# A Systematic Review of Fairness in Artificial Intelligence Algorithms

Khensani Xivuri, Hossana Twinomurinzi

▶ **To cite this version:**

Khensani Xivuri, Hossana Twinomurinzi. A Systematic Review of Fairness in Artificial Intelligence Algorithms. 20th Conference on e-Business, e-Services and e-Society (I3E), Sep 2021, Galway, Ireland. pp.271-284, 10.1007/978-3-030-85447-8_24 . hal-03648158

**HAL Id: hal-03648158**

**https://inria.hal.science/hal-03648158**

Submitted on 21 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License


This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# A Systematic Review of Fairness in Artificial Intelligence algorithms

Khensani Xivuri[1[0000-0002-1041-3648]] and Hossana Twinomurinzi[2[0000-0002-9811-3358]]

[1] University of Johannesburg, Auckland Park, Johannesburg
KhensaniX@gmail.com,Hossanat@uj.ac.za

**Abstract.** Despite being the fastest-growing field because of its ability to enhance competitive advantage, there are concerns about the inherent *fairness* in Artificial Intelligence (AI) algorithms. In this study, a systematic review was performed on AI and the fairness of AI algorithms. 47 articles were reviewed for their focus, method of research, sectors, practices, and location. The key findings, summarized in a table, suggest that there is a lack of formalised AI terminology and definitions which subsequently results in contrasting views of AI algorithmic fairness. Most of the research is conceptual and focused on the technical aspects of narrow AI, compared to general AI or super AI. The public services sector is the target of most research, particularly criminal justice and immigration, followed by the health sector. AI algorithmic fairness is currently more focused on the technical and social/human aspects compared to the economic aspects. There was very little research from Asia, Middle East, Oceania, and Africa. The study makes suggestions for further research.

**Keywords:** AI · Machine Learning · Algorithms · Fairness · Bias · Ethics

## 1 Introduction

Artificial Intelligence (AI) has rapidly been gaining momentum in different industries and is now a part of daily life [1]. AI enables systems to perform tasks that would normally be performed by humans [2] in three different categories; narrow AI, general AI and super AI [3]. Narrow AI performs operational tasks through the use of machine learning tools such as recognising individual faces, driving a car or speech recognition [3, 4]. General AI is designed to be as intelligent as humans with the ability to perform any intelligent tasks [3], but remains computationally complex [5]. General AI solves complex problems and independently controls itself. General AI has the ability to get knowledge, apply it, reason, and think. Super AI is the type that is more intelligent than humans and would do better than humans in almost everything including intelligence and social skills [3]. Super AI has not been developed yet and its implementation although being utopic is feared it could have negative consequences such as human extinction [4].

AI applications collect and process data, and provide results that mimic human intelligence using rules learnt over time. Most AI applications have three common components; input data, a machine learning algorithm that processes the input data, and the output decision which is based on the machine learning process [6]. Machine learning

makes use of trained data and test data for algorithmic models [7]. Algorithms are the coded procedures trained on existing data that transforms input data into expected results [8]. Another type of AI is expert systems which are rule-based and do not use machine learning algorithms to process data [9].

Some of the popular AIs are computer vision, natural language processing, and artificial neural networks. Some well-known AI solutions include Apple's SIRI, Google Maps, Google predictions, Smart replies by Gmail [10].

As the use of AI grows in different industries, organisations also need to consider the ethics and morals relating to the decisions from AI [11]. AI is prone to algorithmic unfairness, that is, making judgmental errors and incorrect assessments based on biased code or data, resulting in operational and reputational damage [12]. AI systems have the potential to be unfair to certain groups of people, especially with regards to racial discrimination [11], yet it is often difficult to prove if the algorithms are being unfair without access to the data and algorithms [12].

The ethics of AI is a rapidly emerging field of ethics concerned with the design, development, and implementation of AI systems [13]. The ethics of AI is important in maintaining trust in AI and ensuring that bias is removed. One of the main principles of AI ethics is fairness, which requires AI systems to be fair in terms of respecting the law, human rights and democratic values and principles. Fairness requires that AI is built in a manner that promotes democratic values and principles such as freedom and equality [14]. Fairness as a behavioral quality also means impartiality in decision-making [15 – 16], that is treatment without any self-interest or prejudice in either the outcomes or the process leading to the outcomes. Fairness further extends beyond equality and perceptions and digs into the underlying reasons.

The objective of this study was therefore to conduct a systematic review of AI algorithmic fairness. The paper sought to identify the existing gaps, challenges, and opportunities for future research on AI algorithmic fairness.

## 2 Methodology: search procedures, coding, and classification

The study adopted Okoli and Schabram's guide [17] to conducting a systematic literature review (SLR). Systematic literature reviews are important for identifying and evaluating the existing body of work and knowledge produced by scholars, researchers and practitioners [17]. This kind of literature review is used to identify gaps and challenges for future research [18]. The SLR includes details on the literature search, screening for inclusion and exclusion, data extraction and analysis of the results. The data for this research was collected between April and September 2020. Refer to Figure 1 for all search results. The results were analyzed against the inclusion, exclusion and quality assessment criteria as detailed in table 1 below [19].

**Table 1.** Inclusion, Exclusion & Quality assessment Criteria

| Inclusion criteria | Exclusion criteria | Quality assessment criteria |
|---|---|---|

| | | |
|---|---|---|
| Full research Articles including AI and fairness | Research performed before 2015 (to ensure that the research is current). | Verifying that the paper was based on research |
| Reputable news | | Verify that an adequate description of the research context was described |
| All industries | Research that included AI but not Fairness | Verify that the methodology was described |
| All locations | | Verify that there is a clear statement of findings |
| Articles written in English | Research that included Fairness but not AI | Verify that the paper described AI |

### 2.1 Screening for inclusion and exclusion

The Boolean operators "AND" and "OR" were used in the search to combine all search elements for articles relating to both Artificial Intelligence and Fairness. The search string that was used on all the search engines was: ((AI OR Artificial Intelligence OR Algorithm OR Algorithmic) AND (Fairness OR Bias OR Discrimination OR Ethics)). The initial search resulted in 295 311 possible articles as shown in figure 1 below. The research results were reviewed based on the inclusion, exclusion and quality criteria assessment defined in table 1. Only 47 articles met all required conditions. The 47 articles were read in detail to identify the current gaps, challenges, and opportunities in algorithmic fairness. The process is presented through the PRISMA flowchart [20] in figure 1. The classification and coding of the articles selected were defined as per section 2.2. Section 2.2 also details the classification framework used for the analysis.

The analysis of the results was based on the classification framework in section 2.2. A correlation analysis using Pearson's coefficient was additionally used to check for any linear relationships between the different classifications and their relationship strength [21].

### 2.2 Classification Framework

A classification system was developed based on Amui et al [18] that includes focus, method of the research, sector, practices or dimensions and origin. The framework follows the following procedures:

- Conduct a survey of available articles published on the fairness of AI algorithms;
- Develop and use a structured classification coding system to provide a structure on the existing knowledge on the fairness of AI algorithms;
- Identify main results of the articles based on the developed coding system; and
- Analyse the identified gaps, opportunities, and challenges for future studies.

The classification codes are based on the following:

- Focus (1), coded on a scale of A to D (i.e. whether the articles focused on General AI, Super AI, Narrow AI or no categorisation), based on the work of [22] and [23].
- Research method (2), coded as A to H (i.e. Quantitative, Qualitative, Conceptual research methodologies etc.,) based on the work of [24].
- Sector analysis (3), coded as A to T (i.e. Agriculture, Basic Metal production, Chemical industries, commerce etc.,) based on the work of [22]. The sectors were taken from [25].

- Practices or dimensions used in the research (4) coded as A to D (i.e. technical aspects, social/human aspects, or economic aspects) based on the work of [26].
- Origin of the research done, coded as A to G, which includes the different continents, based on the work of [27].

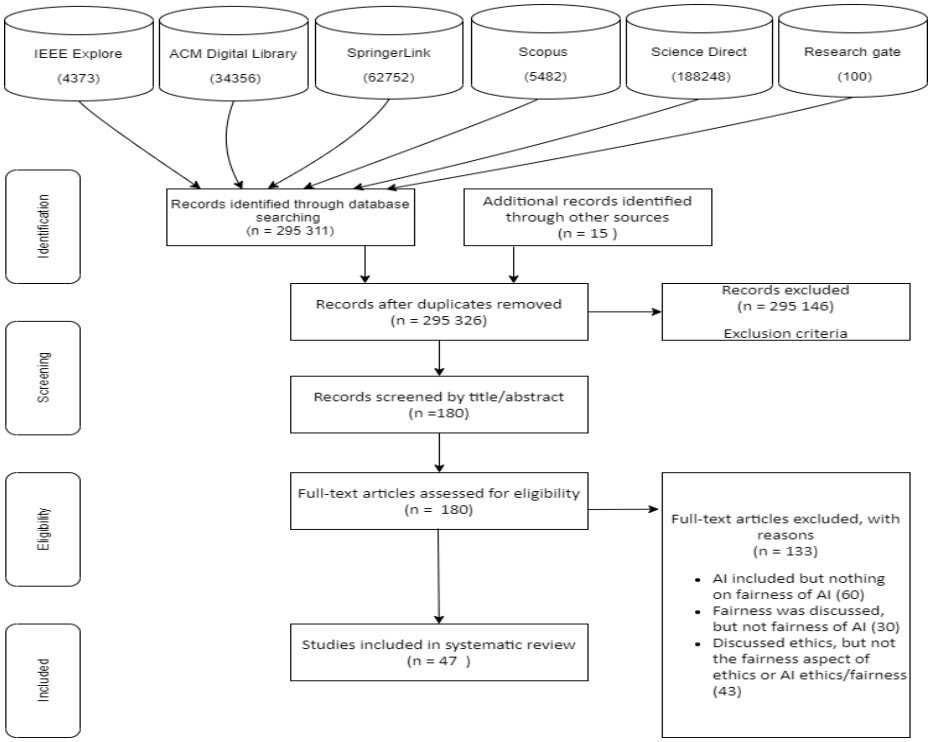


**Fig. 1.** Screening for Inclusion & Exclusion

Due to space limitations, all classifications and codes used are given in Annexure 1. A brief description of the objectives and results of each of the articles is given in Annexure 2. Annexure 3 presents the correlation table and Annexure 4 has a summary of the 47 articles reviewed. This is presented in https://dx.doi.org/10.13140/RG.2.2.26883.43048.

## 3 Analysis and Discussion of Findings

### 3.1 Focus

Algorithmic fairness was discussed using AI, machine learning and algorithms interchangeably in almost all the research papers reviewed. In as much as machine learning has been around since the 1950s, the terminology used by different entities still differs. For example, Beil et al., [28] use both AI and machine learning together, throughout their research paper, whilst Mujtaba and Mahapatra [29] use AI and machine learning

interchangeably, and Dias and Torkamani [30] use AI, machine learning and algorithms separately with a clear differentiation between all three terms. This can be expected as there is a lack of formalised AI terminology and definitions. The lack of formalised AI definitions also creates differences in algorithmic fairness terminology, resulting in the lack of mutual understanding around algorithmic fairness terminology. Many researchers have come up with different technical definitions of machine learning fairness, however, there is no standardised definition of crucial AI terminology [31]. Standards clearly defining AI terminology should be developed to ensure that algorithmic fairness is universally understood and adopted.

57% of the articles focused on fairness in Narrow AI (1C), 38% were non categorised and did not address any specific type of AI (1D), whilst 2% focused on the fairness of both General AI and Narrow AI (1A, 1C). The other 2% considered the fairness of all three types of AI (1A, 1B, 1C) together. The above results are presented in Annexure 5 – Figure 1 found on https://dx.doi.org/ 10.13140/RG.2.2.26883.43048. A correlation analysis revealed that fairness of Narrow AI (1C) was mainly focused on the technical areas (4A) (0.52517), while non-categorised AI (1D) did not focus on the technical areas (4A) (-0.55293).

The findings indicate that a great deal of research on the fairness of AI algorithms is focused on the technical areas of Narrow AI. This might be expected as it is in narrow AI that one will find the algorithmic implementation. Ethical challenges may arise due to the technical characteristics of AI and machine learning and the limited skills in narrow AI [28]. There are technical approaches, fixes, efforts, tools, and solutions that can be used to mitigate bias in AI [32] [33]. Additionally, there are gaps in the current available AI ethical guidelines, including technical detail and explanations [33].

The lower count of research on the fairness of General AI and Super AI indicates both the complexity of the latter two areas and the limited number of implementations. Although Super and General AI are only expected in the future, the lack of research in the two areas could lead to implications more extreme than current implications noted from lack of fairness in Narrow AI.

There is therefore a research opportunity to address the fairness of General AI and Super AI. In as much as General AI currently does not exist, it is important to first understand the difference between Narrow AI and General AI before defining AI, machine learning and their implications [34]. Some risks can be expected from General AI and Super AI automated decision making [35].

### 3.2 Method of the research

Most of the articles used the conceptual research methodology (2G – 45 %), and the qualitative and theoretical research methodologies together (2A, 2C – 17%) with a correlation of 0.359. There is little research using qualitative (2B – 6%), survey (2F – 4%), theoretical (2%), quantitative (2A – 2%), empirical (2D – 2%), and case study/interview (2E – 2%) research methodologies in the study of algorithmic fairness. The above results are presented in Annexure 5 – Figure 2 found on https://dx.doi.org/

10.13140/RG.2.2.26883.43048. Research originating from America (5A) used the theoretical research methodology (0.326) more, compared to Europe which used less of the theoretical research methodology (-0.359), the empirical research methodology (-0.293).

The dominance of the theoretical research methodology could be because the fairness of AI is an emerging phenomenon, and more research is required for example, around policies, legal and consumer protection [36].

There is an opportunity for more research to be conducted using a mixture of different research methodologies across the different regions/continents and sectors. There is a need for theoretical and empirical research around justifications of AI decision making [37]. Performing empirical and philosophical research around this area will aid in getting a more thorough public and individual perspective on AI decision making [37]. There is a need for empirical research that is more focused and rigorous on important questions regarding AI, its adoption, and the consequences thereafter [34].

### 3.3 Sectors

Most of the research was not specific to any sector (3U – 62%). 21% of the articles were focused on the public services sector (3Q), mainly in the criminal justice, immigration, and government. 11% of the articles were from the health sector (3J), 2% from the financial sector (3G) and 2% from the Postal and telecommunications sector (3P). The other 2% considered the fairness of AI in both the health and financial sectors (3G, 3J). The above results are presented in Annexure 5 – Figure 3 found on https://dx.doi.org/ 10.13140/RG.2.2.26883.43048.

The findings indicate that most of the research done was generalised and not specific to any sector. This could be because sector-specific information relating to the fairness of AI is not available or easily attainable. There are currently not enough public data sets and information around AI to allow for a detailed microanalysis of differences in the different sectors and regions [34].

A smaller percentage of the research focused on the criminal justice sector, the financial sector, and the health sector. This could be because more AI benefits have been realised in these sectors compared to the other sectors. The banking and financial sectors are currently at the forefront of AI [38]. AI has brought in a lot of benefits in industries such as health care, transportation, criminal justice, and economic inclusion [39]. This could also be because biases in these sectors could have a larger impact compared to the other sectors. An example relating to the health sector is an algorithm that is used widely in the US health care, affecting millions of people, which was found to be highly biased in that it gave white people greater health care over black people who needed the treatment more as their health conditions were worse [40]. This type of bias could result in consequences such as loss of lives as well as fines from the health regulatory boards. The use of AI in healthcare raises critical ethical issues that are important in avoiding harming patients, liability for healthcare providers and undermining public trust in AI. Another example relating to the finance sector is the use of biases such as gender and race to provide a credit score or determine an individual's credit eligibility, which is illegal in the US [41]. Bias in the financial sector could result in punishment

by courts and fines where there are laws against it. Lastly, an example relating to the criminal justice sector is the recent outrage from the Black Lives Matters movement in the US which argues that the criminal justice sector is biased against black people [42]. The criminal justice sector uses criminal risk assessments algorithms to determine a criminal/defendant's future risk for misconduct [42]. Bias in such algorithms could result in incorrect conviction or sentencing.

There is therefore an opportunity for sector-specific research on the fairness of AI. A sector-specific AI fairness approach is required to oversee, audit and monitor AI technologies in the different sectors [43]. A sector-specific approach will ensure that the sector focuses more on the application of AI and its impacts rather than prioritising the technology. Different sectors have different characteristics and required expertise, and therefore the governance or regulations of one sector may not be appropriate for another sector. Different expertise and knowledge specific to each sector will be required for good AI governance and ensuring the fairness of AI in the different sectors.

### 3.4 Practices or dimensions used in the research

72% of the papers focused on both the technical aspects and the social/human aspects of algorithmic fairness (72 % - 4A, 4B). 19 % of the articles focused on social/human aspects (4B), and 6% on all 3 aspects together (4A, 4B, 4C). The other 2 % considered the economic aspects (4C) of algorithmic fairness. The above results are presented in Annexure 5 Figure 4 found on https://dx.doi.org/ 10.13140/RG.2.2.26883.43048.

The findings indicate that there is not enough research on the economic aspects of algorithmic fairness. This could be a result of the lack of transparency in the AI algorithms used in the financial sector. Gender and race are still used to determine whether a loan should be granted to individuals and businesses in South Africa [44]. Algorithms used for loan decisions are trained using loan history and demographic data of applicants who have been accepted or rejected, which increases the chances of black women being rejected or given high-interest rates for loan applications, resulting in continued financial injustices in the country which may further affect the economy. AI biases have also resulted in black people's loan applications being rejected in the US [44]. Banking services have been investing a lot in AI, however, conversations on the biases of AI used in banking are very limited [44].

There is therefore an opportunity for research to be done on the economic aspects of algorithmic fairness. There is a need for evidence-based research which will provide more detail on how AI will affect economic outcomes [44]. The banking sector was one of the biggest sectors which invested in AI in 2019 globally [44]. The increase in AI investments in this sector calls for more research on the fairness of algorithms in the finance/banking sector and the economic aspects of it.

### 3.5 Origin

Most of the research on algorithmic fairness originated from Europe (5B - 49%), America (5A – 21%) and both Europe and America (5A, 5B - 13%). 9% of the articles orig-

inated from Africa (5D), 6% from Asia (5C), 4% from a combination of America, Europe, and Asia (5A, 5B, 5C), and 2% from a combination of America and Asia (5A, 5C). The other 2% were either global or not specific to any region (5G). The above results are presented in Annexure 5 – Figure 5 found on https://dx.doi.org/10.13140/RG.2.2.26883.43048. An analysis of research originating from Africa revealed that algorithmic fairness in Africa did not focus on the technical aspects of algorithmic fairness. Algorithmic fairness research originating from Africa focused more on uncategorised AI compared to Narrow AI and was done mainly using the conceptual research methodology.

The findings indicate that there is not enough research done on algorithmic fairness in Asia, Middle East, Oceania, and Africa. For example, countries that have issued ethical AI guidelines are economically developed countries such as the USA and the UK, followed by Japan, Germany, France and Finland [45]. There is a lack of AI guidelines originating from Africa, South and Central America, and Asia, which shows that regions are not equally participating in the development of AI ethics.

The findings also illustrate that research done in Africa did not focus on the technical aspects of algorithmic fairness. Research originating from Africa focused on the fairness of generalised AI and not specific AI types. This could be because AI is still an emerging technology in Africa despite it being highly implemented. In as much as there's a lot of researchers, engineers and technology professionals who are ready to explore AI in Africa, AI is still a new concept in Africa [46]. There is a need for expanding AI expertise and building AI solutions in Africa rather than just focusing on the theoretical aspects of it [47].

There is therefore an opportunity for research to be done on algorithmic fairness globally. There is also an opportunity to do research focusing on the fairness of the different types of AI (Narrow AI, Super AI, and General AI) in Africa. For example, it is important for all countries, regardless of their economic conditions, to be fully involved in the development of AI ethics [45]. The involvement of all countries will help avoid neglecting local knowledge, cultural diversity, and the need for global fairness.

## 4 Discussion

Table 2 presents a summary of key gaps, opportunities, and future research.

**Table 2.** Discussion and Conclusion

| Category | Gaps | Opportunities/Further research |
|---|---|---|
| Focus | Lack of formalised AI terminology and definitions | Standards clearly defining AI terminology should be developed to ensure that algorithmic fairness is universally understood and adopted. |
| | Lack of research on the fairness of General and Super AI | There is a research opportunity to address the fairness of General AI and Super AI. |

| | | |
|---|---|---|
| | Lack of Social/human aspects and economic aspects of Narrow AI | There is a research opportunity to address the social/human aspects of Narrow AI and the economic aspects of AI. |
| Research Methodology | Lack of research using a mixture of different research methodologies across the different regions and sectors. Most of the research largely used the conceptual research methodology. | There is an opportunity to perform research using a mixture of different research methodologies across the different sectors and regions. |
| Sectors | Lack of sector-specific research | There is a research opportunity for sector-specific research on the fairness of AI. |
| | Research that was sector-specific was focused on the criminal justice sector, the financial sector, and the health sector | There is a research opportunity to address the fairness of AI in all sectors including the impact of bias in the different sectors. |
| Dimensions | Lack of research on the economic aspects of AI fairness | There is a research opportunity to address the economic aspects of AI fairness |
| Origin | Lack of research on algorithmic fairness in Asia, the Middle East, Oceania, and Africa | There is a research opportunity to address algorithmic fairness globally. |
| | Lack of research on the technical aspects of AI fairness originating from Africa | There is a research opportunity to address the technical aspects of algorithmic fairness in Africa. |

## 5 Conclusions

This study performed a systematic literature review on the fairness of AI algorithms. This type of review is important in structuring available knowledge in a subject area, and the planning of future studies. The results of the study indicate the absence of formalised AI terminology and definitions. Most of the research focused on the fairness of Narrow AI, in no specific sector, in America and Europe, largely using the conceptual research methodology. Less research is available on the economic aspects of algorithmic fairness globally, and the technical aspects of algorithmic fairness in Africa. There is therefore a gap in AI terminology, the algorithmic fairness of Super & General AI, sector-specific algorithmic fairness, and the economic aspects of algorithmic fairness. Standards clearly defining AI terminology should be developed to ensure that algorithmic fairness is universally understood and adopted. Research addressing the technical aspects of algorithmic fairness in Africa should be done.

This research provides a significant implication for research theory and practice. The findings indicate that there is less research on algorithmic fairness in low-income countries. There are opportunities to develop sector-specific theory in the field of algorithmic fairness of AI, including the development of formalised standards clearly defining AI terminology on a global level. In practice, policymakers for AI implementation

should also look at the algorithmic fairness of AI before roll-out, its implications, and how to avoid bias to ensure success and trust from society.

This research contributes to Information systems governance by highlighting gaps, challenges, and opportunities in AI algorithmic fairness research. Research on the fairness of General and Super AI, focusing on the economic aspects of AI, using a mixture of research methodologies.

## 6 References


1. Scherer, M.U.: Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. Havard J. Law Technol. 29, 354–200 (2016).
2. Ghosh, A., Chakraborty, D., Law, A.: Artificial intelligence in Internet of things. CAAI Trans. Intell. Technol. 3, 208–218 (2018).
3. Gherheş, V.: Why Are We Afraid of Artificial Intelligence (Ai)? Eur. Rev. Appl. Sociol. 11, 6–15 (2019).
4. Gurkaynak, G., Yilmaz, I., Haksever, G.: Stifling artificial intelligence: Human perils. Comput. Law Secur. Rev. 32, 749–758 (2016).
5. Pennachin, C., Goertzel, B.: Contemporary Approaches to Artificial General Intelligence. Cogn. Technol. 8, 1–30 (2007).
6. Salah, K., Rehman, M.H.U., Nizamuddin, N., Al-Fuqaha, A.: Blockchain for AI: Review and open research challenges. IEEE Access. 7, 10127–10149 (2019).
7. Hacker, P.: Teaching fairness to artificial intelligence: Existing and novel strategies againstalgorithmic discrimination under EU law. Common Mark. Law Rev. 55, 1143–1185 (2018).
8. Beretta, E., Santangelo, A., Lepri, B., Vetr, A., De Martin, J.C.: The Invisible Power of Fairness. How Machine Learning Shapes Democracy. Lect. Notes Comput. Sci. 11489, 352–358 (2019).
9. Yigin, I.H., Taşkin, H., Cedİmoglu, I.H., Topal, B.: Supplier selection : an expert system approach Supplier selection : an expert system approach. 7287, (2007).
10. Patel, K.N., Raina, S.: Artificial Intelligence and its Models. 0–3 (2020).
11. Sharma, S., Henderson, J., Ghosh, J.: CERTIFAI: Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of Artificial Intelligence models. (2019).
12. Horowitz, M.C.: Artificial Intelligence, International Competition, and the Balance of Power. Texas Natl. Secur. Rev. 1, 37–57 (2018).
13. Strous, L., Johnson, R., Grier, D.A., Swade, D.: Unimagined Futures – ICT Opportunities and Challenges. Springer Nature Switzerland AG, Switzerland (2020).
14. Ienca, M.: Democratizing cognitive technology : a proactive approach. Ethics Inf. Technol. 21, 267–280 (2018).
15. Farnadi, G., Babaki, B., Getoor, L.: Fairness in Relational Domains. AIES 2018 - Proc. 2018 AAAI/ACM Conf. AI, Ethics, Soc. 108–114 (2018).
16. Neuteleers, S., Mulder, M., Hindriks, F.: Assessing fairness of dynamic grid tariffs. Energy Policy. 108, 111–120 (2017).
17. Okoli, C., Schabram, K.: A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts Work. Pap. Inf. Syst. 10, (2010).
18. Amui, L.B.L., Jabbour, C.J.C., de Sousa Jabbour, A.B.L., Kannan, D.: Sustainability as a dynamic organizational capability: a systematic review and a future agenda toward a sustainable transition. J. Clean. Prod. 142, 308–322 (2017).

19. Kusen, E., Strembeck, M.: A decade of security research in ubiquitous computing: results of a systematic literature review. Int. J. Pervasive Comput. Commun. 12, 216–259 (2016).
20. Harris, J.D., Quatman, C.E., Manring, M.M., Siston, R.A., Flanigan, D.C.: How to write a systematic review. Am. J. Sports Med. 42, 2761–2768 (2014).
21. Chok, N.S.: PEARSON'S VERSUS SPEARMAN'S AND KENDALL'S CORRELATION COEFFICIENTS FOR CONTINUOUS DATA. 9, 76–99 (2010).
22. Jabbour, C.J.C.: Environmental training in organisations: From a literature review to a framework for future research. Resour. Conserv. Recycl. 74, 144–155 (2013).
23. Mariano, E.B., Sobreiro, V.A., Rebelatto, D.A. do N.: Human development and data envelopment analysis: A structured literature review. Omega (United Kingdom). 54, 33–49 (2015).
24. Lage, M., Filho, M.G.: Production planning and control for remanufacturing: Literature review and analysis. Prod. Plan. Control. 23, 419–435 (2012).
25. Organisation International Labour: Industries and Sectors, https://www.ilo.org/global/industries-and-sectors/lang--en/index.htm, last accessed 2020/08/11.
26. Jabbour, C.J.C., Jugend, D., De Sousa Jabbour, A.B.L., Gunasekaran, A., Latan, H.: Green product development and performance of Brazilian firms: Measuring the role of human and technical aspects. J. Clean. Prod. 87, 442–451 (2015).
27. Fahimnia, B., Sarkis, J., Davarzani, H.: Green supply chain management: A review and bibliometric analysis. Int. J. Prod. Econ. 162, 101–114 (2015).
28. Beil, M., Proft, I., van Heerden, D., Sviri, S., van Heerden, P.V.: Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med. Exp. 7, 1–13 (2019).
29. Mujtaba, D.F., Mahapatra, N.R.: Ethical considerations in AI-Based Recruitment. Particip. Pedagog. Impact Res. 109–120 (2019).
30. Dias, R., Torkamani, A.: Artificial intelligence in clinical and genomic diagnostics. Genome Med. 11, 1–12 (2019).
31. Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P.: Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. Philos. Technol. 31, 611–627 (2018).
32. Hagendorff, T.: From privacy to anti-discrimination in times of machine learning. Ethics Inf. Technol. 21, 331–343 (2019).
33. Hagendorff, T.: The Ethics of AI Ethics: An Evaluation of Guidelines. Minds Mach. 30, 99–120 (2020).
34. Raj, M., Seamans, R.: Primer on artificial intelligence and robotics. J. Organ. Des. 8, (2019).
35. Gill, K.S.: AI&Society: editorial volume 35.2: the trappings of AI Agency. AI Soc. 35, 289–296 (2020).
36. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., Floridi, L.: Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Sci. Eng. Ethics. 24, 505–528 (2018).
37. de Fine Licht, K., de Fine Licht, J.: Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. AI Soc. (2020).
38. Soluciones Decide: How Different Sectors are Using AI | by Decide Soluciones | Becoming Human: Artificial Intelligence Magazine, https://becominghuman.ai/how-different-sectors-are-using-ai-26470ba334ab, last accessed 2020/09/03.
39. Bundy, A.: Preparing for the future of Artificial Intelligence. Ai Soc. 32, 285–287 (2017).
40. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science (80-. ). 366, 447–453 (2019).

41. Klein, A.: Reducing bias in AI-based financial services, https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/, last accessed 2020/11/25.
42. Rao, A.: Artificial intelligence poses serious risks in the criminal justice system - The Johns Hopkins News-Letter, https://www.jhunewsletter.com/article/2020/09/artificial-intelligence-poses-serious-risks-in-the-criminal-justice-system, last accessed 2020/11/25.
43. Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J., Schwartz, O.: AI Now report. AI Now Inst. (2018).
44. Moosajee, N.: Fix AI's racist, sexist bias - The Mail & Guardian, https://mg.co.za/article/2019-03-14-fix-ais-racist-sexist-bias/, last accessed 2020/11/27.
45. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. 1, 389–399 (2019).
46. Nwankwo, E., Sonna, B.: Africa's social contract with AI. XRDS Crossroads, ACM Mag. Students. 26, 44–48 (2019).
47. Marwala, T.: Review, amend or create policy and legislation enabling the 4IR - The Mail & Guardian, https://mg.co.za/article/2020-04-03-review-amend-or-create-policy-and-legislation-enabling-the-4ir/, last accessed 2020/08/28.
48. de Abreu, J.C.: The role of Artificial Intelligence in the European e-Justice Paradigm - Suiting effective judicial protection demands. Springer International Publishing (2019).
49. Council of Europe - European commission for the efficiency of justice (CEPEJ): European ethical charter on the use of Artificial Intelligence in judicial systems and their environment, https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.
50. Geis, J.R., Brady, A., Wu, C.C., Spencer, J., Ranschaert, E., Jaremko, J.L., Langer, S.G., Kitts, A.B., Birch, J., Shields, W.F., van den Hoven van Genderen, R., Kotter, E., Gichoya, J.W., Cook, T.S., Morgan, M.B., Tang, A., Safdar, N.M., Kohli, M.: Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. Insights Imaging. 10, (2019).
51. Ishii, K.: Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects. AI Soc. 34, 509–533 (2019).
52. Rullo, A., Serra, E., B, J.L.: Policy-Based Autonomic Data Governance. Springer International Publishing (2019).
53. Choraś, M., Pawlicki, M., Puchalski, D., Kozik, R.: Machine learning – the results are not the only thing that matters! what about security, explainability and fairness? Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 12140 LNCS, 615–628 (2020).
54. Thesmar, D., Sraer, D., Pinheiro, L., Dadson, N., Veliche, R., Greenberg, P.: Combining the Power of Artificial Intelligence with the Richness of Healthcare Claims Data: Opportunities and Challenges. Pharmacoeconomics. 37, 745–752 (2019).
55. Završnik, A.: Criminal justice, artificial intelligence systems, and human rights. ERA Forum. 20, 567–583 (2020).
56. Neri, E., Coppola, F., Miele, V., Bibbolino, C., Grassi, R.: Artificial intelligence: Who is responsible for the diagnosis? Radiol. Medica. 125, 517–521 (2020).
57. Currie, G., Hawk, K.E., Rohren, E.M.: Ethical principles for the application of artificial intelligence (AI) in nuclear medicine. Eur. J. Nucl. Med. Mol. Imaging. 47, 748–752 (2020).
58. D'Agostino, M., Durante, M.: Introduction: the Governance of Algorithms. Philos. Technol. 31, 499–505 (2018).
59. Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to Design AI for Social Good: Seven Essential Factors. Sci. Eng. Ethics. 26, 1771–1796 (2020).

60. Lee, M.S.A., Floridi, L.: Algorithmic Fairness in Mortgage Lending: from Absolute Conditions to Relational Trade-offs. Minds Mach. (2020).
61. Miron, M., Tolan, S., Gómez, E., Castillo, C.: Evaluating causes of algorithmic bias in juvenile criminal recidivism. Springer Netherlands (2020).
62. Wong, P.H.: Democratizing Algorithmic Fairness. Philos. Technol. 33, 225–244 (2020).
63. Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Muller, K.-R.: Towards Explainable Artificial Intelligence. Lect. Notes Comput. Sci. 11700, 435 (2019).
64. Iosifidis, V., Fetahu, B., Ntoutsi, E.: FAE: A Fairness-Aware Ensemble Framework. Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019. 1375–1380 (2019).
65. Parsheera, S.: A gendered perspective on Artificial Intelligence. Mach. Learn. a 5G Futur. (ITU K). 1689–1699 (2018).
66. Altman, M., Wood, A., Vayena, E.: A Harm-Reduction Framework for Algorithmic Fairness. IEEE Secur. Priv. 16, 34–45 (2018).
67. Bellamy, R.K.E., Mojsilovic, A., Nagar, S., Ramamurthy, K.N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K.R., Zhang, Y., Dey, K., Hind, M., Hoffman, S.C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S.: AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM J. Res. Dev. 63, (2019).
68. Oneto, L., Chiappa, S.: Fairness in Machine Learning. Stud. Comput. Intell. 896, 155–196 (2020).
69. Antunes, N., Balby, L., Figueiredo, F., Lourenco, N., Meira, W., Santos, W.: Fairness and Transparency of Machine Learning for Trustworthy Cloud Services. Proc. - 48th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Networks Work. DSN-W 2018. 188–193 (2018).
70. Zhang, W., Tang, X., Wang, J.: On fairness-aware learning for non-discriminative decision-making. IEEE Int. Conf. Data Min. Work. ICDMW. 2019-Novem, 1072–1079 (2019).
71. Binns, R.: What Can Political Philosophy Teach Us About Algorithmic fairness ? 73–80 (2018).
72. Nayebare, M.: Artificial intelligence policies in Africa over the next five years. XRDS Crossroads, ACM Mag. Students. 26, 50–54 (2019).
73. Heaven, W.D.: The UK is dropping an immigration algorithm that critics say is racist | MIT Technology Review, https://www.technologyreview.com/2020/08/05/1006034/the-uk-is-dropping-an-immigration-algorithm-that-critics-say-is-racist/, last accessed 2020/08/28.
74. Marwala, T.: South Africa must have a stake in artificial intelligence technology - The Mail & Guardian, https://mg.co.za/article/2020-03-06-south-africa-must-have-a-stake-in-artificial-intelligence-technology/, last accessed 2020/08/28.
75. Žliobaitė, I.: Measuring discrimination in algorithmic decision making. Data Min. Knowl. Discov. 31, 1060–1089 (2017).
76. Ignatiev, A., Cooper, M.C., Siala, M., Hebrard, E., Marques-silva, J.: Towards Formal Fairness in Machine Learning. LNCS. 846–867 (2020).
77. Feuerriegel, S., Dolata, M., Schwabe, G.: Fair AI. Bus. Inf. Syst. Eng. 62, 379–384 (2020).
78. Kapatamoyo, M., Ramos-Gil, Y.T., Márquez Dominiguez, C.: Algorithmic discrimination and responsibility: Selected examples from the United States of America and South America. Commun. Comput. Inf. Sci. 1051 CCIS, 147–157 (2019).
79. Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., Staab, S.: Bias in Data-driven AI Systems -- An Introductory Survey. (2020).