AI and Ethics
https://doi.org/10.1007/s43681-023-00362-w

ORIGINAL RESEARCH

AI and the quest for diversity and inclusion: a systematic literature review

Rifat Ara Shams[1] · **Didar Zowghi**[1] · **Muneera Bano**[1]

Received: 15 July 2023 / Accepted: 3 October 2023

Abstract
The pervasive presence and wide-ranging variety of artificial intelligence (AI) systems underscore the necessity for inclusivity and diversity in their design and implementation, to effectively address critical issues of fairness, trust, bias, and transparency. However, diversity and inclusion (D&I) considerations are significantly neglected in AI systems design, development, and deployment. Ignoring D&I in AI systems can cause digital redlining, discrimination, and algorithmic oppression, leading to AI systems being perceived as untrustworthy and unfair. Therefore, we conducted a systematic literature review (SLR) to identify the challenges and their corresponding solutions (guidelines/ strategies/ approaches/ practices) about D&I in AI and about the applications of AI for D&I practices. Through a rigorous search and selection, 48 relevant academic papers published from 2017 to 2022 were identified. By applying open coding on the extracted data from the selected papers, we identified 55 unique challenges and 33 unique solutions in addressing D&I in AI. We also identified 24 unique challenges and 23 unique solutions for enhancing D&I practices by AI. The result of our analysis and synthesis of the selected studies contributes to a deeper understanding of diversity and inclusion issues and considerations in the design, development and deployment of the AI ecosystem. The findings would play an important role in enhancing awareness and attracting the attention of researchers and practitioners in their quest to embed D&I principles and practices in future AI systems. This study also identifies important gaps in the research literature that will inspire future direction for researchers.

Keywords Diversity · Inclusion · Artificial intelligence · Systematic literature review

1 Introduction

Artificial intelligence (AI) has become a critical part of our society, presenting unique advantages and challenges. The ethical implications of AI, including fairness, trust, bias, and transparency are pressing issues that must be addressed. Research has indicated that AI systems can entrench and even exacerbate existing biases in areas such as criminal justice and recruitment processes [1, 2]. Maintaining trust in AI is crucial for ensuring its widespread adoption, but the blackbox nature of these systems can undermine trust [3, 4]. In response to these challenges, calls have been made for the deployment of “fairness-aware” algorithms that take demographic diversity into account and increase transparency in decision-making processes [5].

The integration of diversity and inclusion (D&I) principles in AI has the potential to mitigate the challenges posed by the lack of fairness and bias [6]. Research suggests that diverse teams increase the likelihood of recognizing and addressing biases in AI systems [1]. From a design perspective, diverse teams bring different perspectives on fairness and can identify additional sources of bias in data or algorithms [5]. From a user's standpoint, involving marginalized communities in AI development can increase the likelihood of the technology being fair and trustworthy for those groups and increase its acceptance among them [7]. Furthermore, ethical concerns for AI technology should also extend beyond privacy and transparency issues to include diversity and inclusion [8]. AI systems should not only benefit from embedding diversity and inclusion principles in their design, development, and deployment but when their development

✉ Rifat Ara Shams
rifat.shams@csiro.au

Didar Zowghi
didar.zowghi@csiro.au

Muneera Bano
muneera.bano@csiro.au

[1] CSIRO's Data61, Clayton, Victoria, Australia

Published online: 13 November 2023

is completed in this manner, they should also be treated as agents of change that could, in turn, improve and accelerate understanding and practices of diversity and inclusion in all aspects of life.

The topics of bias and fairness in AI have received significant attention in recent years. Mehrabi et al. [9] conducted a literature review on the sources of data and algorithm biases in AI applications and the different fairness definitions used to reduce bias in AI. Bertrand et al. [10] conducted an SLR on 37 papers exploring cognitive biases in Explainable AI (XAI) systems, and identified four ways cognitive biases impact XAI-assisted decisions. Another study [11] reviewed 47 articles on fairness in AI algorithms and found a lack of consensus on definitions of AI algorithmic fairness. Obermeyer et al. [12] addressed racial biases through algorithms by providing health and cost predictions for both Black and white patients. Benthall et al. [13] proposed a method for group fairness interventions using unsupervised learning to mitigate racialized social inequality, social segregation, and stratification in machine learning.

In contrast, limited research can be found that has explored the principles of diversity and inclusion (D&I) in AI. To the best of our knowledge no systematic literature review has been conducted on this topic. In this paper, therefore, we fill the above-mentioned research gap and present a systematic literature review that provides the state of the art on AI and diversity and inclusion. Our aim is to explore challenges and solutions (guidelines/strategies/approaches/practices) in the research literature focused on diversity and inclusion in AI (D&I in AI) as well as the applications of AI for enhancing and improving diversity and inclusion practices (AI for D&I). To differentiate "D&I in AI" and "AI for D&I" while extracting challenges and solutions, we followed two different definitions of these two terms. For "D&I in AI", we followed the definition provided by Zowghi and da Rimini [6]: "inclusion of humans with diverse attributes and perspectives in the data, process, system, and governance of the AI ecosystem". We defined "AI for D&I" as "the applications of AI systems to enhance the diversity and inclusion practices in environment".

The main contributions of this SLR include a rigorous search, selection and analysis of 48 articles published in the last six years (2017–2022) on the topic of D&I in AI as well as AI for D&I. We believe that the results of our exploration presented in this paper contribute to a deeper understanding of diversity and inclusion considerations in AI system development and deployment. Our findings from this SLR present:

- 55 unique challenges and 33 unique solutions about D&I in AI as well as 24 unique challenges and 23 unique solutions about the applications of AI for D&I practices.
- The number of studies on AI for D&I are significantly less than the number of studies on D&I in AI. Moreover, not all papers that state challenges also propose solution for each challenge.
- 'Gender' is the prominent diversity attribute in AI, whereas other attributes (e.g., race, ethnicity, language) are given less attention.
- 'Health' is the most discussed domain in the literature, whereas other domains such as law, banking, and transportation are ignored in the literature.
- 'Facial analysis' and 'natural language processing' are the most discussed types of AI systems to address D&I; other AI systems are ignored such as voice recognition and large language models.
- 'Governance'-related challenges and solutions are less discussed both for D&I in AI and AI for D&I.

Paper organization: Sect. 2 describes the background of this research and the related work. Section 3 briefly explains our research method and Sect. 4 reports the findings of this study. We discuss the findings in Sect. 5. Section 6 discusses the possible threats to validity of this research. Finally, the research is concluded with possible future research directions in Sect. 7.

2 Background and related work

AI has emerged as a technological force that is continuously evolving and reshaping various societal structures [14]. In recent years, there has been a heightened focus on the importance of D&I in AI [15], but the literature reveals that D&I concerns are not consistently addressed in AI projects due to the lack of operational tools, and ambiguity around responsibilities in the AI development process [16]. Neglecting D&I can have serious repercussions including harm to users and slowing AI adoption. Therefore, it is crucial for project teams and stakeholders to understand the criticality of D&I in AI. The awareness of D&I in AI will enable them to identify, monitor, and mitigate potential risks and challenges, thereby fostering an AI-literate society that can make informed decisions about the use and participation in AI systems across various contexts.

As the body of AI literature continues to expand, a growing number of traditional and systematic reviews reflects an increased focus on issues related to bias [17, 18], fairness [11, 19], transparency [20], and explainability [21]. This focus arises from the acknowledgment that AI systems have the potential to reproduce and even exacerbate existing societal biases, leading to practices that can be unfairly discriminatory [22, 23]. Bias in AI systems has roots in numerous factors, most notably the utilization of datasets that lack comprehensive representation of the entire society, leading to outcomes that are skewed [22]. Additionally, the homogeneity of AI's development community, primarily being Western and male, can unintentionally inject biases into the design and programming of AI systems [23]. Addressing this imbalance, there is a growing recognition of diversity and inclusion as critical elements in AI development that can significantly contribute to mitigating these biases [6].

Despite the acknowledged importance of diversity and inclusion, there is a gap in the literature regarding how these principles can be practically implemented in AI systems. Fosch-Villaronga and Poulsen [22] define *D&I in AI* as a *multi-faceted concept that addresses both the technical and socio-cultural aspects of AI*. They highlight *diversity* as the representation of individuals with respect to socio-political power differentials such as gender and race. *Inclusion*, they suggest, is the representation of an individual user within a set of instances, with better alignment between a user and the options relevant to them indicating greater inclusion. This concept is further analyzed at three levels: the *technical*, the *community*, and the *user*. The technical level considers if algorithms account for all necessary variables and if they classify users in a discriminatory manner. The community level examines diversity and inclusivity in AI development teams, looking at gender representation and diversity of backgrounds. Finally, the user level focuses on the intended users of the system and how the research and implementation process takes into account the stakeholders and their feedback, emphasizing the principles of Responsible Research and Innovation.

Zowghi and da Rimini [6] provide a more detailed and normative definition of D&I within the context of AI and present a set of guidelines for ensuring these principles are incorporated into the AI development process. The authors focus on a socio-technological perspective, recognizing that addressing issues of bias and unfairness requires a holistic approach that considers cultural dynamics and norms and involves end users and other stakeholders. They defined *D&I in AI* as *'inclusion' of humans with 'diverse' attributes and perspectives in the data, process, system, and governance of the AI ecosystem*. *Diversity* refers to the representation of the differences in attributes of humans in a group or society. *Attributes* are known facets of diversity including (but not limited to) the protected attributes in Article 26 of the International Covenant on Civil and Political Rights (ICCPR), as well as race, color, sex, language, religion, national or social origin, property, birth or other status, and inter-sections of these attributes. *Inclusion* is the process of proactively involving and representing the most relevant humans with diverse attributes; those who are impacted by, and have an impact on, the AI ecosystem context.

According to Zowghi and da Rimini [6], diversity and inclusion in AI can be structured and conceptualized involving five pillars: humans, data, process, system, and governance. The humans pillar focuses on the importance of including individuals with diverse attributes in all stages of AI development. The data pillar highlights the need to be mindful of potential biases in data collection and use. The process pillar emphasizes the need for diversity and inclusion considerations during the development, deployment, and evolution of AI systems. The system pillar recognizes the necessity for the AI system to be tested and monitored to ensure it does not promote non-inclusive behaviors. The governance pillar underlines the importance of structures and processes that ensure AI development is compliant with ethical principles, laws, and regulations.

There is limited literature on how AI can help in enhancing D&I [24–27], but there is no comprehensive definition in literature to present the concept. D&I in AI, and AI for D&I, create a synergistic cycle of progress that enriches both fields and their potential to effect meaningful change. AI, functioning as a mirror, reflects the patterns and prejudices ingrained in our societies, revealing biases that often go unnoticed. This heightened visibility aids in improving D&I by identifying gaps, promoting awareness, and guiding mitigation strategies. On the flip side, the integration of D&I within AI's development process is equally critical. A diverse team of creators and evaluators can identify, understand, and correct underlying biases, resulting in more equitable and inclusive AI systems. Thus, D&I and AI form a continuous, self-enhancing cycle: the use of AI advances D&I, while fostering D&I within AI development ensures more holistic, fair, and representative AI systems.

Even with these insights, many existing AI ethics guidelines remain narrowly focused on fairness, justice, and non-discrimination, with a heavy lean toward compliance-based procedures [28]. Furthermore, there is an evident gap in initiatives that aim to directly impact AI actors' behaviors and foster diversity, equity, and inclusion (DEI) awareness [29]. In terms of inclusivity, it is pertinent to note that the global discourse on AI often lacks voices and perspectives from the Global South and other underrepresented groups, with a marked dominance of Western perspectives [30]. This imbalance affects the development of ethical AI standards and calls for more inclusive practices and deeper consideration of power structures in AI policy formulation [31–33].

Despite the increased awareness of these concerns, there remains a dearth of comprehensive understanding in current research addressing these critical areas. Hence, the urgent need for a systematic literature review that investigates diversity and inclusion in AI. This approach will provide a comprehensive evaluation and synthesis of all existing research on this topic, which traditional literature reviews may fail to capture in their entirety. Consequently, it will help identify the current state of the art, define challenges and solutions, and shape future research directions, thereby addressing this critical gap in the literature.

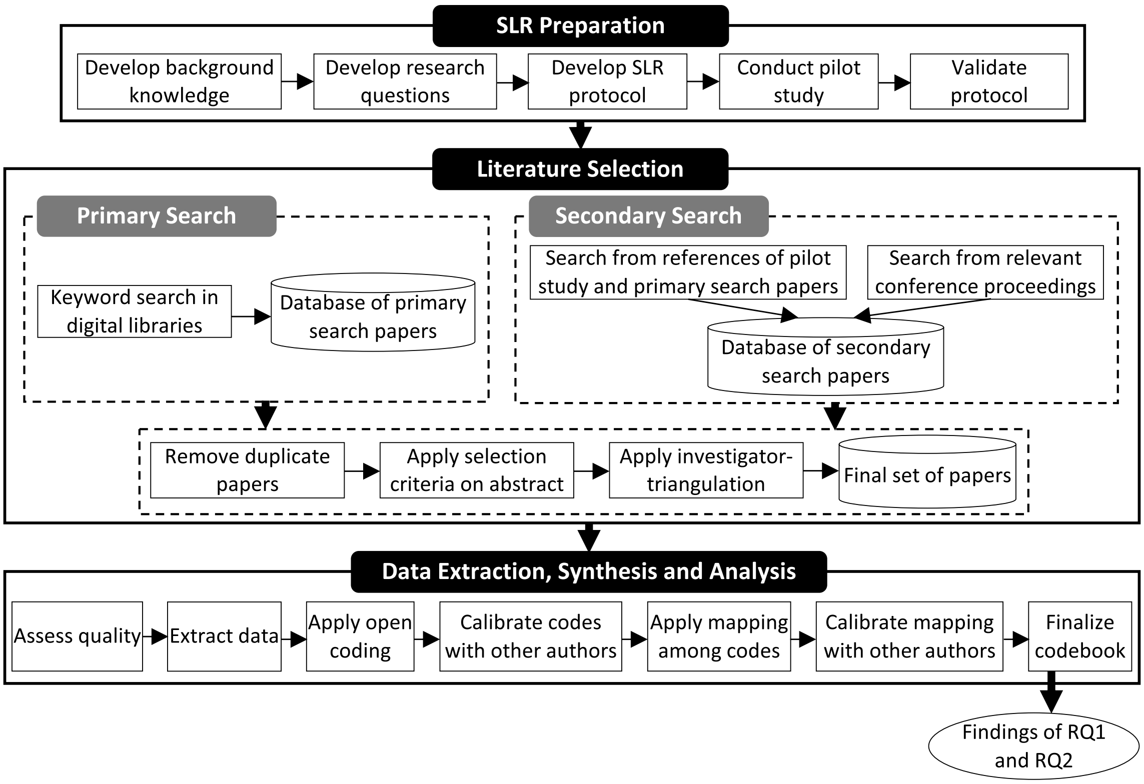


Fig. 1 An overview of the research method

3 Methodology

This study aims to explore and gain a comprehensive understanding of diversity and inclusion in the context of artificial intelligence and the use of artificial intelligence for diversity and inclusion from the published research literature. Our research was guided by the following two research questions.

RQ1. What challenges and solutions are found in the literature about diversity and inclusion in AI (D&I in AI)?

RQ2. What challenges and solutions are found in the literature about the applications of AI for diversity and inclusion practices (AI for D&I)?

We conducted a systematic literature review (SLR) in accordance with the guidelines established by Kitchenham et al. [34] to address the research questions. This approach was chosen to comprehensively identify, evaluate, and interpret existing research in this under-explored area [34]. These guidelines have also undergone numerous reviews and revisions by software engineering communities, thereby enhancing their robustness. The protocol for systematic review for this paper has also been assessed by two SLR experts, who made revisions to comply with the reliability and replicability requirements of systematic reviews. The second and third authors of this paper further augment its credibility, with their extensive experience in conducting SLRs and their publication of highly cited systematic review papers.

The methodology of the SLR is outlined in Fig. 1. The preparation stage of the SLR involved the development of a background understanding of diversity and inclusion (D&I) in AI, the formulation of research questions, the creation of an SLR protocol, and the validation of the protocol through a pilot study. The paper selection summary for the pilot study and the main study (primary search and secondary search [35]) is shown in Fig. 2. As a result of a rigorous search and selection process, we finally identified 48 papers that satisfied inclusion/exclusion criteria and are relevant to D&I in AI or AI for D&I.

To ensure the validity of the data extraction process and the relevance of the search keywords, a pilot study was conducted at the outset of the process. The search string in the five digital libraries (ACM Digital Library, IEEE Xplore, Science Direct, Scopus, and Google Scholar) was formulated using the three primary keywords relevant to the research questions: “artificial intelligence”, “machine learning”, “diversity and inclusion” (see Appendix B). In order to guide the selection of studies, clear inclusion and exclusion criteria were established. The inclusion criteria were: “papers on diversity and inclusion in AI or AI for diversity and inclusion”, “papers in the form of peer reviewed published scientific papers (journal/conference)”, and “papers published in 2017–2022”. The exclusion criteria were: “papers not related to diversity and inclusion in AI or AI for

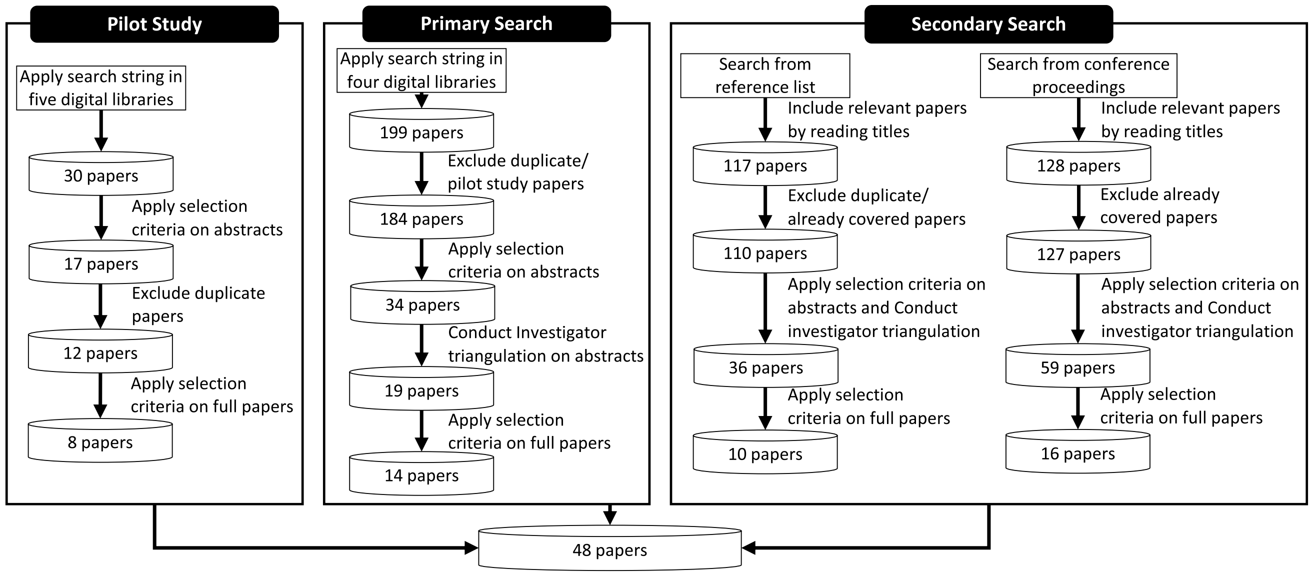


Fig. 2 SLR paper selection summary

diversity and inclusion", "literature review paper", "Tutorial/workshop paper/ArXiv paper/magazine article/book/book chapter", "Master/Ph.D. dissertations", "conference version of a study that has an extended journal version", "papers not written in English", "full papers unavailable online", and "papers already covered in the pilot study". We included papers from 2017–2022, as we did not find many relevant papers prior to 2017. Moreover, we identified only one relevant paper in 2017, whereas the majority of the studies published in 2022.

Considerations for D&I in AI or AI for D&I as a standalone topic of research are scarce in the literature. We experimented with including the terms "bias" and/or "fairness" in our search string which resulted in a very large number of papers. For example, ACM digital library returned 92 research articles on diversity and inclusion in AI (2017–2022). When we added "bias" or "fairness" to the search string, it returned 669305 articles.[1] To ensure the feasibility of the SLR, we decided to narrow the scope and remove "bias" and "fairness" from the search string. Equally, the keywords "dataset", "training", and "developer" were not incorporated into our research, although they could potentially yield results providing greater insight into the AI system development mechanisms. Nevertheless, these keywords often lead to papers that do not essentially focus on D&I. Additionally, considering the study's scope and feasibility, these keywords were deliberately excluded to prevent the return of an overwhelming number of irrelevant articles.

[1] Search date: 29/08/2022.

3.1 Primary search

In the pilot study, we used the keyword "diversity and inclusion" which was restricted to the papers that were based on both "diversity" and "inclusion". After several rounds of discussion among the authors, we decided to include all the papers on "diversity" OR "inclusion", so that no paper was left out which worked on either diversity or inclusion in AI. Therefore, we developed the final search string for our main study using the three main keywords ("diversity", "inclusion" and "artificial intelligence") and their corresponding alternatives. For example, we used "machine learning" as an alternative to "artificial intelligence". Similarly, we used two alternatives of the keyword, "inclusion": "inclusive" and "inclusiveness". The primary search was carried out with this search string in four digital libraries: ACM Digital Library, IEEE Xplore, Science Direct, and Scopus. We also applied the same search string in Google Scholar, but it provided the papers which were already covered in the above-mentioned four digital libraries. The search string was customized depending on the interfaces of different digital libraries. The details of the primary search protocol and the search results are shown in Appendix B.

After eliminating duplicates in the primary search, a total of 184 papers underwent a rigorous application of the study selection criteria on the abstracts, resulting in a selection of 34 relevant papers (Fig. 2). The next stage of selection process was guided by the principle of investigator triangulation [36], where all the authors read the 34 abstracts independently and made decisions on inclusion/exclusion. Finally, they discussed their opinions and agreed on the final selection of 19 papers which later underwent a selection process by reading the full papers. Then, the first author carefully evaluated the full text of each of the included studies and excluded 5 papers, as they were found to be irrelevant to the research questions (such as diverse literature, diverse algorithms, diverse technology). Finally, 14 papers were selected from the primary search for data extraction.

3.2 Secondary search

The secondary searches involved a manual examination of the titles of the references listed in the selected pilot and primary studies. In addition, a manual scan was performed on the proceedings of two most frequent conferences where the pilot and primary studies were published: ACM Conference on Fairness, Accountability, and Transparency and AAAI/ACM Conference on AI, Ethics, and Society. After removing the duplicate papers and the papers already covered in the pilot study and primary search, we came up with a total of 237 papers (110 from the reference list and 127 from the conference proceedings). Then, study selection criteria were applied to the abstracts to yield 95 papers from the secondary search. Investigator triangulation was also met to validate our selection. Then, the first author evaluated the full text of each of the included studies and excluded 69 papers from the secondary search due to their irrelevance to our research objectives, despite appearing promising from their abstracts. Finally, we selected 26 papers from the secondary search (10 from the reference list and 16 from conference proceedings). This provided a total of 48 papers for this SLR for data extraction (see Fig. 2 and the full list of included papers in Appendix 1).

3.3 Quality assessment

To assess the quality of the selected papers, we employed the five-question assessment criteria proposed by Liu et al. [37]. These questions assess the clarity of research aims, appropriateness of research design, clarity of findings and contributions, description of limitations and future work, and empirical nature of the study. Each question was evaluated on a scale of 0 to 1, with 0 indicating "no", 0.5 indicating "partly", and 1 indicating "yes". The overall quality score was calculated by summing the scores of the five questions, and the papers were classified as Good: if the score is between 3 and 4, Fair: if the score is between 2 and 3, Poor: if the score is between 0 and 2. Out of the 48 selected papers, 32 were deemed "Good" quality, 11 were "Fair" quality, and 5 were "Poor" quality, demonstrating the robustness of this review.

3.4 Data extraction

Excel spreadsheet and NVivo software were used to extract demographic and content-related data from the 48 selected papers on D&I in AI and AI for D&I. The demographic data included the source of the paper, title, abstract, authors, affiliated countries of authors, year of publication, venue, and citation. Content-related data included the challenges faced to address D&I in AI and AI for D&I, and the proposed/used solutions (guidelines/ strategies/ approaches/ practices) to those challenges. The data were extracted through manual coding by the first author and cross-checked in weekly meetings with the other authors.

3.5 Data synthesis and analysis

The data synthesis and analysis for RQ1 and RQ2 is outlined in Fig. 1. To answer RQ1 and RQ2, the first author employed open coding to identify the challenges about D&I in AI and AI for D&I, as well as the proposed guidelines/ strategies/ approaches/ practices to address the challenges. The first author established intra-rater reliability by revisiting and cross-checking all the papers and coded data multiple times. All the authors checked the challenges and solutions to ensure inter-rater reliability and had several iterations of discussions to finalize them. Throughout this process, the first author went back to the papers several times to validate the established findings. The solutions were then mapped with the challenges for each paper to get a comprehensive understanding of what guidelines/ strategies/ approaches/ practices are taken for a specific challenge. The initial mapping analysis was undertaken by the first and second authors independently, with each of the contributing authors involved in the review process. The results were finalized after numerous revisions through iterative discussions among all the authors, enabling a consensus to be reached on the final mapping outcome.

4 Results

This section presents the results of the systematic literature review starting with the demographics of our selected 48 studies. We further present the extracted challenges of addressing diversity and inclusion in AI (D&I in AI) and enhancing diversity and inclusion practices in the environment through AI (AI for D&I), as well as the mentioned solutions to address the challenges.

4.1 Demographics

Demographics covers a range of elements, including the publication year, citation count, whether the studies were

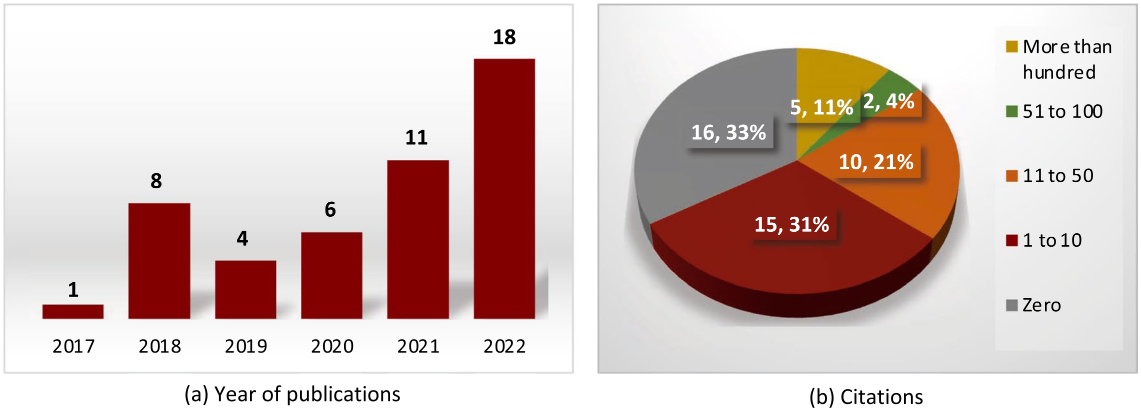


Fig. 3 Year of publications and citations of the selected 48 papers

Table 1 Attributes of diversity and their corresponding paper IDs

Attributes	Number of studies	Paper IDs
Gender	23	S2, S3, S6, S10, S11, S13, S15, S20, S21, S26, S27, S28, S29, S30, S31, S33, S36, S37, S39, S40, S43, S45, S48
Sex	1	S10
Age	6	S6, S10, S15, S21, S28, S33
Race	15	S3, S6, S7, S10, S15, S17, S20, S21, S23, S25, S26, S28, S33, S36, S41
Ethnicity	3	S3, S10, S26
Disability	4	S12, S14, S38, S42
Neurodiversity	1	S32
Skin tone	1	S13, S18, S25, S26
Geographic location	2	S21, S35
Family income and insurance status	1	S25
Language	1	S33

empirical or theoretical in nature, the attributes of diversity addressed in each paper, as well as the affiliated countries of first authors.

The publication year and citations of the 48 selected papers are depicted in Fig. 3. The data reveal that the majority of the papers (18) were published in 2022, followed by 11 papers published in 2021. Only one paper was published in 2017. This trend suggests that the field of D&I in AI is relatively new, and further research in this area is needed. With regards to citations, Fig. 3 reveals that although we only covered last six years, five papers received more than 100 citations, while two papers received 51–100 citations. We also identified the number of empirical and non-empirical studies among the 48 selected studies. 30 of the selected studies are empirical and 18 are non-empirical.

The attributes of diversity analyzed in the selected studies, such as gender, age, and race, are depicted in Table 1. We differentiated the terms "gender" and "sex" in this table based on the terms used in the selected studies. According to Walker et al., "Sex refers to the anatomical or chromosomal categories of male and female. Gender refers to socially constructed roles that are related to sex distinctions" [38]. The results suggest that the majority of the papers focus on gender (23 papers). There are also a good number of papers (15) on race, leaving room for further research on other attributes of diversity, such as age, sex, disability, neurodiversity, geographic location, skin tone, language, and ethnicity.

We also explored the ratio of affiliated countries of the first authors of the selected 48 studies which is presented in Fig. 4. The presence of United States of America (USA) is the maximum (29 out of 48), which reveals that the majority of the D&I in AI or AI for D&I-related work has been conducted in USA. Three of the first authors are affiliated with United Kingdom (UK), two are affiliated with China, Canada, and Belgium each. Rest of the countries have only one occurrence each such as Thailand, Netherlands, Turkey, Qatar, India, Ireland, Japan, Sweden, Australia, and Germany. This findings reveal that diversity and inclusion in AI is the limited explored research area worldwide except

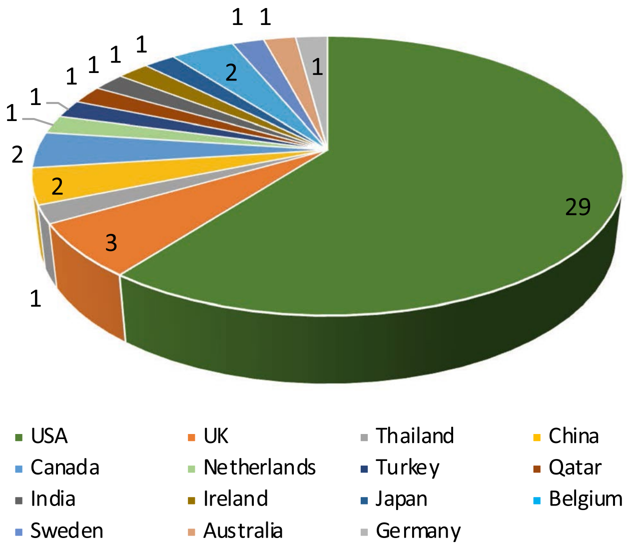


Fig. 4 Affiliated countries of first authors of the selected papers

USA. Therefore, this area should be focused more in future research.

4.2 RQ1: challenges and solutions about diversity and inclusion in AI (D&I in AI)

Table 2 presents the list of challenges about D&I in AI with their corresponding challenge IDs and paper IDs. We identified 55 unique challenges. Among the selected 48 papers, we identified challenges about D&I in AI from 36 papers. We also identified 33 unique solutions to address some of those challenges as shown in Table 3. Among the total of 48 papers, 23 papers discussed the solutions to the specific challenge mentioned. We also mapped the challenges with their corresponding solutions for each of the papers as presented in Appendix C. Some illustrative quotations on challenges and solutions for RQ1 as well some illustrative quotations on their mapping are presented below.

Illustrative quotations on challenges.

Challenge C14: (Lack of diverse race, ethnicity, sex and gender inclusion and representation in the design, development, and implementation of AI system). *"Lack of consideration for race, ethnicity, sex and gender in the design, development, and implementation of AI system in healthcare can lead to marginalization of underrepresented groups from benefiting from such technologies."*- S10

Challenge C15: (The lack of Equity, Diversity, and Inclusion (EDI) principles and indicators). *"The lack of EDI principles and indicators, for example, the presence of sex/gender, and racial/ethnicity bias in healthcare can be defined as differential medical and healthcare delivery and treatment of men, women, non-binary people and one race (dominant) over the others, the impact of which may be positive, negative, or neutral."*- S10

Illustrative quotations on solutions.

Solution L10: (Consider purpose and definition of gender before gender classification in facial analysis or image labeling). *"Before embedding gender classification into a facial analysis service or incorporating gender into image labeling, it is important to consider what purpose gender is serving. Furthermore, it is important to consider how gender will be defined, and whether that perspective is unnecessarily exclusionary (e.g., binary)."*- S11

Solution L9: (Develop policies to prevent discriminatory and nonconsensual gender representations in FA (Facial Analysis) systems). *"Establishing policies for how biometric data and face and body images are collected and used may be the most effective way of mitigating harm to trans people-and also people of marginalized races, ethnicities, and sexualities. Polices that prevent discriminatory and non-consensual gender representations could prevent gender misrepresentation from being incorporated into FA systems in both the data and infrastructure by regulating the use of gender as a category in algorithmic systems. For example, by banning the use of gender from FA-powered advertising and marketing."*- S11

Illustrative quotations on mapping of challenges and solutions.

Challenge C52: (Racial categories are ill-defined in computer vision systems). *"Racial categories are ill-defined, arbitrary and implicitly tied loosely to geographic origin. Second, given that racial categories are implicitly references to geographic origin, their extremely broad, continent-spanning construction would result in individuals with drastically different appearances and ethnic identities being grouped incongruously into the same racial category if the racial categories were interpreted literally. Thus, racial categories must be understood both as references to geographic origin as well as physical characteristics."*- S41

Solution L32 to address C52: (Adopt fair computer vision datasets with different racial categories). *"We empirically study the representation of race through racial categories in fair computer vision datasets, and analyze the crossdataset generalization of these racial categories, as well as their cross-dataset consistency, stereotyping, and self-consistency."*- S41

Table 2 Results of RQ1: challenges about diversity and inclusion in AI

Challenge ID	Challenges	Paper ID
C1	Concern about diversity without considering inclusion in AI	S1
C2	Lack of women role models in AI ecosystem	S2, S43
C3	Lack of culture where women feel welcomed within the AI field	S2
C4	Lack of education on diversity, equity or disparities	S2
C5	Socio-cultural norms and biases and stereotypes about women in AI	S2
C6	Lack of diversity in the composition of AI teams	S2, S10
C7	Lack of Equity, Diversity and Inclusion (EDI) considerations in data set	S5, S8, S10
C8	Lack of diversity and inclusion issues in AI ethics documentation	S5
C9	Algorithm bias in human face processing technology	S7
C10	Data handlers are influenced by their own backgrounds and prejudices	S9
C11	Certain communities' voices are disregarded and not uplifted in AI practice	S9, S36, S37, S44
C12	Bias in the training data sets	S9, S17, S27
C13	Implicit social bias in AI around ageism	S10
C14	Lack of diverse race, ethnicity, sex and gender inclusion and representation in the design, development, and implementation of AI system	S10
C15	Lack of Equity, Diversity, and Inclusion (EDI) principles and indicators	S10
C16	Lack of a definitive and inclusive definition of Equity, Diversity, and Inclusion (EDI)	S10
C17	Under-representation of minority groups in sampling during model training and testing	S10, S25, S42
C18	Unconscious biases by AI, which is a manifestation of its creators' biases	S10, S15, S17, S29
C19	Unrealisticness of having universally inclusive dataset	S11
C20	Lack of broader definitions of any given gender in gender classification system	S11
C21	Absence of an explicit gender classifier in facial analysis (FA) services	S11
C22	Biases in model design, training, or implementation	S11, S25
C23	AI industry is dominated by men	S15
C24	Hiring AI developer and other technical roles is not gender-neutral	S15
C25	Inequalities in training data and lack of balanced inclusion data in machine algorithms	S15, S34
C26	Insufficiency of "diversity" and "inclusion" related terminologies	S16
C27	Lack of AI researcher diversity	S17
C28	Inability to recognize the bias in data source due to lack of knowledge	S17
C29	Data invisibility, incomplete data, missing data	S17
C30	Lack of equitable standards to address diversity in algorithm	S17
C31	Lack of comprehensive and accurate collection and generation of demographic data	S17
C32	The underrepresentation of dark skin images in training data for diagnosing skin pathology	S18
C33	Lack of accuracy in unsupervised classification of diverse audiences	S21
C34	Traditional design processes, tools, and methods are difficult to effectively analyze and deal with complex diversity factors	S22
C35	Neglecting the level of diversity required by contextualized user-sensitive design	S22
C36	Less attention on equity and justice principles in AI design and development	S23
C37	Limited focus on diversity, equity, or disparities in AI-based academic literature	S24
C38	Stakeholder roles and experiences are overlooked in equitable AI design, implementation, and use	S24
C39	Unstability of race and ethnic data labels due to the inconsistency of racial and ethnic categories across geographies	S26
C40	Internal and subjective identities are not discussed when designing algorithms	S27
C41	Subjective nature of gender identity	S27
C42	Algorithm learning discriminatory behavior from human behavior	S29
C43	Ad algorithms are not gender-neutral	S29
C44	Humans' unreliability of, or weakness in, spotting subconscious biases in AI systems	S30
C45	Stereotypical gender concepts are embedded in the data	S31
C46	Systematic imbalance of gender representation in the design of AI systems	S31, S36
C47	Gender and racial discrimination in and by AI development team	S33

Table 2 (continued)

Challenge ID	Challenges	Paper ID
C48	Computer vision systems are not inclusive for all people from different demographics	S35
C49	Overlooking disability considerations in ethical or legal levels of AI algorithms	S38
C50	No comprehensive analysis of how gender is theorized in natural language processing	S39
C51	Less examples of feminist and participatory methodologies to address power inequalities	S40
C52	Racial categories are ill-defined in computer vision systems	S41
C53	Lack of data from disabled populations to train AI systems	S42
C54	Lack of trust in data accuracy in AI-enabled hiring software	S46
C55	Difficulties in measuring diversity in algorithm	S47

Table 3 Results of RQ1: solutions to address the challenges of diversity and inclusion in AI

Solution ID	Solutions	Paper ID
L1	Implement awareness campaigns that tackle socio-cultural norms and biases and stereotypes	S2
L2	Arrange unconscious bias training for teachers and counselors	S2
L3	Foster female-identifying role models in AI	S2, S31
L4	Promote diversity and inclusion by developing methods and tools with diverse datasets that bring diversity and inclusion into engineering practice	S5
L5	Use Lenovo face recognition engine (LeFace) to achieve better performance of racial fairness	S7
L6	Adopt data disaggregation by demographic groups	S8
L7	Adopt more data examples in training for better learning outcomes	S9
L8	Adopt participatory design in AI system development, ensuring that all relevant stakeholders are represented/participated	S10, S27, S37
L9	Develop policies to prevent discriminatory and non-consensual gender representations in FA (Facial Analysis) systems	S11
L10	Consider purpose and definition of gender before gender classification in facial analysis or image labeling	S11
L11	Use diverse and inclusive data to train AI systems to be inclusive	S15
L12	Remove discriminatory bias from job descriptions and resumes	S15
L13	Configure AI to equitably screen candidates by disregarding age, gender, and race in profile assessment	S15
L14	Use Word2Vec approach to visualize related terminologies	S16
L15	Establish accurate standards for the collection of detailed demographic data	S17
L16	Assess data used for models to avoid amplifying and perpetuating racial bias	S17
L17	Analyze publicly available skin image repositories to quantify the underrepresentation of darker skin tones	S18
L18	Enhance diversity-oriented design capacity to increase inclusiveness of diversity requirements	S22
L19	Establish a user-centered machine learning system based on user and context features	S22
L20	Apply new tools, processes, and methods that algorithmically provide appropriate responses to specific needs	S22
L21	Designers require learning on diversity-oriented design	S22
L22	Integrate EDI (Equity, Diversity and Inclusion) and racial justice principles and practice in AI health	S23
L23	Build a responsible culture in innovation and establish ethical building blocks for reliable delivery of equitable AI	S23
L24	Arrange training and education on the use of AI tools for equity promotion	S24
L25	Include diverse voices in training data and design	S25, S27
L26	Partner with ethicists and antiracism experts in developing, training, testing, and implementing models	S25
L27	Let users define their own gender identity before designing AI systems	S27
L28	Use fairness indicators (e.g., harmful label association, geographical diversity and fairness, same-attribute assessment via similarity search) to probe main sources of biases in computer vision models	S35
L29	Use consistent, respectful, and accurate language for gender	S39
L30	Use feminist research methodologies	S39
L31	Adopt framework of data feminism to co-design datasets and machine learning models	S40
L32	Adopt fair computer vision datasets with different racial categories	S41
L33	Adopt diversity by design, by operationalizing and implementing diversity-aware chatbot	S44

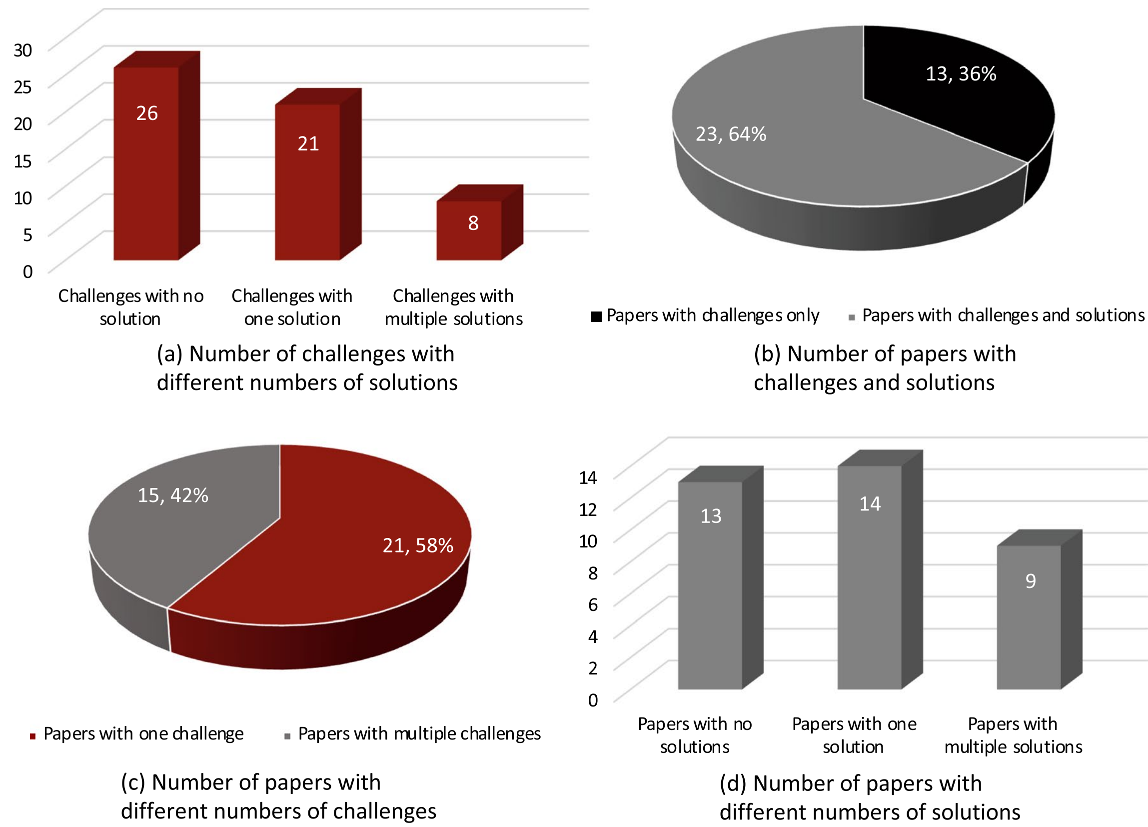


Fig. 5 Analysis of the findings of RQ1

We have further analyzed to explore the findings from the mapping of challenges and solutions. According to Fig. 5(a), nearly half of the challenges (26) have no associated proposed solutions (e.g., C1, C3, C6, C23). 21 challenges have one solution each and the rest of the 8 challenges have more than one solutions. Figure 5(b) shows the number of papers that have challenges with and without solutions. Among the 36 papers that identified the challenges about D&I in AI (see Appendix C), 13 papers discussed about challenges with no solutions such as S1, S21, S42. The rest of the 23 papers discussed the challenges with possible solutions such as S2, S8, S23. Figure 5(c) presents the number of papers with different numbers of challenges. More than half of the papers (21 papers) have only one challenge each (e.g., S1, S7, S47). The rest of the 15 papers mentioned more than one challenges such as S2, S10, S29. The last pie chart (see Fig. 5(d)) shows the ratio of papers with no solution, one solution and multiple solutions. Majority of the papers (14 papers) have one solution each (e.g., S7, S16, S40). However, there are a large number of papers (13 papers) which did not propose any solution at all such as S29, S36, S46. On the other hand, 9 papers proposed more than one solution for the challenges such as S2, S15, S22.

According to Appendix C, the paper S10 discussed the maximum number of challenges (8) about D&I in AI such as "Lack of Equity, Diversity, and Inclusion (EDI) principles and indicators", "Under-representation of minority groups in sampling during model training and testing". The paper S17 also discussed a large number of challenges (7) such as "Bias in the training data sets" and "Lack of comprehensive and accurate collection and generation of demographic data". The paper S22 provided the maximum number of solutions (4) to address the challenges about D&I in AI. For example, "Enhance diversity-oriented design capacity to increase inclusiveness of diversity requirements" and "Establish a user-centered machine learning system based on user and context features". Similarly, three papers (S2, S15, S27) provided three solutions each to address the challenges about D&I in AI. Some of the papers provided multiple solutions for one challenge. For example, S22 provided three solutions for the challenges C35. Similarly, the papers S15, S23, S25 and S39 provided two solutions each to address one challenge (C24, C36, C22, C50, respectively).

We also identified some of the challenges which have been mentioned by more than one paper (see Table 2). For example, C11 and C18 have been mentioned by four papers. Similarly, three papers discussed each of the challenges of C7, C12, and C17. Similar to the challenges,

Table 4 Results of RQ2: challenges of the applications of AI for diversity and inclusion practices

Challenge ID	Challenges	Paper ID
H1	Lack of ability of facial recognition software to assess racial and ethnic diversity in qualitative medical studies	S3
H2	Bias by AI in workplace	S4
H3	AI-based decisions exhibit discrimination based on sensitive attributes such as age, gender, and race	S6
H4	Bias by AI in decisions related to hiring, compensation, and promotion	S8
H5	Gender classifier is not used to mitigate gender bias	S11
H6	Underrepresented genders are not acknowledged by gender classification systems	S11, S13
H7	Inaccurate data label detection	S11
H8	Gender data labeling by gender classification systems offers limited labels to third-party developers	S11
H9	Lack of machine learning process to engage data by co-researchers with learning disabilities (LDs)	S12
H10	Bias by machine algorithms within a diverse pool of personnel	S15
H11	Delayed or incorrect diagnoses of skin cancer for the people of color by early detection system	S18
H12	Lack of use of machine learning technology in organizational diversity research	S19
H13	AI replaces certain jobs that are predominantly held by underrepresented groups	S20
H14	Difficulties in understanding through AI how important Africans, women, and young people are in protecting, restoring, and promoting the sustainable use of terrestrial ecosystems in Africa	S21
H15	Difficulties in identifying corresponding design patterns by machine learning technology after changing design requirements and problems	S22
H16	Less accuracy of facial recognition technology to identify non-binary gender	S27
H17	Difficulties in accuracy of diversity attributes detection in face detection tools	S28
H18	Lack of use of AI technology in delivering support to the children with autism	S32
H19	Difficulties in analyzing disability requirements for AI in recruitment	S38
H20	Bias in hiring with AI toward the people with disability	S38
H21	Disability is not widely studied in mitigation of bias in AI algorithms on ethical, legal or technical levels	S38
H22	Disability-based discrimination by AI technologies	S42
H23	Difficulties in estimating diversity in a given dataset	S47
H24	Lack of use of AI in understanding the diversity of people in any social media activism campaign	S48

three solutions (L3, L8, L25) have been discussed by multiple papers (see Table 3).

4.3 RQ2: challenges and solutions about the applications of AI for diversity and inclusion practices (AI for D&I)

20 out of 48 papers focused on the applications of AI for enhancing D&I practices (AI for D&I). Table 4 presents the list of 24 unique challenges about AI for D&I. The solutions to address the challenges with corresponding paper IDs are presented in Table 5, where we identified 23 solutions. The mapping of challenges with their corresponding solutions for each paper is shown in Appendix D. Some illustrative quotations on challenges and solutions for RQ2 as well some illustrative quotations on their mapping are presented below.

Illustrative quotation on challenge.

Challenge H6: (Underrepresented genders are not acknowledged by gender classification systems). *"When classifying gender, designers of the systems we studied chose to use only two predefined demographic gender categories: male and female. As a result, these presentations are recorded, measured, classified, labeled, and databased for future iterations of binary gender classification."*- S11

Illustrative quotation on solution.

Solution N1: (Use Betaface (Betaface.com) facial analysis software to determine the diversity attributes). *"To determine the rates of diversity within departments, Betaface facial analysis software was used to analyze photos taken from the hospitals' websites. This software was able to determine the race, ethnicity, and gender of the care providers."*- S3

Illustrative quotations on mapping of challenges and solutions.

Challenge H16: (Less accuracy of facial recognition technology to identify non-binary gender). *"This work which positions transgender faces as problematic to facial recognition accuracy, also raised ethical issues related to user privacy as the data for the database was scraped from transgender individuals' videos without their consent or knowledge."*- S27

Table 5 Results of RQ2: solutions of the challenges of AI applications for diversity and inclusion practices

Solution ID	Solutions	Paper ID
N1	Use Betaface (Betaface.com) facial analysis software to determine the diversity attributes	S3
N2	Use AI to bring in the appropriate knowledge on bias reduction techniques and methods	S4
N3	Use AI to adopt fairness standards	S4
N4	Use AI to make use of the relevant information and methods on data and algorithm development	S4
N5	Use data analytics tools to improve decisions related to hiring, compensation, promotion, and retention, which can advance Diversity, Equity and Inclusion (DEI) practice	S8
N6	Design Facial analysis services with the knowledge of the negative consequences of not recognizing genders correctly	S11
N7	Consider feature-based data labeling during design of Facial analysis services	S11
N8	Consider how to provide gender classification functionality to third-party developers with more scrutiny and oversight	S11
N9	Use qualitative data for labeling in gender classification	S11
N10	Adopt five-steps machine learning-based structural topic modeling (STM) co-analysis process for creative, inclusive, and critical engagement of data by co-researchers	S12
N11	Assist in diverse candidate selection via social media by AI	S15
N12	Adopt AI-based skin cancer early detection system for all skin tones using clinical images	S18
N13	Adopt unsupervised machine learning models to text mine and analyze how organizations communicate about Diversity, Equity and Inclusion (DEI) topics	S19
N14	Adopt topic modeling to statistically identify word groups underlying all the diversity documents	S19
N15	Use a combination of self-reported location data, computer vision of social media photographs, natural language processing to estimate the demographics of individuals participating in the Global Landscapes Forum (GLF)	S21
N16	Adopt inclusive design tools, processes, and methods combined with machine learning technology to identify corresponding design patterns	S22
N17	Train automatic gender recognition (AGR) with a variety of gender identities early in the design process, by working with diverse teammembers and adopting participatory design approaches to identify non-binary gender	S27
N18	Use face detection tools such as Face++, IBM Bluemix Visual Recognition, AWS Rekognition, and Microsoft Azure Face API to detect diversity attributes	S28
N19	Use ECHOES that utilizes an AI virtual character to facilitate autistic children's ability to engage in social interaction	S32
N20	Adopt analytical roadmap named "recruitment AI" to help mitigating the bias toward people with disability through ethical, legal and technical analysis	S38
N21	Develop AI-powered accessibility tools to raise accessibility awareness in AI	S42
N22	Adopt an approach called "algorithmic greenlining" to use diversity estimates instead of true diversity scores	S47
N23	Use AI to analyze the diversity attributes from social media data	S48

Solution N17 to address H16: (Train automatic gender recognition (AGR) with social and ethical implications). *"While AGR technology is still in its infancy, the recent integration of facial recognition into already pervasive technologies suggest it could impact large numbers of people in the near future. As technologists continue to develop AGR applications, it is important to understand the social and ethical implications of widespread adoption."*- S27

We have also explored additional findings from the mapping of challenges and solutions for RQ2, which is presented in Fig. 6. As shown in Fig. 6(a), the majority of the challenges (16) have one solution each such as H1, H10, H24. On the other hand, four challenges have no solution at all (H3, H5, H13, H19) and another four challenges have more than one solutions (H2, H7, H8, H12). Out of 20 papers on AI for D&I, majority of the papers (17 papers) provided both challenges and solutions on the applications of AI for enhancing D&I practices (see Fig. 6(b)). Three papers discussed challenges without any solutions. According to Fig. 6(c), two papers (S11, S38) discussed more than one challenge, whereas the rest of the 18 papers provided only one challenge each. S11 discussed the maximum number of challenges (4). As shown in Fig. 6(d), majority of the papers (14 out of 20) provided one solution each to address the challenges related to AI for D&I. On the other hand, three papers did not propose any solution and another three papers provided multiple solutions for the challenges.

4.4 Diversity attributes

Figure 7 illustrates the ratio of diversity attributes (e.g., age, gender, race, ethnicity) discussed in the challenges and solutions about D&I in AI (RQ1). Majority of the challenges (56%) and solutions (54%) did not mention about any attributes at all. Gender has the maximum occurrences (25% for challenges and 23% for solutions). Race is the second highest attribute that was discussed in 7% of the challenges and 14% of the solutions.

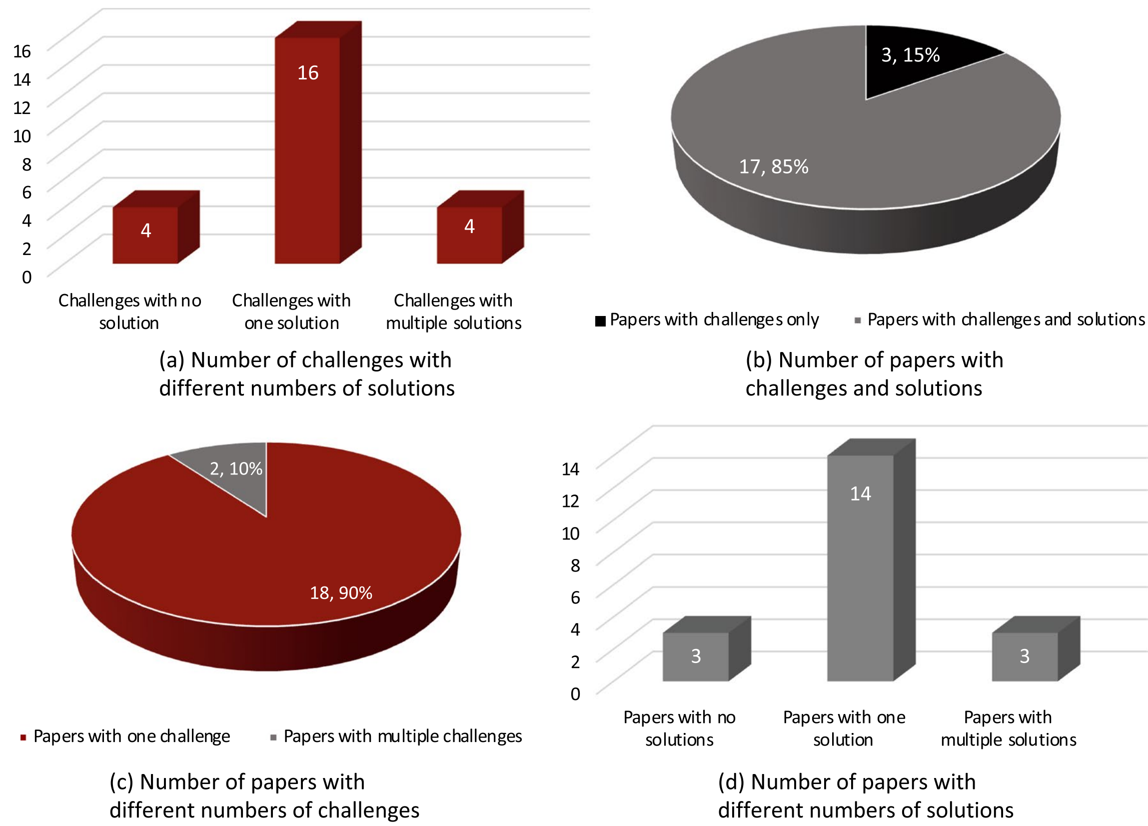


Fig. 6 Analysis of the findings of RQ2

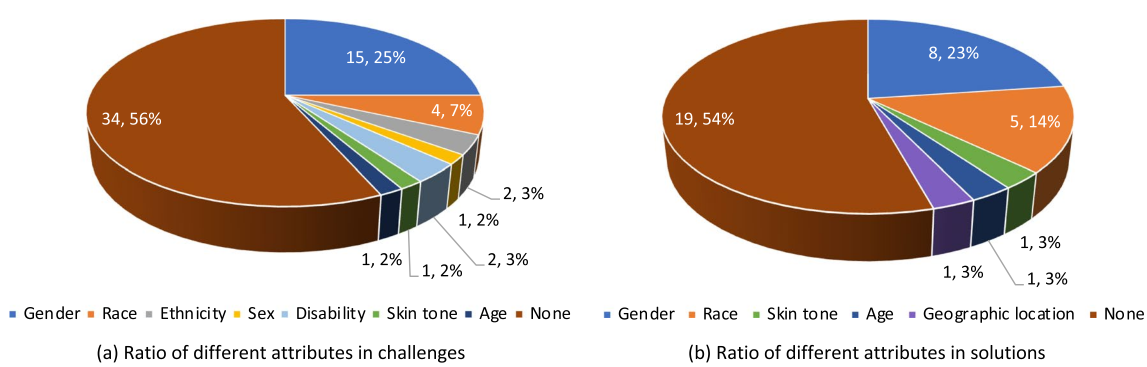
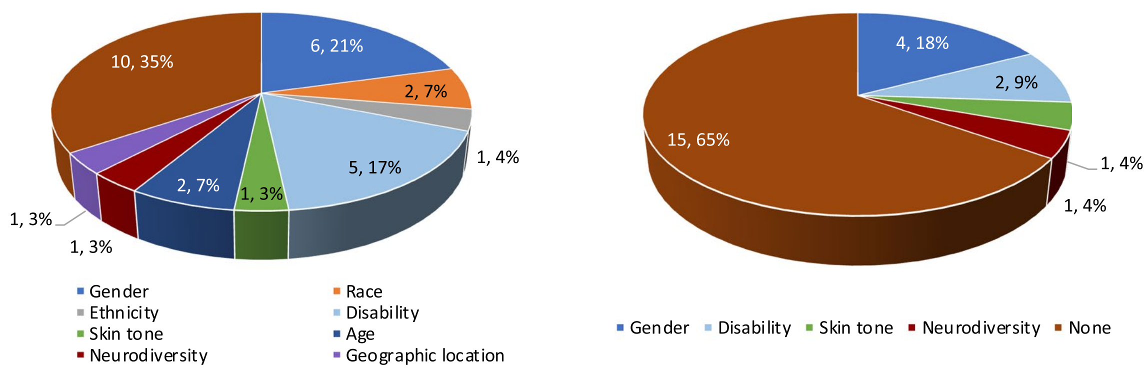


Fig. 7 Ratio of diversity attributes in challenges and solutions for RQ1

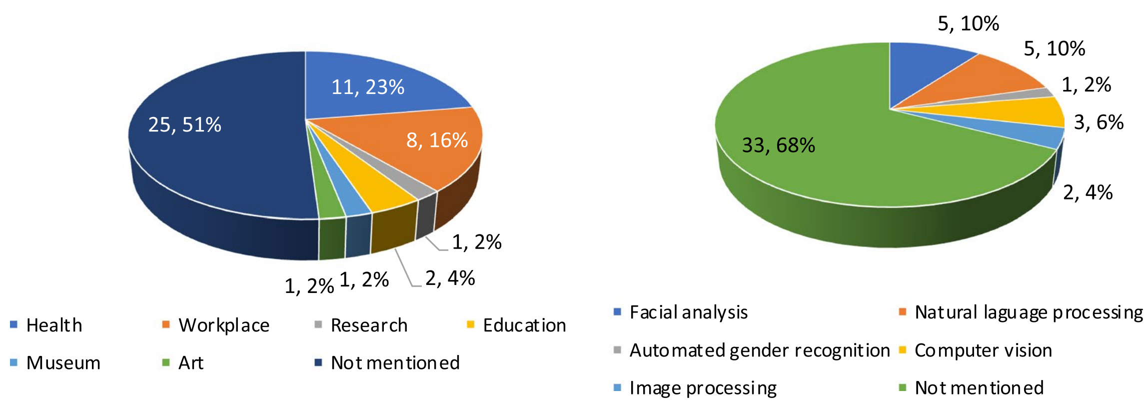
Figure 8 illustrates the ratio of diversity attributes (e.g., age, gender, race, ethnicity) mentioned in the challenges and solutions about the applications of AI for D&I practices (RQ2). According to Fig. 8(a), gender and disability are the two attributes that have the most occurrences (21% for gender and 17% for disability) in challenges. On the other hand, ethnicity, skin tone, neurodiversity, and geographic location have the least occurrences. Figure 8(b) shows that majority of the solutions (65%) do not indicate any attribute explicitly. However, gender has the maximum occurrences (18%), whereas skin tone and neurodiversity have the least occurrences.



(a) Ratio of different attributes in challenges

(b) Ratio of different attributes in solutions

Fig. 8 Ratio of diversity attributes in challenges and solutions for RQ2



(a) Number of papers on different domains

(b) Number of papers on different types of AI systems

Fig. 9 Ratio of different domains and types of AI systems

5 Discussion and implications

5.1 Highlights of the results

'Gender' as the most discussed attribute of diversity. Our analysis reveals that gender has been the top explored diversity attribute, in 23 out of the 48 papers (refer to Table 1). As for the other dimensions of diversity, 15 papers delved into race, 6 investigated age, 4 explored disability, 3 looked into ethnicity, 2 touched on geographic location, while sex, neurodiversity, skin tone, family income and insurance status, and language each took the spotlight in only a single paper. Moreover, when considering the challenges and solutions pertaining to D&I in AI and AI for D&I, gender has been the predominant topic of discussion (see Figs. 7 and 8). For addressing D&I in AI (RQ1), gender was the focus in 15 out of 55 challenges and 8 out of 33 solutions. Similarly, in the context of enhancing D&I practices through AI (RQ2), 6 out of 24 challenges and 4 out of 23 solutions emphasized gender. The other diversity dimensions are largely overlooked. Recent studies [39] have shed light on the challenges women face in AI, including bias, discrimination, a lack of self-confidence, inadequate resources and support, and limited exposure to AI in early education. Another study [40] highlighted the difficulties faced by gender classifiers in recognizing non-binary genders. Despite these studies, there exists a dearth of research addressing other facets of diversity, like age, disability, race, ethnicity, and language. None of our included 48 studies worked on four of the attributes of diversity which was mentioned in the Article 26 of the International Covenant on Civil and Political Rights (ICCPR): religion, birth or other status, property, and national or social origin. Some recent federal laws of different countries such as Australian discrimination law [41] also discussed diversity attributes. None of our selected papers focused on many of the diversity attributes in Australian federal laws on discrimination such as religion, political opinion, and marital status. This

underscores the necessity for more extensive investigations into these areas and the necessity to consider a broader spectrum of diversity, especially the notion of intersectionality, in both AI research and practice.

'Health' as the most discussed domain. Some of the included studies worked on D&I in AI and AI for D&I for some specific domains such as health, workplace, and education. Figure 9(a) shows the ratio of different domains mentioned in the selected papers. More than half of the papers (51%) do not focus on any specific domain, rather they discussed diversity and inclusion in AI in general. However, 'health' is the most discussed domain, 23% of the papers focused specifically on this domain. The second highest is 'workplace' (16%). Only a small number of papers mentioned about other domains such as 'education', 'research', 'museum', and 'art'. As many important domains such as law, banking, and transportation were not focused in any of the paper, more research is needed.

'Facial analysis' and 'natural language processing' as the most discussed type of AI system. Fig. 9(b) illustrates the ratio of different types of AI systems, which were discussed in the selected studies of this SLR. Majority of the papers (68%) did not mention any specific AI systems. Similar number of papers (10%) discussed about facial analysis system and natural language processing system. 6% papers focused on computer vision system, 4% on image processing, and 2% on automated gender recognition system. Other types of AI systems must be studied with the lens of D&I such as voice recognition and large language models.

Global North as predominant region on D&I in AI concept. Being a societal construct, the notions of diversity and inclusion often do not play significant roles in many countries. Many equity, diversity, and inclusion (EDI) policies, initiated in the Global North, address this constructive concept by promoting enhanced representation of Black, Asian, and Minority Ethnic groups within the workforce [42]. However, the Global South does not showcase a similar predominant focus on the concept of diversity and inclusion. Therefore, the field of D&I in AI research exhibits a notable deficiency in geographic diversity, particularly from regions in the Global South. This deficiency results in an insufficient appreciation of various diversity attributes such as language, ethnicity, race, and nationality within the AI ecosystem. Furthermore, the challenges and solutions we have identified based on diversity and inclusion within the AI ecosystem do not adequately represent the unique conditions prevalent in the global South. Consequently, this infers that the specific challenges and solutions pertaining to D&I in AI or AI for D&I, within the context of the global South are yet to be distinctly recognized and documented. Therefore, this represents a substantial gap in this research area, highlighting an urgent requirement for significant improvement.

The correlation between authors' geographic locations and the progression of AI development. USA is the pioneer of AI development [43] and the affiliated geographic location of majority of the authors is also USA. Therefore, it can be argued that the geographic location of the researchers is directly proportional to the location leading for AI development. However, this assertion lacks empirical evidence. For example, China also holds a dominant position in AI development [43], but they are noticeably behind in D &I in AI research. Therefore, this area should be focused more in future research to develop a comprehensive understanding on the issues related to D&I in AI.

Lack of solutions to address D&I in AI. Number of solutions are less than the number of challenges about D&I in AI (55 challenges, 33 solutions). Figure 5(a) also shows that 26 out of 55 challenges have no solution to offer at all. Similarly, Fig. 5(b) shows that 36% papers do not have any solution, whereas all of the papers discussed challenges. Moreover, a large number (18 out of 48) of selected studies are non-empirical. This implies that proposed solutions are not implemented or validated in real settings.

The area of diversity and inclusion in artificial intelligence is a relatively new and less-explored field, with a limited number of studies undertaken. Consequently, there are fewer identified solutions for the challenges linked to D&I in AI or AI for D&I. This scenario is further amplified by minimal awareness about D&I-related issues within AI, leading to a scarcity of solutions to tackle such challenges. Furthermore, D&I principles have not been widely implemented within AI systems, contributing to the limited understanding among researchers and practitioners on how to mitigate associated challenges. The lack of related research emerging from the Global South also plays a part in the deficit of solutions for the D&I challenges rising within this geographical region. The collective impact of these issues underscores the urgent need for further evidence-based intensive research in this area to recommend more solutions to address the challenges. While the existing literature offers solutions for some of the D&I issues in AI, not every challenge has been addressed. AI researchers and developers can leverage these identified gaps to concentrate more on proposing solutions for the challenges presented in addressing D&I in AI and AI for D&I. In an effort to enhance collective problem-solving, we have plans to publicly share the existing challenges identified and their correlated solutions for the benefit of larger audiences facing similar issues.

Insufficient collaborations between developers and researchers. The diversity of AI system developers is a critical factor. If the people who develop AI systems lack diversity, it is likely that the resulting AI systems will mirror this homogeneity. On the other hand, while researchers studying these systems may identify issues related to D&I in AI or AI for D&I, proposing solutions may be challenging since they are not directly engaged in the AI development process. This difficulty could contribute to the relative scarcity of solutions compared to the identified challenges in D&I within AI. However, if the researchers come from diverse backgrounds, they could leverage their varying perspectives to interpret challenges and propose potential solutions. By fostering collaborative relationships between diverse AI developers and researchers, their combined skill sets can be utilized to uncover more D&I issues in AI and propose tailored solutions. This collaboration could ultimately lead to the development of AI systems that are both diverse and inclusive. However, our SLR has not given a clear evidence and positive indication whether this issue has an impact on the lack of solutions to address the challenges of D&I in AI or AI for D&I. We hope that our paper will serve as a bridge, connecting individuals and prompting more widespread exploration of challenges and potential solutions in this field.

Limited research on AI for D&I. Our literature review shows that the majority of the selected papers (36 out of 48) discussed the challenges and some corresponding solutions to address D&I in AI. On the other hand, a few papers (20 papers) discussed the challenges and solutions to enhance D&I practices by AI (AI for D&I). Similarly, the number of solutions to consider D&I in AI is higher that the number of solutions to address AI for D&I. The findings indicate that AI researchers are aware to address D&I in AI, whereas AI for D&I has taken limited attention. Although some recent studies worked on enhancing D&I practices in workplace through AI [14, 25] and enhancing D&I practices in automated gender recognition systems [40, 44], further research needs to be conducted for more comprehensive understanding on AI for D&I.

Low hanging fruits. Our results revealed that various challenges could be tackled immediately with regard to diversity and inclusion in AI. For instance, including the perspectives of marginalized communities, such as individuals with disabilities and the elderly, in the development process, can support more representation in the training data [31, 45]. This can address various challenges, including the "Under-representation of minority groups in sampling during model training and testing" (S10, S25, S42), "Certain communities' voices are disregarded and not uplifted in AI practice" (S9, S36, S37, S44), "Lack of comprehensive and accurate collection and generation of demographic data" (S17), "Overlooking disability considerations in ethical or legal levels of AI algorithms" (S38). Additionally, promoting diversity in the recruitment of AI development teams and among researchers can help combat unconscious biases [39, 45–47]. Raising awareness and promoting education about diversity, equity, and disparities in AI can assist mitigating the knowledge gap about the people, places, and factors that make up the data [39, 47].

5.2 Five pillars of diversity and inclusion in AI

According to Zowghi and da Rimini, the definition of D&I in AI consists of five pillars: *Humans*, *Data*, *Process*, *System*, and *Governance* [6]. We categorized our findings under the five pillars to explore the coverage of the challenges and solutions from this SLR within AI ecosystem and the pillars. We used these pillars for cross analysis and applied thematic coding on the findings for RQ1 and RQ2 to structure the challenges and solutions for D&I in AI and AI for D&I under the five pillars. It should be noted that the challenges and solutions for RQ1 and RQ2 are not necessarily mutually exclusive in relation to the five pillars. Therefore, many of them are listed under more than one pillar. This process was conducted independently by all of the authors and one external expert. An iterative series of discussions were conducted between all authors and the external annotator to ensure that the findings for answering RQ1 and RQ2 were accurately represented under their corresponding pillars. As all of the annotators have previous experience and expertise to this area and they analyzed the challenges and solutions from different disciplinary lens, we did not disregard any of their opinions. Therefore, we took the larger set which means we combined all the pillars categorized by all the annotators. The findings of our analysis are shown in Appendix E. Some examples of challenges and solutions for RQ1 and RQ2 under the five pillars are given below.

Humans: (C11) Certain communities' voices are disregarded and not uplifted in AI practice. - S9, S36, S37, S44

Data: (L6) Adopt data disaggregation by demographic groups. - S8

Process: (H15) Difficulties in identifying corresponding design patterns by machine learning technology after changing design requirements and problems. - S22

System: (N1) Use Betaface (Betaface.com) facial analysis software to determine the diversity attributes. - S3

Governance: (C15) Lack of Equity, Diversity, and Inclusion (EDI) principles and indicators. - S10

The frequencies of different pillars for the challenges and solutions for RQ1 (D&I in AI) and RQ2 (AI for D&I) are illustrated in Fig. 10. The findings revealed that *Human*, not surprisingly, has the maximum occurrences for the challenges about D&I in AI (RQ1) (see Fig. 10(a)). However, *Process* is the highly addressed pillar in solutions to address the challenges about D&I in AI (see Fig. 10(b)). *System* is the most occurred pillar for both challenges and solutions for AI for D&I (RQ2) (see Fig. 10(c and d)), whereas *System* was mentioned less for RQ1.

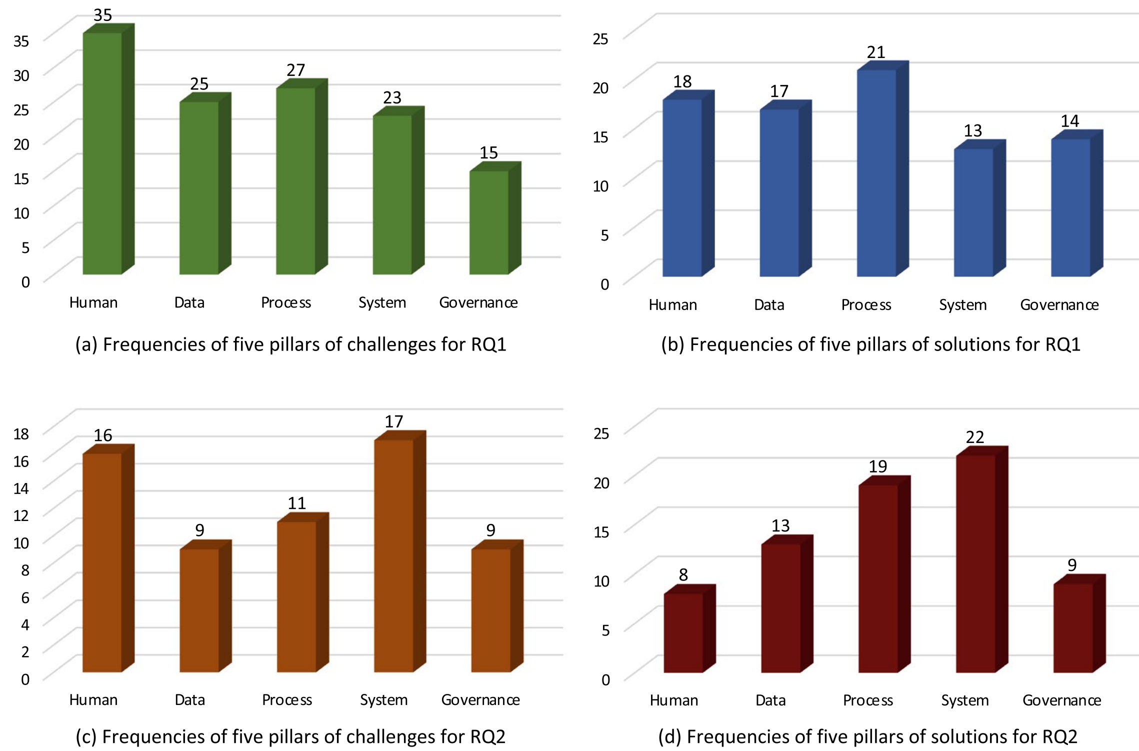


Fig. 10 Frequencies of five pillars for the challenges and solutions for D&I in AI (RQ1) and AI for D&I (RQ2)

On the other hand, we identified the least number of *Governance* related challenges for both RQ1 and RQ2. *Data* related challenges are also minimum for RQ2, whereas many challenges mentioned about *Data* for RQ1. Surprisingly, *Human* was mentioned the least for the solutions for RQ2.

As presented in Appendix E, the literature is limited in its coverage of diversity and inclusion in relation to governance of AI systems. Only a small number of studies mention the governance-related challenges and solutions associated with addressing D&I in AI and AI for D&I, such as "lack of Equity, Diversity, and Inclusion (EDI) principles and indicators (C15)", "Integrate EDI (Equity, Diversity and Inclusion) and racial justice principles and practice in AI health (L22)", "Disability is not widely studied in mitigation of bias in AI algorithms on ethical, legal or technical levels (H21)", "Use AI to adopt fairness standards (N3)". This is likely due to the fact that establishing D&I principles and standards for AI systems often requires long-term planning, whereas addressing the challenges associated with humans, data, process, and systems can be addressed in less time. Therefore, it is crucial that policymakers be made aware of the importance of D&I in AI to establish adequate plans for AI governance (such as standards, regulations, and policies) and principles to address these issues.

5.3 Implications for inclusive AI systems development

In recent years, the importance of diversity and inclusion in AI and the corresponding have become increasingly acknowledged by researchers. Many challenges to address D&I and AI for D&I have been discussed in literature with various proposed solutions. One key solution is to raise awareness and provide training on cultural competency and algorithmic vigilance [39, 48]. This could help address socio-cultural norms, human biases, and stereotypes that may be embedded within AI systems [39, 47]. Another solution involves mitigating bias from job descriptions and resumes through training AI systems to disregard certain demographic information, such as age, gender, and race, while assessing profiles [46].

Inclusive design practices have also been suggested as a way to address D&I in AI. This could involve adopting participatory design processes that involve diverse communities in the data collection and design process [31]. Another approach involves combining inclusive design tools and methods with machine learning technology to changes design requirements and identify corresponding design patterns [49]. Additionally, policy makers have a crucial role to play in addressing D&I in AI. One suggestion is to establish more explicit policy documentation to ensure transparency on the policies [45].

Although we extracted and presented paper-wise solutions to address the challenges of D&I in AI and AI for D&I, some solutions from different papers could address a specific challenge. For example, the challenge, "Underrepresented genders are not acknowledged by gender classification systems (H6)" identified from the paper S11 and S13, could be addressed by the solutions from different papers such as "Train automatic gender recognition (AGR) with a variety of gender identities early in the design process, by working with diverse teammembers and adopting participatory design approaches to identify non-binary gender (N17)" (S27). This, along with other solutions, can help to ensure that AI systems are designed and developed in a manner that is inclusive and equitable for all.

6 Threats to validity

Limitations. Although we have rigorously adhered to the comprehensive search strategy dictated by the evidence-based SLR guidelines, ensuring a comprehensive selection of our samples, there's still a possibility that certain papers might not have been incorporated into our data collection. This may result from their inaccessibility or non-existence on electronic platforms, of which we might be unaware.

In the creation of our search strings, the key terms "fairness", "bias", "dataset", "training", and "developer" were deliberately omitted based on the insights from our pilot study and testing, with the objective of minimizing a large number of unrelated results. While we recognize this could have excluded certain relevant papers from our sample, we employed a meticulous secondary search strategy to counterbalance this limitation. This strategy, we believe, largely made up for the potential drawbacks of not using these terms initially. Nonetheless, we accept the possibility that some potentially relevant research might have been missed due to this strategic decision, though we stand firm in the overall effectiveness of our implemented research approach.

Another shortcoming of this paper is the absence of a detailed analysis of the diversity attributes of each author from all the selected papers. Thoroughly examining all the diversity attributes of the authors of the selected papers would undoubtedly provide us with more comprehensive insights. However, accurately identifying every diversity characteristic of all authors is impossible. Additionally, the process also carries a substantial risk of misidentification. For instance, gender identifiers do not always identify gender correctly.

Internal validity. A potential threat could arise from the small number of selected papers and the restricted time span. As D&I in AI and AI for D&I are relatively new fields of research, we did not find many relevant papers prior to 2017. The majority of the papers were published recently (2022), and only 1 paper was published in 2017. However, in future studies, we will expand our time frame to check if there are more studies in this area. Another significant threat could arise from the bias in study selection and bias in data extraction. However, we mitigated this threat by adopting the investigator triangulation technique.

Construct validity. A potential construct threat could arise from the irrelevance of the selected papers with our research objectives. We selected many papers by reading the abstracts where there was a chance of getting information about D&I in AI or AI for D&I. However, many of them were removed after reading the full papers due to their irrelevance with our objectives. There is another potential threat to the subjective interpretation of the extracted data. Both of the threats were mitigated by adopting the investigator triangulation technique. In addition, conducting the preliminary mapping analysis of challenges and their associated solutions solely by the first author could potentially present a construct threat. Nevertheless, this threat was mitigated by incorporating all the authors in the revision process through several iterations of discussions.

External validity. An external threat could arise from the generalizability of our findings. Although the results of this SLR may not be generalized for all types of AI technology, they can be considered representative within the specific domain of AI system development.

7 Conclusions and future work

We conducted a systematic literature review with the goal to develop a comprehensive understanding of the challenges and corresponding solutions in addressing diversity and inclusion in artificial intelligence (D&I in AI) and enhancing diversity and inclusion practices by artificial intelligence (AI for D&I). After a rigorous process, we selected 48 academic papers published from 2017 to 2022, from which we extracted data and applied open coding on the data to explore information relevant to the challenges and solutions. Finally, we identified 55 unique challenges and 33 unique solutions in addressing D&I in AI, and 24 unique challenges and 23 unique solutions in addressing AI for D&I.

The analysis of the findings revealed that the integration of AI with diversity and inclusion is a less-explored area of research, as we found only a limited number of papers. Majority of these studies discussed the challenges of addressing D&I in AI, but provided limited attention to the solutions to address those challenges. Moreover, a large number of solutions were proposed by some non-empirical studies without any implementation or validation in real life settings. Our study reveals that there is a lack of guidance for operationalizing the proposed solutions. We identified less challenges and solutions to address AI for D&I from a limited number of papers compared to the number of challenges and solutions to address D&I in AI. Hence, further research is required on AI for D&I in particular and solutions of challenges for D&I in AI.

Our results suggest that ‘gender’ is the most discussed attribute of diversity in AI, which leads to the necessity of further research on other attributes such as race, ethnicity, language, ageism, and religion. Similarly, ‘health’ is the most discussed domain, and ‘facial analysis’ and ‘natural language processing’ are the most discussed AI systems in the analyzed literature on D&I in AI and AI for D&I, whereas other domains and types of AI systems are significantly ignored. We also identified that *Governance* related issues are less discussed in the challenges and solutions to address D&I in AI and AI for D&I.

The results of our SLR have provided much-needed evidence for the advocacy of embedding and integrating D&I practices and principles in the AI ecosystem. The gaps in the literature identified are the starting point for our holistic and comprehensive approach to tackling the D&I related issues in the overall AI ethics and Responsible AI body of knowledge. We have recognized the need for D&I in AI guidelines and as a result, parallel to the conduct of this SLR, we have also performed a multi-vocal analysis of academic and gray literature to develop a comprehensive set of guidelines [6]. Our next step is to design and develop a risk-based framework for practitioners from the findings of this SLR that would incorporate a risk assessment checklist and context-specific recommendations for tackling the related issues at different stages of the AI development lifecycle. Our plan will include co-designing this framework by applying human-centered design and evidence-based approaches involving AI practitioners and relevant stakeholders.

Appendix A: list of 48 included studies

S1 Mitchell, Margaret, Dylan Baker, Nyalleng Moorosi, Emily Denton, Ben Hutchinson, Alex Hanna, Timnit Gebru, and Jamie Morgenstern. “Diversity and inclusion metrics in subset selection.” In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, pp. 117-123. 2020.

S2 Roopaei, Mehdi, Justine Horst, Emilee Klaas, Gwen Foster, Tammy J. Salmon-Stephens, and Jodean Grunow. “Women in ai: Barriers and solutions.” In 2021 IEEE World AI IoT Congress (AIIoT), pp. 0497-0503. IEEE, 2021.

S3 Mathis, Michelle S., Tosin E. Badewa, Ruth N. Obiarinze, Linda T. Wilkinson, and Colin A. Martin. “A novel use of artificial intelligence to examine diversity and hospital performance.” Journal of Surgical Research 260 (2021): 377-382.

S4 Tongkachok, Korakod, Shaifali Garg, Veena Prasad Vemuri, Vijesh Chaudhary, Poonam Vitthal Koli, and K. Suresh Kumar. “The Role of Artificial Intelligence on Organisational support Programmes to Enhance work outcome and Employees Behaviour.” Materials Today: Proceedings 56 (2022): 2383-2387.

S5 Chi, Nicole, Emma Lurie, and Deirdre K. Mulligan. “Reconfiguring diversity and inclusion for AI ethics.” In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 447-457. 2021.

S6 Srinivasan, Ramya, and Kanji Uchino. “The Role of Arts in Shaping AI Ethics.” In AAAI Workshop on reframing diversity in AI: Representation, inclusion, and power. CEUR Workshop Proceedings (CEUR-WS. org). 2021.

S7 Shi, Sheng, Shanshan Wei, Zhongchao Shi, Yangzhou Du, Wei Fan, Jianping Fan, Yolanda Conyers, and Feiyu Xu. “Algorithm Bias Detection and Mitigation in Lenovo Face Recognition Engine.” In Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14-18, 2020, Proceedings, Part II 9, pp. 442-453. Springer International Publishing, 2020.

S8 Chauhan, Preeti S., and Nir Kshetri. “The Role of Data and Artificial Intelligence in Driving Diversity, Equity, and Inclusion.” Computer 55, no. 4 (2022): 88-93.

S9 Huang, Han-Yin, and Cynthia CS Liem. “Social Inclusion in Curated Contexts: Insights from Museum Practices.” In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 300-309. 2022.

S10 Nyariro, Milka, Elham Emami, and Samira Abbasgholizadeh Rahimi. “Integrating Equity, Diversity, and Inclusion throughout the lifecycle of Artificial Intelligence in health.” In 13th Augmented Human International Conference, pp. 1-4. 2022.

S11 Scheuerman, Morgan Klaus, Jacob M. Paul, and Jed R. Brubaker. “How computers see gender: An evaluation of gender classification in commercial facial analysis services.” Proceedings of the ACM on Human-Computer Interaction 3, no. CSCW (2019): 1-33.

S12 Chapko, Dorota, Pedro Andrés Andrés Pérez Rothstein, Lizzie Emeh, Pino Frumiento, Donald Kennedy, David McNicholas, Ifeoma Orjiekwe et al. “Supporting Remote Survey Data Analysis by Co-researchers with

Learning Disabilities through Inclusive and Creative Practices and Data Science Approaches." In Designing Interactive Systems Conference 2021, pp. 1668-1681. 2021.

S13 Celis, L. Elisa, and Vijay Keswani. "Implicit diversity in image summarization." Proceedings of the ACM on Human-Computer Interaction 4, no. CSCW2 (2020): 1-28.

S14 Kurnaz, Sefer, and Maalim AH Aljabery. "Predict the type of hearing aid of audiology patients using data mining techniques." In Proceedings of the Fourth International Conference on Engineering MIS 2018, pp. 1-6. 2018.

S15 Jora, Rachna Bansal, Kavneet Kaur Sodhi, Prabhat Mittal, and Parul Saxena. "Role of Artificial Intelligence (AI) In meeting Diversity, Equality and Inclusion (DEI) Goals." In 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 1687-1690. IEEE, 2022.

S16 Bhaduri, Sreyoshi, and Tamoghna Roy. "A word-space visualization approach to study college of engineering mission statements." In 2017 IEEE Frontiers in Education Conference (FIE), pp. 1-5. IEEE, 2017.

S17 Dankwa-Mullan, Irene, and Dilhan Weeraratne. "Artificial intelligence and machine learning technologies in cancer care: Addressing disparities, bias, and data diversity." Cancer Discovery 12, no. 6 (2022): 1423-1427.

S18 Rezk, Eman, Mohamed Eltorki, and Wael El-Dakhakhni. "Leveraging Artificial Intelligence to Improve the Diversity of Dermatological Skin Color Pathology: Protocol for an Algorithm Development and Validation Study." JMIR Research Protocols 11, no. 3 (2022): e34896.

S19 Wang, Wei, Julie V. Dinh, Kisha S. Jones, Siddharth Upadhyay, and Jun Yang. "Corporate diversity statements and employees' online DEI ratings: An unsupervised machine-learning text-mining analysis." Journal of Business and Psychology (2022): 1-17.

S20 Ozkazanc-Pan, Banu. "Diversity and future of work: inequality abound or opportunities for all?." Management Decision 59, no. 11 (2021): 2645-2659.

S21 Brandt, John, Kathleen Buckingham, Cody Buntain, Will Anderson, Sabin Ray, John-Rob Pool, and Natasha Ferrari. "Identifying social media user demographics and topic diversity with computational social science: a case study of a major international policy forum." Journal of Computational Social Science 3 (2020): 167-188.

S22 Li, Fang, Hua Dong, and Long Liu. "Using AI to Enable Design for Diversity: A Perspective." In Advances in Industrial Design: Proceedings of the AHFE 2020 Virtual Conferences on Design for Inclusion, Affective and Pleasurable Design, Interdisciplinary Practice in Industrial Design, Kansei Engineering, and Human Factors for Apparel and Textile Engineering, July 16-20, 2020, USA, pp. 77-84. Springer International Publishing, 2020.

S23 Dankwa-Mullan, Irene, Elisabeth Lee Scheufele, Michael E. Matheny, Yuri Quintana, Wendy W. Chapman, Gretchen Jackson, and Brett R. South. "A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle." Journal of Health Care for the Poor and Underserved 32, no. 2 (2021): 300-317.

S24 Clark, Cheryl R., Consuelo Hopkins Wilkins, Jorge A. Rodriguez, Anita M. Preininger, Joyce Harris, Spencer DesAutels, Hema Karunakaram, Kyu Rhee, David W. Bates, and Irene Dankwa-Mullan. "Health care equity in the use of advanced analytics and artificial intelligence technologies in primary care." Journal of General Internal Medicine 36 (2021): 3188-3193.

S25 Sveen, William, Maya Dewan, and Judith W. Dexheimer. "The Risk of Coding Racism into Pediatric Sepsis Care: The Necessity of Antiracism in Machine Learning." The Journal of Pediatrics 247 (2022): 129-132.

S26 Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." In Conference on fairness, accountability and transparency, pp. 77-91. PMLR, 2018.

S27 Hamidi, Foad, Morgan Klaus Scheuerman, and Stacy M. Branham. "Gender recognition or gender reductionism? The social implications of embedded gender recognition systems." In Proceedings of the 2018 chi conference on human factors in computing systems, pp. 1-13. 2018.

S28 Jung, Soon-Gyo, Jisun An, Haewoon Kwak, Joni Salminen, and Bernard Jim Jansen. "Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race." In Twelfth international AAAI conference on web and social media. 2018.

S29 Lambrecht, Anja, and Catherine E. Tucker. "Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM careerads." An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads (March 9, 2018) (2018).

S30 Cohen, Tammy. "How to leverage artificial intelligence to meet your diversity goals." Strategic HR Review (2019).

S31 Leavy, Susan. "Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning." In Proceedings of the 1st international workshop on gender equality in software engineering, pp. 14-16. 2018.

S32 Porayska-Pomsta, Kaśka, Alyssa M. Alcorn, Katerina Avramides, Sandra Beale, Sara Bernardini, Mary Ellen Foster, Christopher Frauenberger et al. "Blending human and artificial intelligence to support autistic children's social communication skills." ACM Transactions on Computer-Human Interaction (TOCHI) 25, no. 6 (2018): 1-35.

S33 Kong, Youjin. "Are "Intersectionally Fair" AI Algorithms Really Fair to Women of Color? A Philosophical Analysis." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 485-494. 2022.

S34 Salem, Jad, Deven Desai, and Swati Gupta. "Don't let Ricci v. DeStefano Hold You Back: A Bias-Aware Legal Solution to the Hiring Paradox." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 651-666. 2022.

S35 Goyal, Priya, Adriana Romero Soriano, Caner Hazirbas, Levent Sagun, and Nicolas Usunier. "Fairness indicators for systematic assessments of visual feature extractors." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 70-88. 2022.

S36 S36 Hirota, Yusuke, Yuta Nakashima, and Noa Garcia. "Gender and racial bias in visual question answering datasets." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 1280-1292. 2022.

S37 Vlasceanu, Madalina, Miroslav Dudik, and Ida Momennejad. "Interdisciplinarity, Gender Diversity, and Network Structure Predict the Centrality of AI Organizations." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 1-10. 2022.

S38 Buyl, Maarten, Christina Cociancig, Cristina Frattone, and Nele Roekens. "Tackling Algorithmic Disability Discrimination in the Hiring Process: An Ethical, Legal and Technical Analysis." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 1071-1082. 2022.

S39 Devinney, Hannah, Jenny Björklund, and Henrik Björklund. "Theories of "gender" in NLP bias research." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 2083-2102. 2022.

S40 Suresh, Harini, Rajiv Movva, Amelia Lee Dogan, Rahul Bhargava, Isadora Cruxen, Ángeles Martinez Cuba, Guilia Taurino, Wonyoung So, and Catherine D'Ignazio. "Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection." In 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 667-678. 2022.

S41 Khan, Zaid, and Yun Fu. "One label, one billion faces: Usage and consistency of racial categories in computer vision." In Proceedings of the 2021 acm conference on fairness, accountability, and transparency, pp. 587-597. 2021.

S42 Park, Joon Sung, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. "Designing an online infrastructure for collecting AI data from people with disabilities." In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 52-63. 2021.

S43 Cheong, Marc, Kobi Leins, and Simon Coghlan. "Computer science communities: Who is speaking, and who is listening to the women? Using an ethics of care to promote diverse voices." In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 106-115. 2021.

S44 Helm, Paula, Loizos Michael, and Laura Schelenz. "Diversity by Design? Balancing the Inclusion and Protection of Users in an Online Social Platform." In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, pp. 324-334. 2022.

S45 Siapka, Anastasia. "Towards a Feminist Metaethics of AI." In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, pp. 665-674. 2022.

S46 Li, Lan, Tina Lassiter, Joohee Oh, and Min Kyung Lee. "Algorithmic hiring in practice: Recruiter and HR Professional's perspectives on AI use in hiring." In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 166-176. 2021.

S47 Borgs, Christian, Jennifer Chayes, Nika Haghtalab, Adam Tauman Kalai, and Ellen Vitercik. "Algorithmic greenlining: An approach to increase diversity." In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 69-76. 2019.

S48 Karbasian, Habib, Hemant Purohit, Rajat Handa, Aqdas Malik, and Aditya Johri. "Real-time inference of user types to assist with more inclusive and diverse social media activism campaigns." In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 171-177. 2018.

Appendix B: search string in digital libraries and their corresponding results for pilot and main study

Pilot study

Digital library	Search string	Search within	Time frame	Search time	No. of papers returned	Selected by reading abstract
ACM Digital Library	[[Abstract: "artificial intelligence"] OR [Abstract: "machine learning"]] AND [Abstract: "diversity and inclusion"]	Research article (Abstract)	2017–2022	15/05/2022 (2:47 am)	2	2
IEEE Xplore	("Abstract":"artificial intelligence" OR "Abstract":"machine learning") AND ("Abstract":"diversity and inclusion")	Conference, Journal (Abstract)	2017–2022	15/05/2022 (2:47 am)	2	2
Science Direct	("artificial intelligence" OR "machine learning") AND ("diversity and inclusion")	Research articles (Title, Abstract, Keywords)	2017–2022	15/05/2022 (2:58 am)	8	2
Scopus	(TITLE-ABS-KEY ("artificial intelligence" OR "machine learning") AND TITLE-ABS-KEY ("diversity and inclusion"))	Conference paper, Article (Title, Abstract, Keywords)	2017–2022	15/05/2022 (1:02 pm)	13	6
Google Scholar	allintitle: artificial AND intelligence AND diversity AND inclusion	All (Title, Abstract, Keywords)	2017–2022	15/05/2022 (1:46 pm)	5	5

Main study

Digital library	Search string	Search within	Time frame	Search time	No. of papers
ACM Digital Library	[[Abstract: "artificial intelligence"] OR [Abstract: "machine learning"]] AND [Abstract: diversity] AND [[Abstract: inclusion] OR [Abstract: inclusive] OR [Abstract: inclusiveness]]	Research article (Abstract)	2017–2022	11/07/2022 (4:23 pm)	92
IEEE Xplore	("Abstract":"artificial intelligence" OR "Abstract":"machine learning") AND ("Abstract":diversity) AND ("Abstract":inclusion OR "Abstract":inclusive OR "Abstract":inclusiveness)	Conference, Journal (Abstract)	2017–2022	07/07/2022 (3:29 am)	8
Science Direct	("artificial intelligence" OR "machine learning") AND "diversity" AND ("inclusion" OR "inclusive" OR "inclusiveness")	Research articles (Title, Abstract, Keywords)	2017–2022	12/07/2022 (5:49 pm)	12
Scopus	(TITLE-ABS-KEY ("artificial intelligence" OR "machine learning") AND TITLE-ABS-KEY (diversity) AND TITLE-ABS-KEY (inclusion OR inclusive OR inclusiveness))	Conference paper, Article (Title, Abstract, Keywords)	2017–2022	12/07/2022 (6:33 pm)	87

Appendix C: mapping of challenges with corresponding solutions about D&I in AI (RQ1)

(NoS=No solution, N/A=Not applicable, N=None, G=Gender, S=Sex, A=Age, R=Race, E=Ethnicity, D=Disability, K=Skin tone, L=Geographic location)

Paper ID	Challenge ID	Attributes of challenges	Solution ID	Attributes of solutions
S1	C1	N	NoS	N/A
S2	C2	G	L3	G
S2	C3	G	NoS	N/A
S2	C4	N	L2	N
S2	C5	G	L1	N
S2	C6	N	NoS	N/A
S5	C7	N	L4	N
S5	C8	N	NoS	N/A
S7	C9	N	L5	R
S8	C7	N	L6	N
S9	C10	N	NoS	N/A
S9	C11	N	NoS	N/A
S9	C12	N	L7	N
S10	C13	A	NoS	N/A
S10	C14	R, E, S, G	L8	N
S10	C15	N	NoS	N/A
S10	C16	N	NoS	N/A
S10	C17	N	NoS	N/A
S10	C18	N	NoS	N/A
S10	C6	N	NoS	N/A
S10	C7	N	NoS	N/A
S11	C19	N	NoS	N/A
S11	C20	G	L10	G
S11	C21	G	L9	G
S11	C22	N	NoS	N/A
S15	C23	G	NoS	N/A
S15	C24	G	L12	N
S15	C24	G	L13	A, G, R
S15	C25	N	L11	N
S15	C18	N	NoS	N/A
S16	C26	N	L14	N
S17	C27	N	NoS	N/A
S17	C28	N	NoS	N/A
S17	C29	N	NoS	N/A
S17	C30	N	L15	N
S17	C31	N	L15	N
S17	C12	N	L16	R
S17	C18	N	NoS	N/A
S18	C32	K	L17	K
S21	C33	N	NoS	N/A
S22	C34	N	L20	N
S22	C35	N	L18	N
S22	C35	N	L19	N
S22	C35	N	L21	N
S23	C36	N	L22	R
S23	C36	N	L23	N
S24	C37	N	L24	N
S24	C38	N	NoS	N/A
S25	C17	N	L25	N
S25	C22	N	L25	N
S25	C22	N	L26	N
S26	C39	R, E	NoS	N/A
S27	C40	N	L8	N
S27	C41	G	L27	G
S27	C12	N	L25	N
S29	C42	N	NoS	N/A
S29	C43	G	NoS	N/A
S29	C18	N	NoS	N/A
S30	C44	N	NoS	N/A
S31	C45	G	NoS	N/A
S31	C46	G	L3	G
S33	C47	G, R	NoS	N/A
S34	C25	N	NoS	N/A
S35	C48	N	L28	L
S36	C11	N	NoS	N/A
S36	C46	G	NoS	N/A
S37	C11	N	L8	N
S38	C49	D	NoS	N/A
S39	C50	G	L29	G
S39	C50	G	L30	G
S40	C51	G	L31	G
S41	C52	R	L32	R
S42	C53	D	NoS	N/A
S42	C17	N	NoS	N/A
S43	C2	G	NoS	N/A
S44	C11	N	L33	N
S46	C54	N	NoS	N/A
S47	C55	N	NoS	N/A

Appendix D: mapping of challenges with corresponding solutions about AI for D&I (RQ2)

(NoS=No solution, N/A=Not applicable, N=None, G=Gender, S=Sex, A=Age, R=Race, E=Ethnicity, D=Disability, K=Skin tone, L=Geographic location)

Paper ID	Challenge ID	Attributes of challenges	Solution ID	Attributes of solutions
S3	H1	R, E	N1	N
S4	H2	N	N2	N
S4	H2	N	N3	N
S4	H2	N	N4	N
S6	H3	A, G, R	NoS	N/A
S8	H4	N	N5	N
S11	H5	G	NoS	N/A
S11	H6	G	N6	G
S11	H7	N	N7	N
S11	H7	N	N9	N/A
S11	H8	G	N8	G
S11	H8	G	N7	N
S11	H8	G	N9	G
S12	H9	D	N10	N
S13	H6	G	NoS	N/A
S15	H10	N	N11	N
S18	H11	K	N12	K
S19	H12	N	N13	N
S19	H12	N	N14	N
S20	H13	N	NoS	N/A
S21	H14	L, G, A	N15	N
S22	H15	N	N16	N
S27	H16	G	N17	G
S28	H17	N	N18	N
S32	H18	U	N19	U
S38	H19	D	NoS	N/A
S38	H20	D	N20	D
S38	H21	D	N20	D
S42	H22	D	N21	D
S47	H23	N	N22	N
S48	H24	N	N23	N

Appendix E: challenges and solutions with corresponding D&I pillars for D&I in AI (RQ1) and AI for D&I (RQ2)

(H=Humans, D=Data, P=Process, S=System, G=Governance, O=Other)

D &I in AI (RQ1)				AI for D &I (RQ2)			
Challenge ID	Pillar	Solution ID	Pillar	Challenge ID	Pillar	Solution ID	Pillar
C1	H, G	L1	H, G	H1	H, S	N1	P, S
C2	H	L2	H, G	H2	S, G	N2	H, S, P
C3	H, G	L3	H, G	H3	H, D, P, S	N3	P, S, G
C4	H, G	L4	D, P, S	H4	H, G	N4	D, P, S
C5	H, G	L5	P, S	H5	H, D, P, S	N5	H, D, S, G
C6	H	L6	D	H6	H, D, S	N6	H, P, S, G
C7	D	L7	D	H7	D	N7	D, P, S
C8	H, G	L8	H, S, P	H8	H, D	N8	D, P, G
C9	P, S	L9	S, G	H9	H, D, P	N9	D, S
C10	H, D	L10	H, D, P, G	H10	H, P, S, G	N10	H, D, P, S
C11	H, P	L11	D	H11	H, S, P	N11	H, D, P, S, G
C12	D	L12	D, P, S, G	H12	S, G	N12	D, P, S
C13	H, P	L13	H, D, P, G	H13	H, G	N13	D, P, S, G
C14	H, D, P, S	L14	P, S	H14	H, G	N14	D, P, S, G
C15	H, P, G	L15	D, G	H15	P, S	N15	D, P, S
C16	H, P, G	L16	D, P	H16	D, P, S	N16	P, S
C17	H, D, P, S	L17	D, P	H17	P, S	N17	H, D, P, S
C18	H, D, P, S	L18	H, D, P	H18	H, S, P	N18	P, S
C19	D	L19	H, S, P	H19	D, P, S	N19	H, S, P
C20	H, D, G	L20	P, S	H20	H, S, G	N20	P, S, G
C21	S	L21	H, G	H21	H, S, G	N21	S, G

D &I in AI (RQ1)				AI for D &I (RQ2)			
Challenge ID	Pillar	Solution ID	Pillar	Challenge ID	Pillar	Solution ID	Pillar
C22	D, P, S	L22	H, G	H22	H, S, G	N22	P, S
C23	H	L23	H, P, G	H23	D, P	N23	H, D, S
C24	H, G	L24	H, S, G	H24	H, S, P	–	–
C25	D, P, S	L25	H, D, P, S	–	–	–	–
C26	H, P, G	L26	H, D, P, S, G	–	–	–	–
C27	H, P	L27	H, D	–	–	–	–
C28	H, D	L28	P, S	–	–	-	–
C29	D	L29	H, D, P	–	–	–	–
C30	H, S, G	L30	H, P	–	–	–	–
C31	D, P	L31	H, D, P, G	–	–	–	–
C32	H, D, S	L32	D, P	–	–	–	–
C33	D, S	L33	P, S	–	–	–	–
C34	P, S	–	–	–	–	–	–
C35	H, P	–	–	–	–	–	–
C36	P, S, G	–	–	–	–	–	–
C37	H	–	–	–	–	–	–
C38	H, P, S	–	–	–	–	–	–
C39	D	–	–	–	–	–	–
C40	D, P, S	–	–	–	–	–	–
C41	H, G	–	–	–	–	–	–
C42	H, D, P, S	–	–	–	–	–	–
C43	D, P, S	–	–	–	–	–	–
C44	H, D, P	–	–	–	–	–	–
C45	D	–	–	–	–	–	–
C46	H, S, P	–	–	–	–	–	–
C47	H	–	–	–	–	–	–
C48	H, D, S	–	–	–	–	–	–
C49	P, S, G	–	–	–	–	–	–
C50	P, S	–	–	–	–	–	–
C51	H, P, G	–	–	–	–	–	–
C52	D, P, S	–	–	–	–	–	–
C53	H, D, S	–	–	–	–	–	–
C54	H, D, S	–	–	–	–	–	–
C55	P, S	–	–	–	–	–	–

Funding Open access funding provided by CSIRO Library Services. No funding was received to assist with the preparation of this manuscript. No funding was received for conducting this study. No funds or grants was received.

Declarations

Conflict of interest All the authors are full-time paid employee of Australian National Science Agency (CSIRO's Data61). The authors have no competing interests to declare that are relevant to the content of this article. The authors have no financial or proprietary interests in any material discussed in this article. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Bellamy, R.K., Dey, K., Hind, M., Hoffman, S.C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., et al.: Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943 (2018)

2. Dattner, B., Chamorro-Premuzic, T., Buchband, R., Schettler, L.: The legal and ethical implications of using ai in hiring. Harvard Business Review **25** (2019)
3. Schmidt, P., Biessmann, F., Teubner, T.: Transparency and trust in artificial intelligence systems. Journal of Decision Systems **29**(4), 260–278 (2020)
4. Eschenbach, W.J.: Transparency and the black box problem: Why we do not trust ai. Philosophy & Technology **34**(4), 1607–1622 (2021)
5. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., Vertesi, J.: Fairness and abstraction in sociotechnical systems. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59–68 (2019)
6. Zowghi, D., Rimini, F.: Diversity and inclusion in artificial intelligence. arXiv preprint arXiv:2305.12728 (2023)
7. Srinivasan, R., Chander, A.: Biases in ai systems. Communications of the ACM **64**(8), 44–49 (2021)
8. Saheb, T.: Ethically contentious aspects of artificial intelligence surveillance: a social science perspective. AI and Ethics **3**(2), 369–379 (2023)
9. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR) **54**(6), 1–35 (2021)
10. Bertrand, A., Belloum, R., Eagan, J.R., Maxwell, W.: How cognitive biases affect xai-assisted decision-making: A systematic review. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, pp. 78–91 (2022)
11. Xivuri, K., Twinomurinzi, H.: A systematic review of fairness in artificial intelligence algorithms. In: Responsible AI and Analytics for an Ethical and Inclusive Digitized Society: 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2021, Galway, Ireland, September 1–3, 2021, Proceedings 20, pp. 271–284 (2021). Springer
12. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science **366**(6464), 447–453 (2019)
13. Benthall, S., Haynes, B.D.: Racial categories in machine learning. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 289–298 (2019)
14. Pereira, V., Hadjielias, E., Christofi, M., Vrontis, D.: A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. Human Resource Management Review **33**(1), 100857 (2023)
15. Avellan, T., Sharma, S., Turunen, M.: Ai for all: defining the what, why, and how of inclusive ai. In: Proceedings of the 23rd International Conference on Academic Mindtrek, pp. 142–144 (2020)
16. Stoyanovich, J., Howe, B.: Follow the data! algorithmic transparency starts with data transparency. Politics and Public Policy, Harvard Kennedy School, Shorenstein Center on Media (2018)
17. Shrestha, S., Das, S.: Exploring gender biases in ml and ai academic research through systematic literature review. Frontiers in artificial intelligence **5**, 976838 (2022)
18. Varsha, P.: How can we manage biases in artificial intelligence systems-a systematic literature review. International Journal of Information Management Data Insights **3**(1), 100165 (2023)
19. Richardson, B., Gilbert, J.E.: A framework for fairness: A systematic review of existing fair ai solutions. arXiv preprint arXiv:2112.05700 (2021)
20. Laato, S., Tiainen, M., Najmul Islam, A., Mäntymäki, M.: How to explain ai systems to end users: a systematic literature review and research agenda. Internet Research **32**(7), 1–31 (2022)
21. Anjomshoae, S., Najjar, A., Calvaresi, D., Främling, K.: Explainable agents and robots: Results from a systematic literature review. In: 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019, pp. 1078–1088 (2019). International Foundation for Autonomous Agents and Multiagent Systems
22. Fosch-Villaronga, E., Poulsen, A.: Diversity and inclusion in artificial intelligence. Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice, 109–134 (2022)
23. Nadeem, A., Marjanovic, O., Abedin, B., et al.: Gender bias in ai-based decision-making systems: a systematic literature review. Australasian Journal of Information Systems **26** (2022)
24. Mathis, M.S., Badewa, T.E., Obiarinze, R.N., Wilkinson, L.T., Martin, C.A.: A novel use of artificial intelligence to examine diversity and hospital performance. Journal of Surgical Research **260**, 377–382 (2021)
25. Chauhan, P.S., Kshetri, N.: The role of data and artificial intelligence in driving diversity, equity, and inclusion. Computer **55**(4), 88–93 (2022)
26. Buyl, M., Cociancig, C., Frattone, C., Roekens, N.: Tackling algorithmic disability discrimination in the hiring process: An ethical, legal and technical analysis. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 1071–1082 (2022)
27. Borgs, C., Chayes, J., Haghtalab, N., Kalai, A.T., Vitercik, E.: Algorithmic greenlining: An approach to increase diversity. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 69–76 (2019)
28. Cachat-Rosset, G., Klarsfeld, A.: Diversity, equity, and inclusion in artificial intelligence: An evaluation of guidelines. Applied Artificial Intelligence **37**(1), 2176618 (2023)
29. Robert, L.P., Pierce, C., Marquis, L., Kim, S., Alahmad, R.: Designing fair ai for managing employees in organizations: a review, critique, and design agenda. Human-Computer Interaction **35**(5–6), 545–575 (2020)
30. Roche, C., Wall, P., Lewis, D.: Ethics and diversity in artificial intelligence policies, strategies and initiatives. AI and Ethics, 1–21 (2022)
31. Nyariro, M., Emami, E., Abbasgholizadeh Rahimi, S.: Integrating equity, diversity, and inclusion throughout the lifecycle of artificial intelligence in health. In: 13th Augmented Human International Conference, pp. 1–4 (2022)
32. Mhlambi, S., Tiribelli, S.: Decolonizing ai ethics: Relational autonomy as a means to counter ai harms. Topoi, 1–14 (2023)
33. Ormond, E.: Governance of ai ethics: Perspective from the global south (africa). Available at SSRN 4368020 (2023)
34. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering (2007)
35. Jalali, S., Wohlin, C.: Systematic literature studies: database searches vs. backward snowballing. In: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, pp. 29–38 (2012)
36. Archibald, M.M.: Investigator triangulation: A collaborative strategy with potential for mixed methods research. Journal of mixed methods research **10**(3), 228–250 (2016)
37. Liu, Y., Lu, Q., Zhu, L., Paik, H.-Y., Staples, M.: A systematic literature review on blockchain governance. Journal of Systems and Software, 111576 (2022)
38. Walker, P.L., Cook, D.C.: Brief communication: Gender and sex: Vive la difference. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists **106**(2), 255–259 (1998)
39. Roopaei, M., Horst, J., Klaas, E., Foster, G., Salmon-Stephens, T.J., Grunow, J.: Women in ai: Barriers and solutions. In: 2021 IEEE World AI IoT Congress (AIIoT), pp. 0497–0503 (2021). IEEE
40. Scheuerman, M.K., Paul, J.M., Brubaker, J.R.: How computers see gender: An evaluation of gender classification in commercial facial analysis services. Proceedings of the ACM on Human-Computer Interaction 3(CSCW), 1–33 (2019)

41. Australian human rights commission: A quick guide to australian discrimination laws (2014)
42. Umeh, C., Cornelius, N., Wallace, J.: Exploring equality, diversity, and inclusion in multiethnic settings: A context-sensitive approach. Human Resource Management Journal **33**(1), 148–169 (2023)
43. Brattberg, E., Rugova, V., Csernatoni, R.: Europe and AI: Leading, Lagging Behind, or Carving Its Own Way? vol. 9. Carnegie endowment for international peace Washington, DC, USA, ??? (2020)
44. Hamidi, F., Scheuerman, M.K., Branham, S.M.: Gender recognition or gender reductionism? the social implications of embedded gender recognition systems. In: Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems, pp. 1–13 (2018)
45. Huang, H.-Y., Liem, C.C.: Social inclusion in curated contexts: Insights from museum practices. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 300–309 (2022)
46. Jora, R.B., Sodhi, K.K., Mittal, P., Saxena, P.: Role of artificial intelligence (ai) in meeting diversity, equality and inclusion (dei) goals. In: 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 1687–1690 (2022). IEEE
47. Dankwa-Mullan, I., Weeraratne, D.: Artificial intelligence and machine learning technologies in cancer care: Addressing disparities, bias, and data diversity. Cancer Discovery **12**(6), 1423–1427 (2022)
48. Clark, C.R., Wilkins, C.H., Rodriguez, J.A., Preininger, A.M., Harris, J., DesAutels, S., Karunakaram, H., Rhee, K., Bates, D.W., Dankwa-Mullan, I.: Health care equity in the use of advanced analytics and artificial intelligence technologies in primary care. Journal of General Internal Medicine **36**, 3188–3193 (2021)
49. Li, F., Dong, H., Liu, L.: Using ai to enable design for diversity: A perspective. In: Advances in Industrial Design: Proceedings of the AHFE 2020 Virtual Conferences on Design for Inclusion, Affective and Pleasurable Design, Interdisciplinary Practice in Industrial Design, Kansei Engineering, and Human Factors for Apparel and Textile Engineering, July 16–20, 2020, USA, pp. 77–84 (2020). Springer