.

In addition, as globalization progresses more rapidly, the relationship between technological development and social inequality becomes increasingly complex. Most of the existing schemes, for instance, several studies (in the application of algorithmic governance), show that ethics checklists, algorithmic auditing, and transparency requirements originate from Western developed countries [13-15]. The effectiveness of these governance frameworks in the context of developing countries has been questioned [10, 11], while these governance frameworks have some applicability in the context of Western societies, in terms of technological development level, social and cultural background, institutional norms, and other aspects, developing countries and developed countries differ. It might be difficult to solve the unique ethical and social challenges of developing countries based on copying Western models [16, 17]. As a result, exploring a localized ethical governance framework for AI, especially as it applies to developing countries, has become an important focus of current research on AI ethics.

Moreover, the global study of AI ethics is characterized by a significant imbalance in its geographical distribution. Research on AI ethics is primarily led by research institutions and scholars in North America and Europe. They set questions and shape discourse systems, thereby constructing research paradigms [10, 11, 18, 19]. This regional imbalance makes the discourse on algorithmic bias governance inevitably influenced by geopolitics, while the voices and perspectives of the Global South are relatively marginalized, especially in Africa [20, 21]. In this context, how to rethink global inequality in AI ethics research from the perspective of spatial justice has become an important issue that needs to be further explored [22, 23]. At the same time, this study argues that to fundamentally solve the problem of algorithm bias and social inequality, it is necessary to shift from "algorithm correction" to "sociotechnical system reconstruction," that is, to re-imagine the interaction between technology and society at the institutional level and promote the fundamental reform of sociotechnical systems [24].

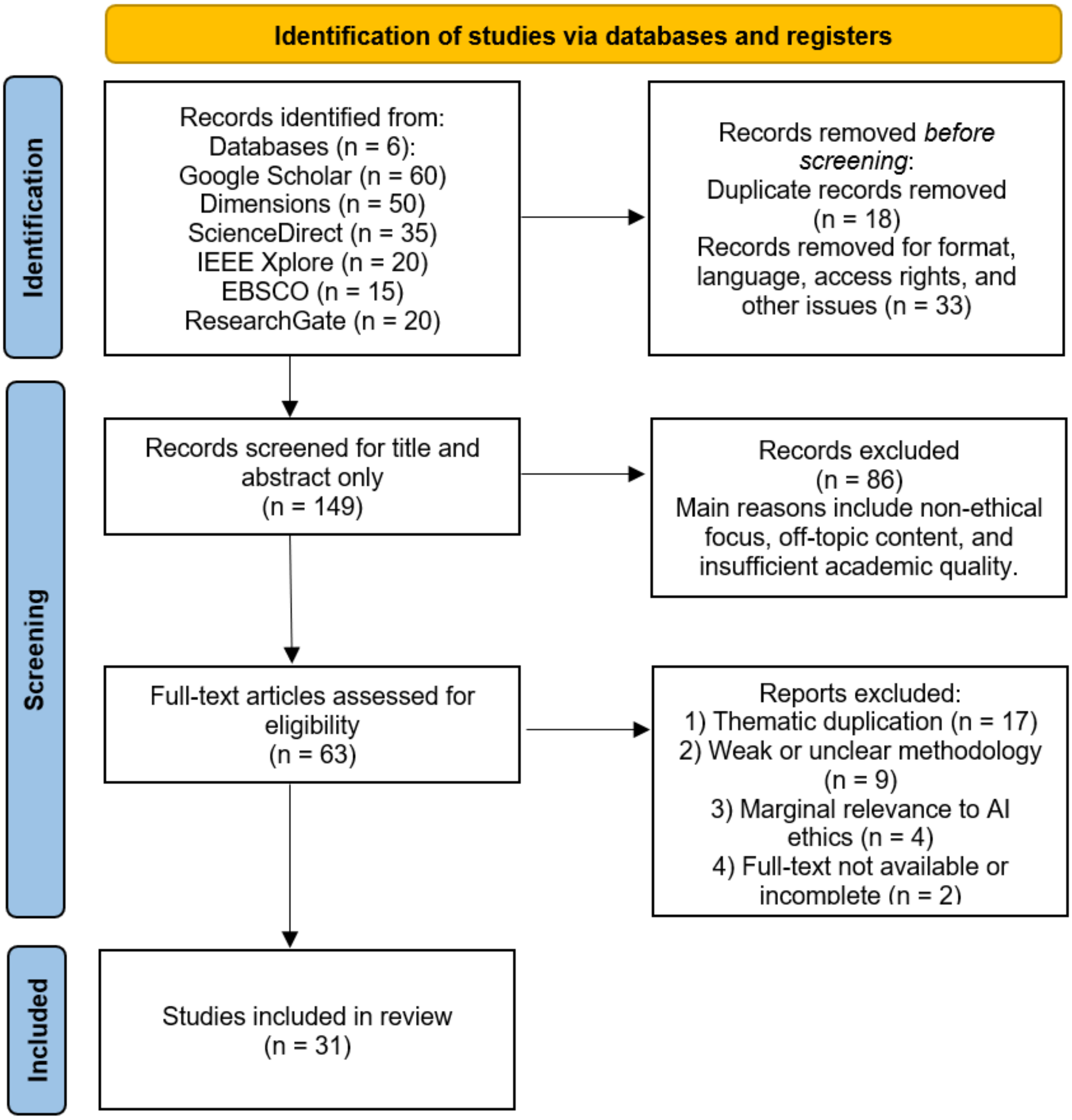


Figure 1.
PRISMA Flow Chart.

2. Method

In this study, the systematic literature review (SLR) method is combined with bibliometric visualization techniques to comprehensively discuss the current state and development trends of artificial intelligence ethics research. The process is as follows: 1. SLR: Used to systematically search and analyze literature related to AI ethics; 2. Bibliometric visualization techniques: Display research results through scientific data cleaning and visualization technology.

2.1. SLR

2.1.1. Research Purpose

SLR is a research method that collects, identifies, and critically analyzes existing research through a systematic process [25, 26]. Its main purpose is to identify and analyze key approaches, research trends, and research challenges in the field of AI ethics research. Through systematic search and screening, a comprehensive research basis was formed.

Table 1.
PICOC Framework.

Description	Example (PICOC)	Example (Synonyms)
Population	Researchers in the AI Ethics field	AI Morality, Algorithmic Ethics, Machine Ethics
Intervention	SLR Method	Bibliometric Analysis, Literature Synthesis
Comparison	Traditional Literature Review Methods	Qualitative Review, Expert Commentary, Thematic Analysis
Outcome	Research hotspots and trends in AI Ethics	AI Ethics Framework, Ethical Challenges, Research Frontiers
Context	Global AI Ethics Research	Sociotechnical Systems, Ethical Governance, Algorithm Fairness

2.2. Research Protocol

According to the standard process of SLR, the PICOC model (Population, Intervention, Comparison, Outcomes, Context) was adopted as the research protocol in this study to guide the literature search strategy [25, 27]. The details are shown in Table 1.

3. Literature Retrieval

Systematic research was conducted in the following databases: Google Scholar, Dimensions, EBSCO, ResearchGate, ScienceDirect, and IEEE Xplore. The years of publication range from 2019 to 2025, and the search keyword used was "AI Ethics".

Table 2.
Literature Screening and Quality Assessment.

Inclusion criteria	Exclusion criteria
Academic resources are published in prestigious journals, peer-reviewed conferences, or high-impact preprint platforms with recognized academic value.	Literature with only abstracts
The research topic focuses on the ethics of AI.	literature whose content is not relevant to AI ethics research.
The article is in full text and published in English.	Studies with vague, unclear, or missing methodology sections.
In general, the paper sets out clear research methods and future directions.	Papers with weak theoretical contribution or limited academic quality.

3.1. Literature Screening and Quality Assessment

As a result, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram shown in Figure 1 was followed for the screening process to ensure transparency and methodological rigor. The diagram was generated using the Shiny App, which is available online at: https://estech.shinyapps.io/prisma_flowdiagram/ [28].

3.2. Data Extraction and Synthesis

Table 3 in Appendix I makes an in-depth analysis of the selected literature and summarizes the research methods, research topics, and future development directions. By combining the research methods, the applicability and limitations are analyzed.

3.3. Bibliometric Visualization Techniques

3.3.1. Data Collection and Cleaning

To provide a comprehensive bibliometric overview complementing the literature review, 2,500 published records were exported from the Dimensions database (2022–2025) using the keyword “AI Ethics.” Data preprocessing followed these steps:

1. Search for the keyword “AI Ethics” from Dimensions and ensure that the search results are sorted by relevance, as shown in Figure 2.
2. Export results for bibliometric mapping.
3. Data screening and cleaning results:
   - Total data size: 2500 publicised literatures
   - Publication year: 2022-2025
   - Selected features: Authors' Affiliations - Country of Research organization
   - Check for missing values: 102 articles show null, and 2 are publications with more than 100 authors, exported without author names and affiliations. These 104 articles will be excluded. Thus, the cleaned data set contains 2,396 articles.
   - Data collation: Of these, 735 had multiple authors from different countries. To avoid double-counting and reduce data complexity, this study will be based on the country of the first author, as they are usually the main contributors to the study.

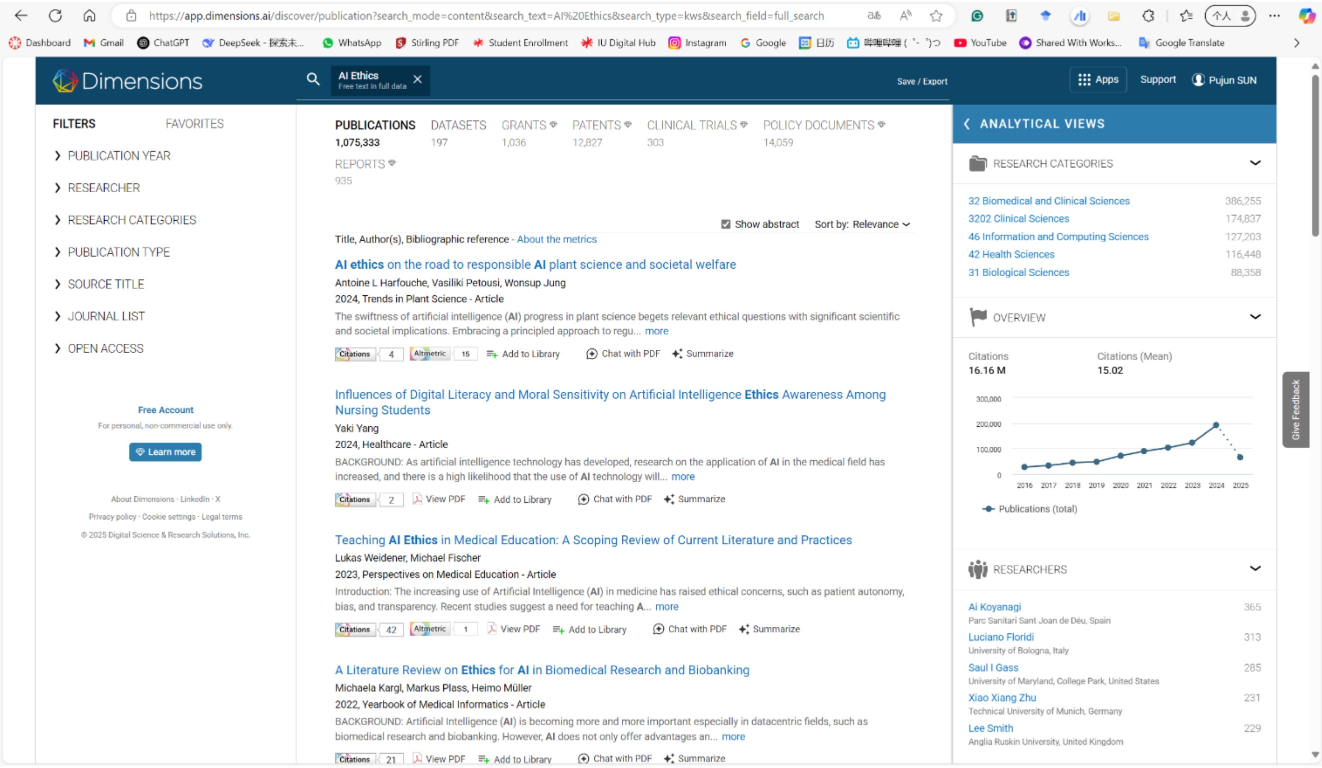


Figure 2.
Search Result by Dimensions.

3.4. Data Import and Visualization Generation

After which the cleaned data was imported to Tableau Software [29] for visualization using 1:110m Cultural Vectors by Countries [30]. The main visualizations include:

1. World Map Distribution: Illustrates the top countries that published the AI Ethics research literature.
2. Map of Research Institutions: Shows the level of research activity in different countries.

Spatial Data used in Tableau: 1:110m Cultural Vectors by Countries, can be accessed from: Natural Earth » 1:110m Cultural Vectors - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales [30].

3.5. Visual Presentation and Analysis

The global distribution and trends of AI ethics research are analyzed using visualizations in Tableau, which are then discussed with research-intensive areas and potential research gaps.

4. Results and Discussion

4.1. Structural Mechanisms of Algorithmic Bias

4.1.1. Social structure and technological mediation

Existing research has shown that algorithm-based bias is not only due to statistical biases in data but also because technological systems themselves are products and re-producers of social structures [24, 31].

Machine Habitus Airoldi [24] starts by pointing out that algorithmic systems (as much as humans) build upon sociocultural and power relations while developing and training, forming habits that are tendencies. This is socializing technology, on the theoretical level, being the reproduction and standardization of a specific social background. Second, Situated Algorithms Draude et al. [4] help us understand that algorithms are not independent of the sociocultural context in which they exist, but are situated in particular technological contexts, and are highly influenced by the contexts and power structures of these contexts. According to this perspective, any fairness and transparency analysis of algorithms cannot be detached from the social context in which they operate. In addition, one research emphasized that it is necessary to go beyond single technology optimization and conduct comprehensive governance from the perspective of a sociotechnical system. Their proposed generative AI system security assessment framework, including capability assessment, human-computer interaction assessment, and system impact assessment, emphasizes the complexity of social and technological mutual nesting and calls for the integration of social equity considerations in the algorithm design stage [13].

4.1.2. Case Study: Algorithmic Bias in Multiple Domains

4.1.2.1. Health Care and Public Services: Assessment of Perceived Injustice and Discrimination

In health care, gene and genome sequencing technologies (GSTs), despite their potential in disease diagnosis, often carry social biases, especially epistemic injustice in the cognitive evaluation of patients through "epistemic capture" and "value partitioning," to transform ambiguous information into definitive knowledge and reduce complex values to binary opposites that compromise the patient's ability to function as a cognitive subject. Similarly, in the field of child protection, algorithm-based models make risk assessments based on historical data from social service agencies, which themselves contain institutional bias [1]. For example, labeling people as 'at-risk' families simply helps the algorithm to further reinforce discrimination against certain groups with its distorted loop of feedback. Consequently, in health and social welfare, algorithmic bias not only serves to diagnose individual instances of bias but also embeds social discriminatory bias as technological bias in data-driven mechanisms.

4.1.2.2. Social Media And Content Moderation: Algorithmic Sexist Mechanisms

In social media content moderation, platforms automatically filter and recommend content through algorithms, but these algorithms often reinforce gender bias, especially concerning female image and body representation, with significant imbalances [31]. Researchers pointed out that the algorithm's sensitivity to keywords and its preference for "normative gender roles" lead to the exclusion of minority forms of gender expression [31].

4.1.2.3. Big Data Finance and Social Stratification: Algorithm-Driven Diffusion of Injustice

In the big data financial model, the opacity of the financial algorithm model and the automation of decision-making have long disadvantaged socially vulnerable groups in algorithm evaluation. For example, in loan risk assessment, algorithms often label low-income individuals as high-risk based on factors such as zip code and income level [2]. This "Weapons of Math Destruction" not only deprives poor groups of access to credit but also solidifies social stratification.

4.1.2.4. Contextualization Bias of Intelligent Recommendation Systems

Social media recommendation algorithms push content based on users' historical behaviors, which makes it easy to form stereotypical recommendations on gender and ethnic issues [4]. This algorithmic mechanism creates a cultural reproduction of bias. In other words, the embeddedness of algorithms in intelligent recommendations further solidifies the bias through technological intermediation.

4.2. Inequality And Marginalization in the Southern Hemisphere

4.2.1. Double Effects of AI Technology and the Global South Perspective

In promoting the SDG10 goals, AI technology has a dual effect. While it can improve efficiency in education, health, and finance, it also increases inequality due to the digital divide and global cognitive asymmetry [7, 8].

First, AI technology has the potential to promote equality. In areas such as healthcare, education, and financial services, AI technology can optimize resource allocation and service accuracy through intelligent analytics and data mining [6]. However, AI technology does not have a positive effect in all situations. Especially in the countries of the Global South, due to weak technical infrastructure and insufficient data resources, the persistence of inequality and algorithmic discrimination often occurs in algorithmic decision-making [10]. One research emphasizes that AI technologies do not exist in isolation but are part of a sociotechnical system whose development and application are influenced by larger structural factors, as well as economic and political systems [7]. This structural inequality places the countries of the South at a long-term disadvantage in terms of the benefits of AI technology. Among them, insufficient localization and technology monopoly are the main problems faced by the countries of the global South. Countries in the global South often lack representation and initiative in technological innovation and application, and the existing algorithmic governance frameworks mostly originate from North America and Europe, ignoring the unique social, cultural background and the realistic needs of developing countries [32]. This Western-centered governance model exacerbates the dilemma faced by southern countries in technological governance.

4.3. Case Study: Challenges of AI Technology Application Under the SDG10 Framework

4.3.1. Education and health: Digital divide and equity dilemma

One study has uncovered resource inequality in digital learning during the pandemic when studying the digital learning experiences of university students in South Africa. Urban students are relatively well-off, while rural students fall prey to educational inequality due to inadequate equipment and poor Internet connections [12]. Similarly, Han and Kumwenda [11] pointed out that there is a significant digital divide in online medical education between the North and the South of the world, and the long-term shortage of digital skills and technical resources in southern countries leads to limited educational equity. In addition, Akpan et al. [10] pointed out that the global South lags far behind the global North in research on virtual education technology and e-learning systems, especially in sub-Saharan Africa and Latin America, and that technology investment and research and development capacity are seriously insufficient, resulting in the difficult realization of educational equity. These cases show that while online education promotes knowledge sharing, the digital divide is exacerbating educational inequality, especially in countries of the global South. This technological divide is also marginalized in ethical discussions, making it difficult for the countries of the South to have a voice in global educational technology research.

4.3.2. Platform Economy and Financial Inclusion: Value Autonomy and Risk Intensification

One researcher proposed the concept of algorithmic inequality and pointed out that the asymmetry between data privacy and algorithmic decision-making may exacerbate the consumption induction and data exploitation of vulnerable groups [33]. In the global South, the risks of data misuse and algorithmic manipulation are more pronounced due to weak consumer protection mechanisms. Another researcher points out that the development of digital platforms in southern countries is often simplified as a process of financial inclusion but conceals the risk of financialization [32]. Mobile money services in Africa, while improving financial access, are creating new forms of economic exploitation due to platform monopolies and capital controls. This paradox of digital inclusion highlights that the indigenous economic structure and socio-cultural characteristics of southern countries have long been ignored in the study of the platform economy. The theoretical monogeneity of platform economy research weakens the cognition of the particularity of southern countries and makes the global AI ethics research lose diversified support.

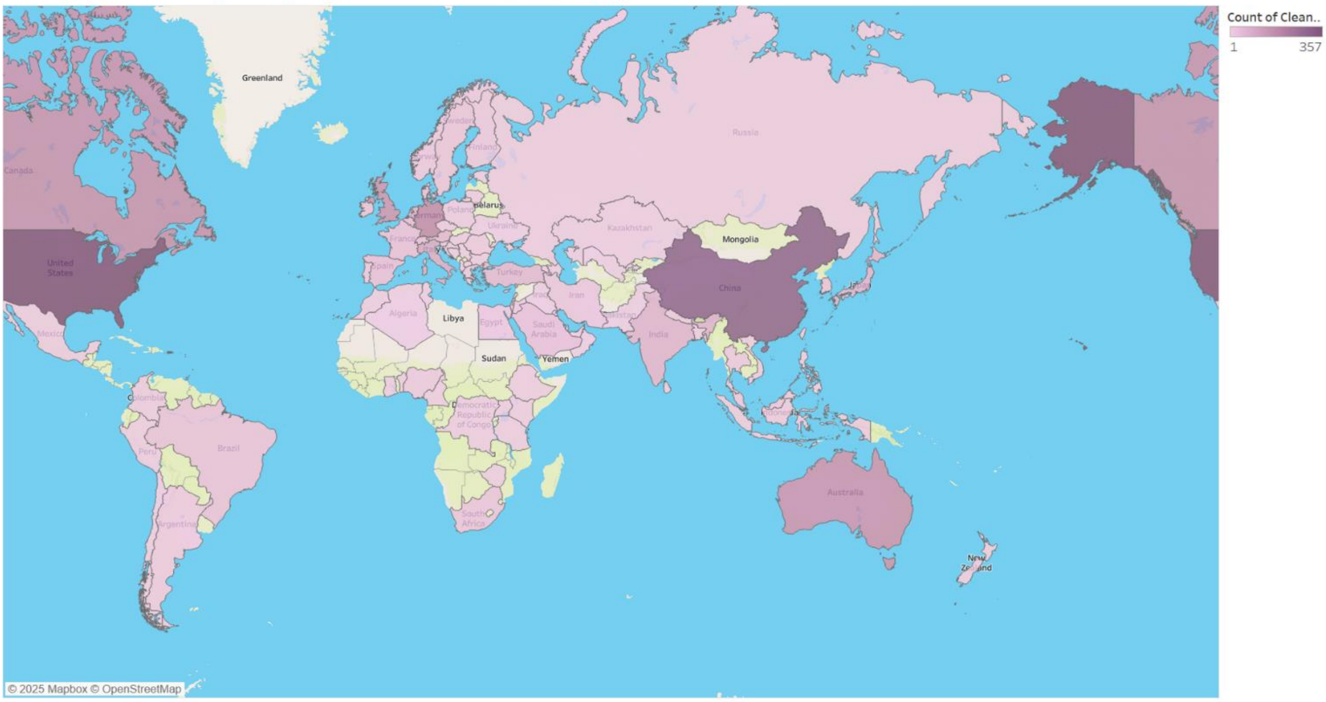
4.3.3. Technological Sovereignty and Data Security: Rebalancing Control of Digital Platforms

Some countries in Latin America, when introducing smart government platforms, found that the main technology infrastructure was controlled by multinational platform companies, with risks of loss of data sovereignty and digital colonization [34]. For example, in Chile's e-government reform, the platform data server is located in Europe and the United States, and the public service data faces threats of data leakage and abuse. This external dependence on technological control presents challenges for developing countries in maintaining technological sovereignty. Therefore, algorithmic governance frameworks in developing countries should emphasize local technology development and data management to reduce reliance on transnational digital platforms.

Figure 3 was created using Tableau Software [29] based on bibliometric data exported from Dimensions on 2,396 AI ethics-related publications from 2022 to 2025. The visualization illustrates both the distribution of country-level affiliations and the first authors' institutional affiliations. It highlights geographic disparities in academic output. The number of publications associated with each country is represented by color intensity, ranging from 1 to 357 articles, as indicated by the scale bar on the left.

This figure (Figure 3) illustrates the global North-South imbalance in academic output in AI ethics research, with North American and European countries dominating. In contrast, many countries in Africa and South America have published less work in AI ethics research, and some countries have not produced relevant academic output. Additionally, there are only a few studies in developing regions such as the Middle East and South Asia. Notably, China represents a major contributor from the Global South, which deviates from general patterns of a lack of representation across developing regions. Consequently, this disparity directly reflects the academic weakness at the level of discourse power of such studies on the research of digital technology and educational ethics in Southern countries.

<The country distribution of authors of articles published between 2022 and 2025, including "AI Ethics">



Map based on Longitude (generated) and Latitude (generated). Color shows count of Cleaned- Authors Affiliations -. Details are shown for Authors Affiliations - Country of Research organization.

Figure 3..
Global Distribution of AI Ethics Research Publications (2022-2025).

On the other hand, some researchers highlight in their research that the ethical wisdom of the South, such as Ubuntu and Buen Vivir, tends to have collectivist and eco-ethical qualities in the sustainable lifestyle and consumption debate. However, in global AI ethics research, these cultural traditions are underrepresented, and the individualistic ethic of northern countries dominates [35]. This cultural discourse inequality weakens the contribution of southern countries in the construction of algorithm ethics.

Figure 4 was also generated by Tableau Software [29] has directly compares the number of published research papers on AI ethics in various countries. It can be seen from the figure that northern countries such as the United States, China, the United Kingdom, and Germany occupy a dominant position in AI ethics research, while southern countries such as Brazil, South Africa, and India account for a small proportion. It further validates the marginalization of Southern ethical wisdom in global discussions.

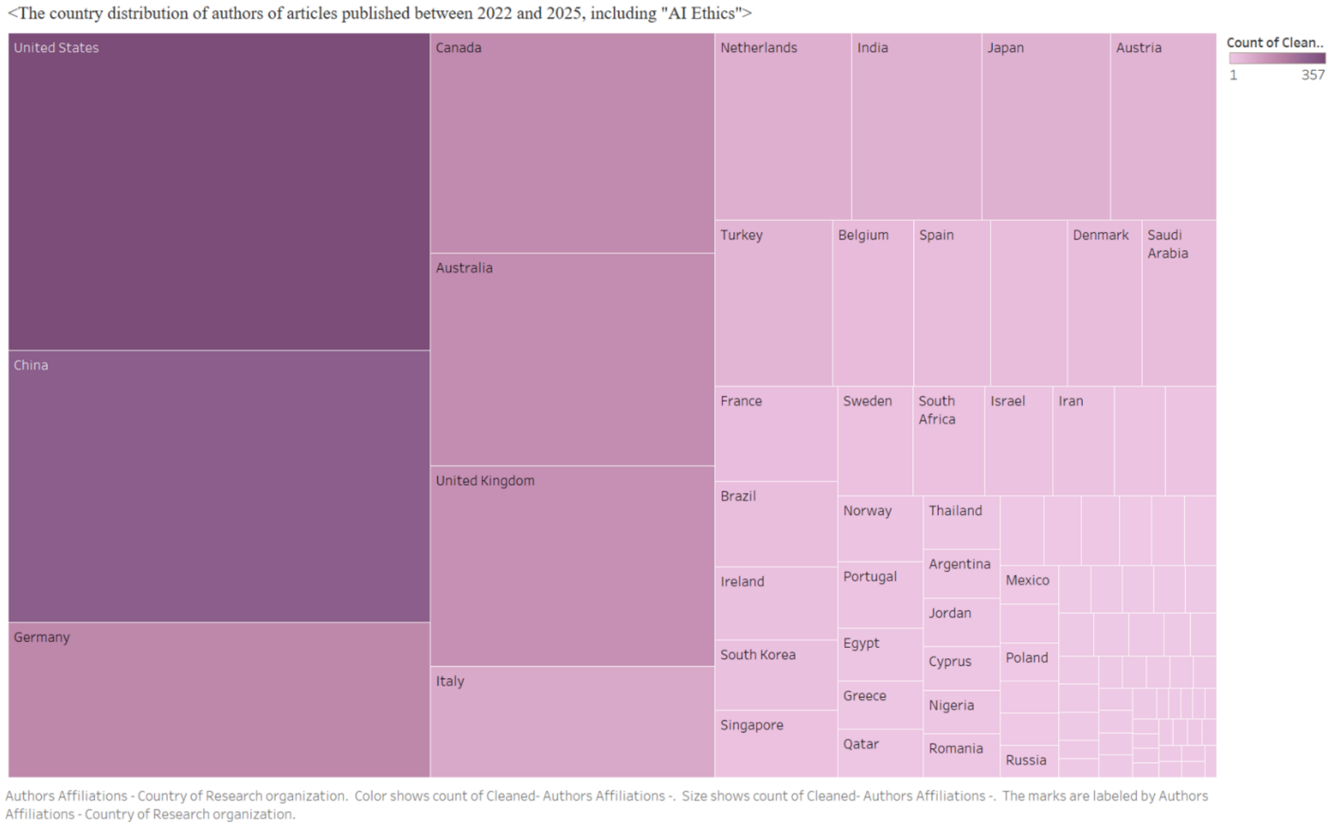


Figure 4.
Global AI ethics Research Publication Country Comparison chart (2022-2025).

4.4. Indigenous Improvement of Algorithmic Governance Frameworks

4.4.4. Localization Governance and Technology Sovereignty

First, the ethical framework has a significant need for localization. Among them, the three-layer framework proposed emphasizes the comprehensibility of sociotechnical systems, but the framework mainly focuses on safety at the technical level and fails to consider ethical risks under different socio-cultural backgrounds [13]. Therefore, relying solely on the Western ethical model is difficult to adapt to the complex social situations in developing countries.

Moreover, digital technologies still present a crisis of trust and governance challenges. The theory of "Mediated trust" reveals the crisis of trust caused by digitalization and globalization [36]. In algorithmic governance, the construction of trust mechanisms often relies on technical intermediaries, and the reliability and transparency of these intermediary services are questioned in developing countries. The transfer of trust brought about by digital technologies requires enhanced community participation and pluralistic oversight within a localized governance framework to foster greater trust among local populations in technology.

Finally, the expansion of the power of digital platforms runs parallel to the loss of technological sovereignty. There is a researcher who proposes the concept of "Cloud Empires," in which digital platforms transcend states to become new centers of power and have a profound impact on society by controlling data and algorithms. In developing countries, the technology monopoly of digital platforms has exacerbated the loss of technological sovereignty and social inequality, making it urgent to introduce the principle of technological autonomy into the localization governance framework to safeguard digital sovereignty and social equity.

5. Future Work

Through a comprehensive review of existing research on AI ethics, this study reveals the multiple challenges of algorithmic bias in the context of the global North-South technology divide, especially the importance of localization, improvement of algorithmic governance frameworks, and the restructuring of sociotechnical systems. Future research should continue to deepen and expand in the following aspects.

First, theory deepening and model building remain important tasks in AI ethics research. Most existing studies focus on algorithm modification and lack in-depth analysis of the structural root causes of algorithm bias [24, 31]. In the future, future studies should analyze the generation mechanism of algorithmic bias to uncover the role institutional factors, socio-cultural bias, and power relations play in technological development [37]. Accordingly, based on a theoretical model, we can combine sociological theory with technical analysis so that a comprehensive analysis framework for the reconstruction of a sociotechnical system is suggested. The integration of the principle of equity with the analysis of social structure is organically integrated into the analysis of technological governance, facilitating its two-way development of theory and practice [4, 13]. Besides, theoretical innovation also includes paying attention to the introduction of the concepts of “machine habit” and “contextualized algorithm”; understanding the socialization process of algorithmic bias; and exploring new routes of algorithmic governance.

Secondly, the dominant ethnic background of AI is the result of Western moral hegemony over AI established through pseudo-moral narratives [38]. This moral hegemony marginalizes the voices and needs of the global South in ethical discourse. Han and Kumwenda [11] found that countries in the global South have lacked representation in online medical education and virtual learning technology research for a long time, resulting in inadequate adaptability of the technology governance framework in southern countries. In AI ethics research, this lack of discourse further exacerbates ethical blind spots in technology governance, especially on issues such as algorithmic bias, the digital divide, and cultural diversity. In response to this phenomenon, ethics should return to philosophy, break away from the simple function of communication, and cannot be monopolized by communication experts [38]. This point of view reminds us that when discussing AI ethics, we should avoid being limited by the Western moral discourse system, in order to respond more comprehensively to the ethical demands in different social and cultural contexts.

An example is a study that aims to bridge the global North-South gap in sustainable lifestyles and ethical frameworks; it is necessary to pay attention to the cultural values of countries in the global South [35]. Similarly, in AI ethics research, the spatial justice perspective calls for a break with single-centered research structures and a multi-centered, inclusive, and culturally sensitive research paradigm that enables countries in the global South to have an equal say in agenda-setting and research methodology [35]. The rich ethical traditions and cultural values of the countries of the South should be an important component of global ethical research, not marginalized. Specifically, the South's voice channels in algorithmic governance and technology ethics research should be enhanced through international collaborative research projects and regional academic platforms. At the same time, localization algorithm governance tools and technical support systems should be developed to reduce dependence on northern technical standards [34, 36].

Finally, the transformation from algorithm correction to sociotechnical system reconstruction is a key link of equitable technological governance [37]. This, therefore, demands social embedding in application scenarios as well as social equity and ethical values in algorithmic design [4]. Interactive ethical impact assessment tools can be developed in practice and are used to enable policymakers to monitor AI ethical risks and increase the social acceptability of algorithm applications [13]. In addition, we can also combine the concept of contextualized algorithms to explore the reproduction mechanism of algorithmic bias within a specific social background. By adopting the sociotechnical system design method, ethical reflection is incorporated into the entire process of technological development to ensure the comprehensiveness and inclusiveness of algorithm governance.

6. Conclusion

This study critically analyzes the structural biases in AI ethics research and the challenges faced by the global North-South divide in the development and utilization of AI technology. The research conducted and the discussion held in this study are based on SLR and bibliometric visualization techniques. It identified that current AI ethics research is primarily based on a Western-centric paradigm, which does not consider the social and cultural specificities of Global South countries. The results show that algorithmic bias is a social, cultural, and political issue and should be approached in a context-sensitive manner.

This study contributes one of the key insights that currently utilized AI governance frameworks, stemming from North American and European contexts, are not easily transferable to the diverse socio-cultural settings of the global South. This can lead to poor governance models, which fall short of addressing the problems encountered by developing countries, such as the evolution of the digital divide, local cultural values, and socio-economic weaknesses.

Further, the study suggests that spatial justice is key to research for upholding AI ethics and asserts that the prevalence of voices and viewpoints from the global north is disempowering for those of the global south. In addition to this, the distribution of research output is unequal, and the concentration of academic power in a small number of countries adds to such uneven distribution. Therefore, this study calls for an inclusive, multiple-perspective approach to AI ethics research that accounts for cultural diversity.

Thus, making algorithmic corrections alone should not suffice to move beyond the problem; therefore, we propose a paradigm shift from “algorithmic adjustment” to “social-technical system reconstruction.” To achieve this transformation, ethical considerations must be woven into the core of the way AI systems are designed and deployed, and the relationship between local contexts and cultural nuances must be carefully considered.

Finally, both cementing the dominant Western-centric governance model of AI ethics in the global South and relying on the hegemony of these models as the overarching solutions for global AI ethics and justice are seen as causes for concern.

Through support for intercontinental collaboration between North and South scholars and policymakers and encouragement of place-adaptive frameworks of AI ethics, we ensure that AI ethics becomes an inclusive and just field.

References

[1] E. Keddell, "Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice," *Social Sciences*, vol. 8, no. 10, p. 281, 2019. https://doi.org/10.3390/socsci8100281

[2] S. Verma, "Weapons of math destruction: How big data increases inequality and threatens democracy," *Vikalpa*, vol. 44, no. 2, pp. 97-98, 2019. https://doi.org/10.1177/0256090919853933

[3] D. Allhutter and B. Berendt, "Deconstructing FAT," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.

[4] C. Draude, G. Klumbyte, P. Lücking, and P. Treusch, "Situated algorithms: A sociotechnical systemic approach to bias," *Online Information Review*, vol. 44, no. 2, pp. 325-342, 2020. https://doi.org/10.1108/OIR-10-2018-0332

[5] J. Sherman, R. Morrison, L. Klein, and D. Rosner, "The power of absence: Thinking with archival theory in algorithmic design," presented at the Proceedings of the 2024 ACM Designing Interactive Systems Conference, Copenhagen, Denmark, 2024.

[6] C. G. Bocean, "Sustainable development in the digital age: Harnessing emerging digital technologies to catalyze global SDG achievement," *Applied Sciences*, vol. 15, no. 2, p. 816, 2025. https://doi.org/10.3390/app15020816

[7] H. S. Sætra, "AI in context and the sustainable development goals: Factoring in the unsustainability of the sociotechnical system," *Sustainability*, vol. 13, no. 4, p. 1738, 2021. https://doi.org/10.3390/su13041738

[8] M. Shahvaroughi Farahani and G. Ghasemi, "Artificial intelligence and inequality: Challenges and opportunities," *Qeios*, 2024. https://doi.org/10.32388/7HWUZ2

[9] P. V. Kumar, S. Priya, D. Gunapriya, and M. Batumalay, "lower limb prosthetics using optimized deep learning model – a pathway towards sdg good health and well being," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1147-1161, 2024.

[10] I. J. Akpan, O. F. Offodile, A. C. Akpanobong, and Y. M. Kobara, "A comparative analysis of virtual education technology, E-learning systems research advances, and digital divide in the global South," *Informatics*, vol. 11, no. 3, p. 53, 2024. https://doi.org/10.3390/informatics11030053

[11] S. P. Han and B. Kumwenda, "Bridging the digital divide: Promoting equal access to online learning for health professions in an unequal world," *Medical Education*, vol. 59, no. 1, pp. 56-64, 2025. https://doi.org/10.1111/medu.15455

[12] D. A. Reddy Moonasamy and P. G. Mulliah Naidoo, "Digital learning: Challenges experienced by South African university students' during the COVID-19 pandemic," *The Independent Journal of Teaching and Learning*, vol. 17, no. 2, pp. 76-90, 2024. https://doi.org/10.17159/fwk81482

[13] L. Weidinger *et al.*, "Sociotechnical safety evaluation of generative ai systems," *arXiv preprint arXiv:2310.11986*, 2023. https://doi.org/10.48550/arXiv.2310.11986

[14] S. Shukla, "Principles governing ethical development and deployment of AI," *International Journal of Engineering, Business And Management*, vol. 8, no. 2, pp. 26-46, 2024. https://doi.org/10.22161/ijebm.8.2.5

[15] M. Hickok, "Lessons learned from AI ethics principles for future actions," *AI and Ethics*, vol. 1, no. 1, pp. 41-47, 2021. https://doi.org/10.1007/s43681-020-00008-1

[16] S. Ness, M. Sarker, M. Volkivskyi, and N. Singh, "The legal and political implications of AI bias: An international comparative study," *American Journal of Computing and Engineering*, vol. 7, no. 1, pp. 37–45, 2024. https://doi.org/10.47672/ajce.1879

[17] I. Oladele, A. Orelaja, and O. T. Akinwande, "Ethical implications and governance of artificial intelligence in business decisions: A deep dive into the ethical challenges and governance issues surrounding the use of artificial intelligence in making critical business decisions," *International Journal of Latest Technology in Engineering Management & Applied Science*, vol. XIII, no. 1, pp. 48–56, 2024.

[18] C. Roche, D. Lewis, and P. Wall, "Artificial intelligence ethics: An inclusive global discourse?," *arXiv preprint arXiv:2108.09959*, 2021.

[19] A. M. Astobiza, T. Ausín, B. Liedo, M. Toboso, M. Aparicio, and D. López, "Ethical governance of AI in the global south: A human rights approach to responsible use of AI," in *Proceedings of The 2021 Summit of the International Society for the Study of Information (p. 136)*, 2022.

[20] S. T. Segun, "Critically engaging the ethics of AI for a global audience," *Ethics and Information Technology*, vol. 23, no. 2, pp. 99-105, 2021. https://doi.org/10.1007/s10676-020-09570-y

[21] E. Ormond, "Governance of ai ethics: Perspective from the global South (Africa)," *Available at SSRN 4368020*, 2023. https://doi.org/10.2139/ssrn.4368020

[22] M. Walker and J. Winders, "Where is artificial intelligence? Geographies, ethics, and practices of AI," *Space and Polity*, vol. 25, no. 2, pp. 163-166, 2021. https://doi.org/10.1080/13562576.2021.1985869

[23] S. T.-I. Wu, D. Demetriou, and R. A. Husain, "Honor ethics: The challenge of globalizing value alignment in AI," presented at the Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 2023.

[24] M. Airoldi, *Machine habitus*. Hoboken, NJ: John Wiley & Sons, 2021.

[25] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," *MethodsX*, vol. 9, p. 101895, 2022. https://doi.org/10.1016/j.mex.2022.101895

[26] L. W. Yen, R. Thinakaran, and J. Somasekar, "Machine learning-based denoising techniques for Monte Carlo rendering: A literature review," *Machine Learning*, vol. 16, no. 2, pp. 1-9, 2025.

[27] R. Thinakaran and S. Chuprat, "Students' characteristics of student model in intelligent programming tutor for learning programming: A systematic literature review," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 669-676, 2022.

[28] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, "PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis," *Campbell Systematic Reviews*, vol. 18, no. 2, p. e1230, 2022. https://doi.org/10.1002/cl2.1230

[29] Tableau Software, *Tableau desktop, version 2025.1.0*. Seattle, WA, USA: Tableau Software, 2025.

[30] Natural Earth, "“Natural Earth» 1:110m Cultural Vectors - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales. aturalearthdata.com," 2021. https://www.naturalearthdata.com/downloads/110m-cultural-vectors/. [Accessed Apr. 2025]
[31] Y. Gerrard and H. Thornham, "Content moderation: Social media’s sexist assemblages," *New Media & Society*, vol. 22, no. 7, pp. 1266-1286, 2020. https://doi.org/10.1177/1461444820912540
[32] J. Roitman, "Platform economies: Beyond the North-South divide," *Finance and Society*, vol. 9, no. 1, pp. 1-13, 2023. https://doi.org/10.2218/finsoc.8089
[33] Z. Liu, M. Sockin, and W. Xiong, "Data privacy and algorithmic inequality. NBER Working Paper No. w31250," 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4448301
[34] V. Lehdonvirta, *Cloud empires*. Cambridge, MA: The MIT Press, 2022.
[35] B. Hayward and J. Roy, "Sustainable living: Bridging the North-South divide in lifestyles and consumption debates," *Annual Review of Environment and Resources*, vol. 44, no. 44, 2019, pp. 157-175, 2019. https://doi.org/10.1146/annurev-environ-101718-033119
[36] B. Bodó, "Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators," *New Media & Society*, vol. 23, no. 9, pp. 2668-2690, 2021. https://doi.org/10.1177/1461444820939922
[37] B. Mittelstadt, S. Wachter, and C. Russell, "The unfairness of fair machine learning: Leveling down and strict egalitarianism by default," *arXiv:2302.02404*, 2023. https://arxiv.org/abs/2302.02404
[38] E. R. Goffi, "Escaping the Western cosm-ethical hegemony: The importance of cultural diversity in the ethical assessment of artificial intelligence," *The AI Ethics Journal*, vol. 2, no. 2, p. 1, 2021. https://doi.org/10.47289/AIEJ20210716-1

Appendix I

Table 3.
Literature Review Summary.

Author	Research Methods	Research Topics	Future Directions
Keddell [1]	Critical perspective	Algorithmic justice in child protection	Balancing statistical fairness with social justice in algorithmic decision-making
Verma [2]	Case study and critical analysis	Big data and algorithmic bias	Developing regulatory frameworks to audit and govern algorithms
Allhutter and Berendt [3]	Mind scripting: Discourse analysis	Hidden assumptions and power relations in FAT research	Reflecting on implicit power dynamics in technical norms
Draude, et al. [4]	Conceptual paper	Sociotechnical systemic approach to algorithmic bias	Integrating gender and diversity studies into algorithmic fairness
Sherman, et al. [5]	Critical archival theory	Application of archival theory in algorithmic bias design	Rethinking the nature of bias and exploring inclusive algorithm design
Bocean [6]	Empirical analysis (Structural Equation Modeling)	Impact of digital technologies on Sustainable Development Goals (SDGs)	Adaptive policies for digital inclusion and sustainable growth
Sætra [7]	Conceptual framework analysis	AI impact on Sustainable Development Goals (SDGs)	Addressing the sociotechnical system’s unsustainability within AI development
Shahvaroughi Farahani and Ghasemi [8]	Qualitative analysis	Impact of AI on inequality and countermeasures	Developing socially fair and inclusive AI systems
Akpan, et al. [10]	Quantitative bibliometric performance analysis	Virtual education technology and e-learning in the Global South	Addressing the digital divide and promoting equitable R&D funding
Han and Kumwenda [11]	Literature review	Bridging the digital divide: Promoting equal access to online learning in an unequal world	Developing strategies to ensure equal access to online education resources, with a focus on technological support and infrastructure in resource-poor areas
Reddy Moonasamy and Mulliah Naidoo [12]	Quantitative research (Survey)	Challenges faced by South African university students in transitioning to online learning during COVID-19	Developing strategic plans for digital literacy and resource availability in higher education institutions
Weidinger, et al. [13]	Sociotechnical safety evaluation	Safety assessment of generative AI systems	Improving safety evaluation frameworks for multimodal AI systems
Shukla [14]	Systematic literature review	Ethical development and deployment of AI	Building global ethical governance frameworks
Hickok [15]	Opinion paper	Lessons learned from AI ethics principles for future actions	Promoting diversity and inclusivity in AI ethics frameworks; moving from principles to practice
Ness, et al. [16]	Comparative analysis	Legal and political impacts of AI bias	Establishing robust regulatory frameworks for AI bias
Oladele, et al. [17]	Qualitative research	Ethical and governance issues in AI-based decisions	Enhancing moral and governance frameworks in business
Roche, et al. [18]	Inclusive ethical analysis	Absence of Global South and female perspectives in AI ethics	Increasing inclusiveness and reducing bias
Astobiza, et al. [19]	Theoretical and conceptual analysis	Ethical governance of AI in the Global South: A human rights approach	Developing inclusive ethical AI governance frameworks with a focus on human rights
Segun [20]	Editorial perspective	Global engagement in AI ethics from diverse cultural perspectives	Developing intercultural AI ethics frameworks
Ormond [21]	Qualitative exploratory study	Governance of AI ethics from a Global South perspective	Developing context-sensitive AI ethics frameworks for Africa
Walker and	Theoretical and conceptual analysis	Geographies, ethics, and practices of	Developing geographically grounded perspectives on

Author	Research Methods	Research Topics	Future Directions
Winders [22]		AI	AI applications and their spatial impacts
Wu, et al. [23]	Value alignment analysis	Role of honor ethics in AI value alignment	Expanding honor ethics in global AI ethics
Airoldi [24]	Theoretical construction (based on Bourdieu's habitus), case study (IAQOS), interdisciplinary literature review	- Algorithmic socialization - Machine habitus theory - Culture in the code & code in the culture - Techno-social reproduction processes	- Develop a "sociology of algorithms" - Treat algorithms as social agents - Analyze feedback in algorithm-society interactions
Gerrard and Thornham [31]	Case study analysis	Sexist assemblages in social media content moderation	Exploring intersectionality in content moderation algorithms
Roitman [32]	Conceptual analysis	Platform economies and financial inclusion in the Global South	Addressing value subjugation and autonomization in digital platforms
Liu, et al. [33]	- Theoretical modeling - Utility-based economic modeling (temptation utility framework) - Welfare analysis and equilibrium modeling - Quantitative simulation (calibration with real-world data)	- Algorithmic inequality from behavioral vulnerabilities - Data-sharing externalities in digital platforms - Impact of data privacy regimes (GDPR-like opt-in/opt-out) - Trade-off between social welfare and protection of vulnerable consumers	- Design privacy regulation to mitigate algorithmic inequality - Address coordination failure in data-sharing decisions - Develop policies accounting for heterogeneous consumer vulnerabilities - Study dynamic implications of data sharing across generations
Lehdonvirta [34]	Interdisciplinary theoretical analysis, case studies, platform document review, qualitative interviews	- Rise of platform empires - Platform governance and legitimacy -Algorithmic management and digital labor - Platforms vs. state authority	- Democratize platform governance - Investigate digital colonialism and Global South perspectives - Develop cross-border regulatory frameworks - Study sociotechnical design and power dynamics
Hayward and Roy [35]	Critical literature review	Sustainable consumption in the global North and South	Nuanced categorization of countries; collective sustainable living perspectives
Bodó [36]	Theoretical framework development	Technological trust mediation and digital intermediation	Understanding trustworthiness of technological trust mediators
Mittelstadt, et al. [37]	Theoretical and empirical analysis (philosophical critique, toolkit experiments: FairLearn & IBM 360)	Critique of "levelling down" in algorithmic fairness and default strict egalitarianism in fairML	Proposing a shift to "levelling up" via minimum rate constraints and a harms-based framework for substantive equality
Goffi [38]	Ethical diversity analysis	Influence of Western ethical hegemony on AI	Application of multicultural ethics in AI