International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 5
ISSN: 2088-5334

Perspectives of Defining Algorithmic Fairness in Customer-oriented Applications: A Systematic Literature Review

Maw Maw [a], Su-Cheng Haw [a,*], Kok-Why Ng [a]

[a] *Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, Cyberjaya, Selangor, Malaysia*
*Corresponding author: *sucheng@mmu.edu.my*

Abstract*—Automated decision-making systems are massively engaged in different types of businesses, including customer-oriented sectors, and bring countless achievements in persuading customers with more personalized experiences. However, it was observed that the decisions made by the algorithms could bring unfairness to a person or a group of people, according to recent studies. Thus, algorithmic fairness has become a spotlight research area, and defining a concrete version of fairness notions has also become significant research. In existing literature, there are more than 21 definitions of algorithmic fairness. Many studies have shown that each notion has an incompatibility problem, and it is still necessary to make those notions more adaptable to the legal and social principles of the desired sectors. Yet, the constraints of algorithmic fairness for customer-oriented areas have not been thoroughly studied. This motivates us to work on a systematic literature review to investigate the sectors concerned about algorithmic fairness as a significant matter when using machine-based decision-making systems, what are the well-applied algorithmic fairness notions, and why they can or cannot be directly applicable to the customer-oriented sectors, what are the possible algorithmic fairness constraints for the customer-oriented sectors. By applying the standard guidelines of systematic literature review, we explored 65 prominent articles thoroughly. The findings show 43 different ways of algorithmic fairness notions in the varieties of domains. We also identified the three important perspectives to be considered for enhancing algorithmic fairness notions in the customer-oriented sectors.

***Keywords*—Algorithmic fairness; artificial intelligence in business customer-oriented; fairness notions; systematic literature review.**

Manuscript received 5 Dec. 2023; revised 17 Apr. 2024; accepted 23 Aug. 2024. Date of publication 31 Oct. 2024.

I. INTRODUCTION

Due to the increased transformation from traditional decision-making systems, customer relationship managers have to depend on machine learning (ML) based classifiers and predictors to upgrade the financial gains and to attract more customers[1]. Besides the issues in using ML-based models such as lack of necessary data and producing inaccurate results due to the wrong choice of algorithms, another hot issue is the biased outcomes which are unfavored to the protected group of people. As a consequence, it has become a significant concern about algorithmic fairness (AF) to obtain unbiased and fairer predictions or classification outcomes. Thus, it would prevent possible life-changing impacts on certain groups of people in the real world. Existing research studies are devoted to group fairness which investigates whether ML-based automated decisions have an impact on a legally protected group of people or a minority group of people. The legally protected group can vary from one country to another but race, ethnicity, gender, religion, etc. However, the attributes that seem to be used in favor or discriminated against a group of people in a specific domain are considered sensitive attributes. For example, although income is not a legally protected characteristic it turns out to be a sensitive attribute in deciding on granting a loan, and likewise, age turns out to be a sensitive attribute in the recruiting process. In line with the studies of AF, setting the boundaries of fairness or defining "how" to be fair in a particular domain is one of the important requirements to obtain fairer prediction outcomes within the desired framework. Based on traditional assumptions, political oppression, and culture, for example, in some countries, women are less favored than men, or groups of people with a certain religious belief are discriminated against, the definitions of fairness could be different. Hence, it is crucial to define the fairness notions or boundaries for a particular region or sector carefully, especially in machine-based decision-making systems where fairness is essentially on demand. This explains why there exist several number of definitions of fairness in the state-of-the-art literature on mitigating AF in different types of machine-based decision-making systems[2].

Customer-oriented (CO) businesses care about customer satisfaction with the products or services provided by the service providers. In other words, CO businesses need to pay more attention to fair treatment towards the customers as loyal customers bring more profits. One unfair treatment or discrimination against a group of minorities could lead to the reputation and sustainability of the business. Unlike other sectors, the CO sector has more chances to provide unfair treatment to the customers due to the demands and compatibility of the market. Covid-19 has caused many changes in people's lifestyles as well as business patterns. For example, there is a drop in travel rates, but people tend to buy more travel insurance[3] and using more contactless food ordering by using QR food ordering systems [4]. People have been more engaging with the internet, and on the other hand, the nature of business is more online-based, especially during and after the Covid-19 period. From shopping, learning news, and conducting meetings to studying and getting treatments, recommender systems provide more personalized suggestions to users or customers. To provide more accurate recommendations, demographics data, which include sensitive data such as gender, spending rate, and address/area, are required to collect[5]. Since using sensitive data in ML models could yield discriminatory outcomes, it is critically important to consider AF in CO businesses.

As mentioned earlier, there is a rich literature of AF definitions and notions in the existing literature. Once AF boundaries are set, the outcomes of the predictions or classifications are measured to determine whether any unfair outcomes go beyond the borders of AF or not. Therefore, fairness definitions and fairness measures are applied interchangeably in most situations. Existing bunches of AF measures are accepted as working well in the presented domains. Yet, no universal measure is the best measure to apply in every situation. In some cases, one measure contradicts another. On the other hand, one noteworthy fact is that CO is mainly two-sided since there are service providers and consumers. Hence, it is necessary to consider fairness towards both parties. This fact encourages investigating whether there is a need for different notions of fairness in the CO sector.

Despite several notions of fairness to date, many pieces of literature point out that there is still an incompatibility for different domains and that adjustments are still needed to work well in various sectors. Building the basic notion of algorithmic fairness in the interested domain is first and essential in reducing possible algorithmic bias in the outcomes. The primary purpose of this article is to investigate the reaches of algorithmic fairness principles specifically for CO domains, identify the gaps to formulate algorithmic fairness for CO sectors, and highlight the key factors to be considered in defining AF in CO domains. Several good review articles have been published in the area of interest. Yet, those emphasized the general domains and were more focused on an overall review of definitions of AF, approaches for mitigating bias in the predictive models, and AF evaluation metrics other than uniquely based on CO domains. Therefore, this Systematic Literature Review (SLR) will provide a comprehensive view of trends and gaps of AF in CO sectors to practitioners and future researchers. It also contributes to the existing research by highlighting directions of coping AF in upcoming customer-oriented ML applications.

This SLR is organized as follows. In section II, we provide research methodology, including research questions, detailed procedures of SLR, and the search process. In section III, we visualize the results based on the research questions (RQs) specifically. After that, we discuss the important findings in response to each RQ in the same session. In section IV, we conclude the SLR by summarizing the key results and implications.

II. Materials and Method

This investigation is a secondary study and is conducted as an SLR by systematically following the guidelines instructed by the very popular scholar Kitchenham[6]. All primary studies are carefully selected, stored, analyzed, and evaluated with the aid of Mendeley Desktop Software. The critical elements of an SLR fall under three categories: planning, analyzing the studies, and reporting the results. For the planning step, we first designate the research questions (RQs), deciding the targeted information sources, keywords to be applied, search period, etc. In analyzing studies, we first screen the titles, then conduct the review and analyze the selected articles thoroughly to get the responses to the designated research questions. Finally, we respond to the research questions through the results obtained from analyzing the articles.

A. Research Questions

We formulated the following relevant RQs to gather adequate evidence for our investigation of algorithmic fairness constraints in customer-oriented sectors.

- RQ 1: In which sectors/areas of machine-driven decision-making applications is algorithmic fairness primarily concerned?
- RQ 2: How do the most significant AF definitions contribute to the existing literature?
- RQ 3: Are existing algorithmic fairness notions/definitions directly applicable to the profit-oriented business domains? Why or why not?
- RQ 4: What are the factors/constraints to consider when defining algorithmic fairness in the profit-oriented business domains?

First of all, we would like to know the background status of algorithmic fairness in terms of its application areas. Recently, machine learning applications have been widely applied in almost every sector, but we would like to investigate how many of them have been concerned about algorithmic fairness. In other words, we would like to examine the percentage of concern for customer-oriented and profit-centered areas. That is the main purpose of formulating RQ 1 in our study.

We created RQ 2 since we would like to systematically organize and survey the existing algorithmic fairness. We work on this by extracting the applied algorithmic fairness definitions and the algorithmic constraints regardless of whether newly proposed or reused from the literature.

Regarding RQ 3, we would like to investigate whether the notions of existing algorithmic fairness could be brought into the sector of data-driven decision-making systems without modifying them or if it is still required to expand the existing algorithmic fairness definitions or formulate a new definition. This will lead us to compare their strengths and weaknesses and highlight the research gap in algorithmic fairness based on the requirements of the different contexts of applications.

In relationship with RQ 3, RQ 4 is set since we would like to determine whether we could bring the existing algorithmic fairness definitions into machine-driven decision-making applications in the profit-centered and customer-oriented sectors. In relationship with RQ 4, RQ 5 is created to determine the best-fit constraints or principles to be applied in the customer-oriented and profit-centered business sectors.

B. Search Process

Since the SLR study's main characteristic is finding evidence from the primary literature, we carefully selected the primary literature sources among the popular electronic databases. We included some databases for the literature search: ArXiv.org, IEEE Xplore, Science Direct, ACM Digital Library, and SpringerLink. We searched all relevant literature from the mentioned databases without limiting it to a journal or a conference paper. Since we intend to find out the definitions of algorithmic fairness applied in different sectors and their practices in the business sector, we emphasize the literature search on algorithmic fairness regardless of the industries or applied area. Therefore, the keywords for the literature search are algorithmic fairness and algorithmic bias. After we applied the AND/OR Boolean character, the search query became algorithmic "AND" fairness "OR" bias.

We determine the search period as well. Since algorithmic fairness studies were introduced in 2009, thus not to miss out on any evidence, we set the search period from January 2009 to August 2022. Therefore, our search period is around 14 years, over a decade of study. We believe we could bring thorough evidence to our research interest. The articles from arXiv.org are included in this SLR as arXiv.org collects AI systems' latest and innovative research tasks. The research articles are well-written and led by the leading AI researchers. Therefore, we include the articles from arXiv.org if their works do not violate our inclusion criteria.

C. Inclusion and Exclusion Criteria

After we have applied the designated keywords to the selected electronic databases, the total initial search gives 78,797 papers to be screened. Therefore, we set the criteria for selecting the most relevant articles. Our inclusion and exclusion criteria are as follows.

TABLE I
INCLUSION AND EXCLUSION CRITERIA

No	Inclusion Criteria (IC)	Exclusion Criteria (EC)
1	Relevance to the designated research questions?	1 non-English articles
2	Algorithmic fairness concerns in the specific sector?	2 Articles focused on fairness and transparency policies
3	Emphasis on definitions of AF	3 Articles with very little relevance to the research questions

Since our initial search results were tremendous, we created some refining processes to eliminate the irrelevant articles. The detailed steps for the refining process are shown in Fig 1.

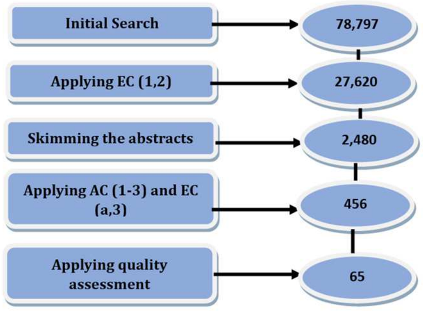

Fig. 1 Refining steps for the final selection of the articles

D. Quality Assessment

We conducted a quality assessment to finalize the selection of the articles since we collected a considerable number of articles as an initial result. The purpose of applying a quality assessment here is to gather the most relevant study that properly reflects our research questions. To do this, we created three survey questions as follows. We follow the procedures from the guidelines of [1].

- Does the study emphasize the problem of the algorithmic fairness constraints or definition in a specific area?
- Does the study mention a specific algorithmic fairness definition applied to evaluate or test out the machine-driven decision-making application?
- Does the study provide any suggestions or principles for defining fairness in ML-based decision-making systems?
- Does the study reveal the gap that necessitates the requirement of new constraints of algorithmic fairness in the customer-oriented sectors?

If the studies could settle the quality mentioned above assessment questions, we include them in our selection, but we would eliminate them if they do not reflect positive answers.

E. Data Extraction

We created a matrix table to extract the relevant data from the selected articles. We focused on the responses to our research questions and extracted the relevant data carefully, emphasizing the following information, as shown in Figure 2.

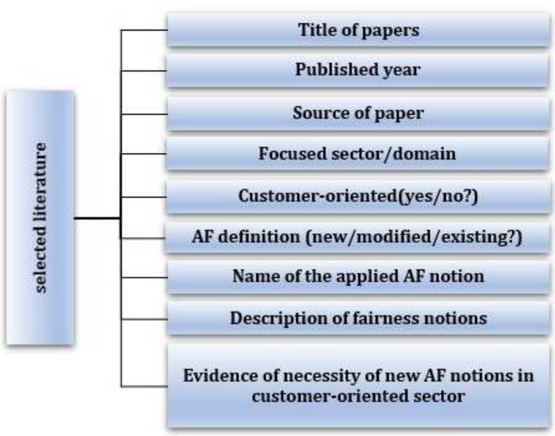

Fig. 2 Extracted data from the selected literature

F. Data Analysis

Define Data analysis is conducted to qualify the included articles. In other words, it is done to ensure the selected literature is most applicable to provide an SLR of the interested domain. To do so, after filtering out the non-applicable articles by using exclusion and inclusion criteria, we examine the selected primary studies pressing on the factors that: what are the sectors that ML-based decision-making applications have prioritized about AF; out of proposed AF definitions, what are the most significant and pervasive ones to date; challenges why existing AF constraints are not directly applicable to the CO domain; and finally the essential points to be focused in defining AF in CO domains.

III. Results and Discussion

In this section, the results of each RQ are provided, and the essential highlights of the key findings are discussed. Fig 3 illustrates the number of selected literature and their focusing area for AF concerns. A total of 43 definitions of AF are identified, and Table 2 enlists the new proposed and applied (existing) definitions, measures, and notions of AF. Table 2 is provided at the end of the article before the reference session. Fig 4 shows the statistical data of their focusing area for AF concerns. The table enlists the new proposed and applied (the existing ones) AF definitions/measures/notions. Fig 4 provides the statistical data of mainly applied AF notions. Out of 43 identified AF definitions, nine notions are used more than one and are ordered based on the reuse frequency. Fig 5 and 6 organize the percentage of selected literature, which emphasizes the CO domain and status of AF notions in the literature chosen, respectively.

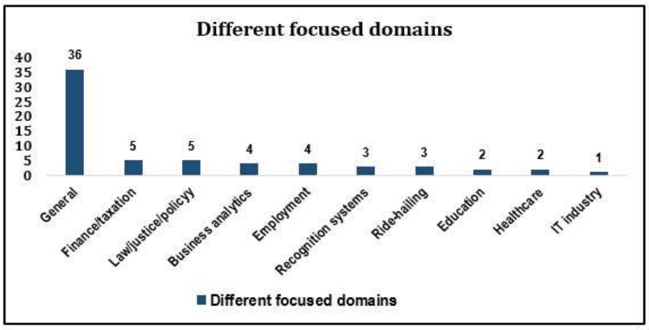

Fig. 3 Focused sectors of existing AF applications from the selected literature

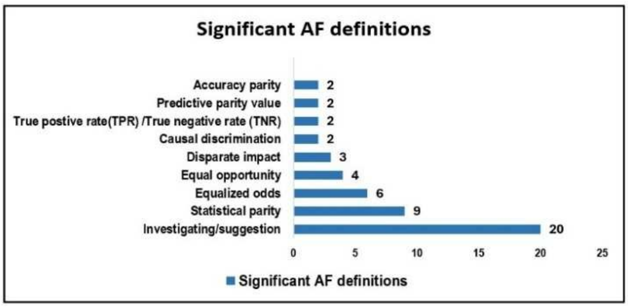

Fig. 4 Most pervasively applied AF notions in the literature

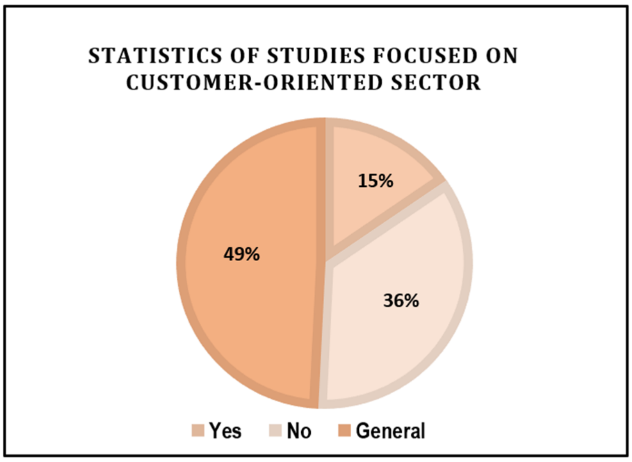

Fig. 5 Statistics of previous research focus on customer-oriented sectors

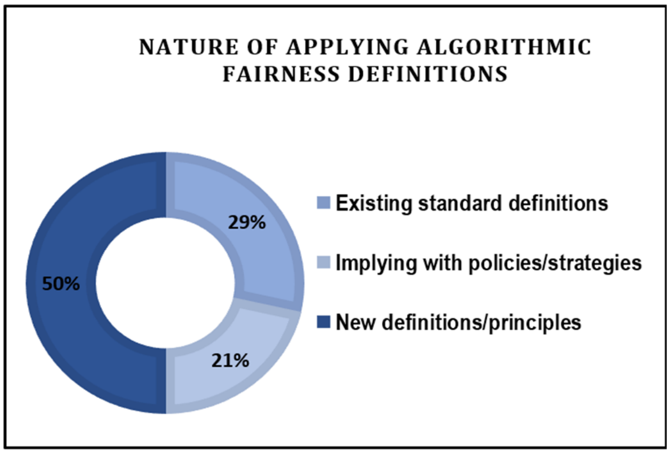

Fig. 6 Distribution of how AF notions are applied in the selected literature

TABLE II
ALGORITHMIC FAIRNESS NOTIONS/DEFINITIONS FROM THE SELECTED LITERATURE

No.	Name of fairness definitions	Description	No. of studies	Reference
1.	Moral value-based	A new approach to plug in the moral values in the decision-making systems by considering the five value-laden questions including how are the pros and cons towards the decision maker and decision subjects upon an algorithmic decision, which group of people are not receiving the utility as they deserved and should they tolerate upon not getting the same or not, and to which extent the fairness should be a trade-off with the utility of the decision-maker.	1	[7]
2.	Social injustice-based	The authors suggested the requirements to consider social justice in defining algorithmic fairness by discussing to which extent algorithmic fairness can be formulated considering morals and politics. They pointed out the necessity of collaboration between ethical, social, and political groups and computer scientists in combating structural injustices.	1	[8]

No.	Name of fairness definitions	Description	No. of studies	Reference
3.	Vertical equity-based	Vertical equity means to treat different individuals appropriately and differently.	1	[9]
4.	Fairness based on phenotype definitions	The evaluating practices are based on reducing three types of bias in health care, diagnosis biases, treatment biases, and access to care biases.	1	[10]
5.	Statistical parity/group fairness/demographic parity	It compares the decisions of both protected and unprotected groups whether they both have an equal probability of assigning to the positively predicted class or not.	9	[11][12][13][14][15][16][17][18][19]
6.	Conditional statistical parity	It compares the decisions of both protected and unprotected groups whether they both have an equal probability of assigning to the positive predicted class or not by controlling a set of legitimate factors.	1	[13]
7.	Equality of opportunity	It compares the decisions of both protected and unprotected groups whether they both have an equal false negative rate (FNR) or not.	4	[11][20][21][22]
8.	Equalized odds	It compares the decisions of both protected and unprotected groups whether they both have an equal true positive rate (TPR) and false positive rate (FPR) or not.	6	[12][23][14][21][24][25]
9.	Rawl's principle of equality of opportunity	It defines that equally talented people should have equal prospects of success.	1	[22]
10.	Casual discrimination	It defines the subjects from different groups with the same attributes that should be assigned to similar classes.	2	[26][27]
11.	True positive rate (TPR)	TPR or sensitivity (or recall) compares the decisions of whether the actual positive cases are truly assigned to the positive class or not.	2	[28][17]
12.	True negative rate (TNR)	TNR compares the decisions of whether the actual negative cases are truly assigned to the negative class or not.	2	[28][17]
13.	Consistency	It defines to predict the outcomes accurately and fairly by making sure the results are not influenced by the protected attributes.	1	[13]
15	Positive predictive value (PPV)	PPV or precision compares the decisions of both protected and unprotected groups and whether they both are equally predicted to the positive class.	2	[29][14]
16.	False omission rate (FOR)	FOR means the probability of the negatively predicted cases being assigned to the positive class.	1	[29]
17.	Sufficiency	This means that the rate of accuracy between the protected and unprotected groups should be equal.	1	[29]
18.	Error rate	This defines the rates of wrongly assigned cases that should be equal between the protected and unprotected groups.	1	[23]
19.	Product-policy-and implementation based	It focuses on fairness in the production system. It defines fairness by checking normative and descriptive questions at three levels, product, policy, and implementation.	1	[30]
20.	Gini aggregation rate for biometric equitability (GARBE)	It focuses on fairness in facial recognition systems. It is a summative aggregation of FDR and IR measures in which the bound can be reasonably controlled but it does not add or subtract error values that exist on markedly different scales.	1	[31]
21.	Positive/negative feedback probability	It focuses on the fairness of recommender systems for premium users. If the user clicks, bookmarks, and applications are taken, it is treated as positive feedback, it is treated as negative feedback if the user deletes the interactions.	1	[32]
22.	Conditional fairness	For the same conditions of a group of people, the cases for both positive and negative should have the same output.	1	[33]
23.	Allocative fairness	It is unfair if the candidates in one group have an inherently higher probability of receiving the resource candidates in another.	1	[34]
24.	Distributional fairness	The fairness notion focuses on getting the even distribution of beneficial outcomes across all members of an attribute group.	1	[35]
25.	Algorithmic fairness (regional-based)	The fairness definition is based on the cultures and social norms in India. The authors found out the different sensitive variables like skin tone, expenditure, language, etc.	1	[36]
26.	Social norms bias-based	It is designed to focus on a special type of harmful impact on subgroups based on gender to prevent social discrimination.[37]	1	[37]
27.	Contrastive fairness	It is the expansion of counterfactual logic which is designed to answer the contrastive question "Why this and not that? Upon the decision made by the algorithms.	1	[38]

No.	Name of fairness definitions	Description	No. of studies	Reference
28.	Accuracy parity/disparate mistreatment	Disparate mistreatment is intentional discrimination towards a protected group of people.	2	[39][35]
29.	Oblivious method	This fairness notion is based on the joint statistics of the predictor, the target, and the protected attribute but does not depend on the interpretation of individual features.	1	[40]
30.	Disparate Impact	It is a measurement of group fairness and means that the prediction rates for two different groups should not be different by more than 80%.	3	[41][35][42]
31.	Individual fairness	It means two individuals with similar attributes should have the same results.	1	[41]
32.	Calibration calibrated fairness	This measure checks the fairness between two individuals with a similar proportion in their merit.	2	[24][43]
33.	Differential fairness	This measure aims to ensure equitable treatment by an algorithm for all intersecting subgroups of a set of protected categories	1	[44]
34.	Subgroup fairness	This measure aims to prevent fairness gerrymandering at the intersections of protected groups.	1	[44]
35.	Intersectional fairness	It focuses on the multiple groups of protected attributes and considers the systematic unfair societal processes.	1	[45]
36.	FairKM	This measure applies multiple sensitive attributes and it preserves a proportional representation of protected classes within clusters as its representation in the whole dataset.	1	[46]
37.	Algorithmic discrimination	The discrimination formed by the algorithm's predictions as the results of influences of conformity to assumed social norms.	1	[47]
38.	AFR-based fairness	This means the algorithmic fairness based on accountability for reasonableness (AFR) which is a modified measure by considering the publicity condition, relevance condition, revision and appeals condition, and regulative condition.	1	[48]
39	Investigation/suggestions for designing a better algorithmic fairness definition	Specific fairness notions are not proposed but highlighting the gaps in defining algorithmic fairness in a particular domain or proposing the general principles or guidelines in defining fairness	21	[49][50][51][2][52][53][54][55][56][57][58][59][60][61][62][63][64][65][66][64][67]
40.	Task-specific similarity metric	A notion of setting the extent to extract the best similar individuals to classify. (need to make it smooth)	1	[68]
41.	Geographic fairness or passenger-side fairness (PSF) and Income-fairness or driver-side fairness(DSF))	PSF- The probability of receiving a ride should not depend on the passengers' origin or destination. DSF- All drivers should earn similar incomes	1	[19]
42.	Intra-fairness	Based on proportional fairness, it requires that over time, the accumulated utility proportional to the workload for every worker should be equalized.	1	[69]
43.	Inter-fairness	Based on demographic parity, it needs to ensure that the proportion of accumulated utility to the workload should be equalized across the subgroups.	1	[69]

A. RQ 1: In which sectors/areas of machine-driven decision-making applications is AF a concern?

We formulated this RQ to investigate the statistics of AF considerations in different areas. The results are elaborated on and discussed accordingly for four commonly applied sectors among ten identified focused sectors.

1) Overview: Over half of the selected literature did not focus on any particular sector or domain. Among the carefully selected articles, 55.4% of research discussed general topics which means they did not emphasize any specific area. Instead, they apply their proposed approaches and methods to the existing publicly available datasets which are from a variety of fields but those approaches are not particularly designed for a specific sector. Out of 55.4%, some of the studies highlighted the loopholes in existing AF definitions and discussed the gaps by providing principles and guidance in defining AF notions regardless of the specific focused domains [21], [67], [70]. From this, we could remark that most researchers desire to embody the emergence of generalized AF, which could be achieved in different types of domains.

2) Second focused sectors: The second leading sectors are law/justice/policy and finance/taxation, sharing 7.7% each. The former involves machine-based decision-making systems applied in making court decisions and also in the domain of gender discrimination [17], [61] etc., while the latter covers the granting of loans, income tax, and credit scoring systems [2], [40]etc. The percentage of AF consideration on law and justice is considered to be less, as the statistics show it is only under 10%. Although AF awareness in ML-based applications is increased significantly via the case studies of criminal justice [71]. Yet only under 10% of studies worked in the sector of law and justice. Finance and taxation play a critical role in a country, and banks are transforming more into ML-based decision-making systems. However, fair classifiers are not widely applied in these domains yet.

3) Third-focused sectors: Business analytics and employment sectors are mainly applied sectors. The term business here organizes the areas or industries looking forward to direct and explicit financial gains or profits, such as business analytics and e-commerce applications [1], [12]. The sectors such as healthcare, education, and finance could be considered profit-oriented. However, those areas are counted separately since the benefactors could be government or non-profit organizations, such as machine learning-based diagnosis systems, student scholarship programs, home loan programs for the citizens, etc. Since it is not clearly described that those applications target profitable or non-profitable purposes, we do not count into the profit-focused groups in our study.

The existing researchers also emphasized indirect profit-oriented sectors, including face recognition systems and employment areas. Most face recognition systems and job recruiting/hunting applications can be used freely. Still, app creators' profit from advertisements even though we do not count them in the list of business sectors here. One interesting thing we observed via our study is that some of the studies focusing on the direct profit-oriented sectors did not propose new measures tailored to the business requirements, leading to new algorithmic fairness definitions.

B. RQ 2 How do the most significant AF definitions contribute to the existing literature?

Existing scholars have contributed significantly to the literature on AF notions as 43 different AF notions are identified through our selected studies. The brief definitions of However, it was observed that not all of them are popular and repeatedly applied. The most significant and most applicable ones are statistical parity with 9 times [11], [12], [13], [14], [15], [16], [17], [18], [19], equalized odds with 6 times [12], [14], [21], [23], [24], [25], and equality of opportunity with four times [11], [20], [21], [22] reuse. These AF measures are briefly explained with an easy example to provide a more comprehensive understanding. In our example, an ML model will predict which customers are switching the service (churners) and which customers are staying with the existing brand (non-churners). In consideration of AF in the model, we do not want the model to give less favor to the group of low-income families. In that case, the most applicable AF notions could be explained in the following way.

1) Statistical parity (SP): This measure is quite popular and frequent due to its simplicity. It is usually applied to specific people or categories rather than emphasizing individual fairness. It compares the decisions of both protected and unprotected groups to determine whether they have an equal probability of assigning to the positively predicted class. With our example, SP should give equal results for predicting churn for customers of different income levels. Its popularity might be due to the simplicity of the measure and suitability since most domains look for fairness for customers or users at a group level. On the other hand, this measure aligns with the legal practices in housing and employment.

2) Equalized odds: Hardt et al. [40] proposed the notion of equalized odds due to the shortcomings of statistical parity. It compares the decisions of both protected and unprotected groups to determine whether they have equal true positive rates (TPR) and false positive rates (FPR). With this notion, the classifier should give equal true positive rates (TPR) and false positive rates (FPR) to both churn and non-churn groups.

3) Equality of opportunity: It compares the decisions of protected and unprotected groups to determine whether they have an equal false negative rate (FNR). This example checks if there is an equal probability of predicting non-churners to be churners (and vice versa) between customers with low-income and high-income levels. Each AF notion has its strengths and weaknesses. A critical expectation of fairer ML models is to provide accurate outcomes while producing unbiased decisions. To achieve this goal in different domains, new AF notions have emerged competitively.

C. RQ 3: Are existing algorithmic fairness notions/definitions directly applicable to customer-oriented domains?

Existing algorithmic fairness research works start with seeking the metrics to measure fairness in specific contexts and formularizing them into mathematical forms. If the prediction systems are not met with the algorithmic fairness formulations, it would be assumed that the systems produce unfairness [7]. In practice, defining fairness is more complex due to its subjective nature. Thus, it needs to be considered from different perspectives, legally, socially, and ethically, based on the contexts of the applications. Recently, according to the era's demands, traditional businesses have transformed into business analytics systems excessively. Hence, the decisions rely more on data-driven and machine-based decision-making systems. However, from the perspective of fairness implications, these systems bring unfairness and injustices to a particular group of people with or without intentions.

In the BA domain, most applications focus on predictive analytics, which requires using historical data to predict the desired outcomes. Thus it is prone to suffer from one type of bias in the training data named skewed example[72]. On the other hand, some predictive applications, such as product recommender systems, use social media data as training data. Therefore, there is a high risk of producing unfair predictions since social media data could be biased.

Based on what one wants to let algorithms decide, it is essential to provide straightforward definitions that make the outcomes of the algorithms evaluable [64]. The profit-centered and customer-oriented business sectors aim to increase financial gain and enhance customers' trust in the businesses. Thus, these two business requirements should mainly be considered when defining AF so that it would produce fair decisions for both parties. Since there is a different spectrum of businesses in [1]. The authors discussed the importance of pinpointing which business sectors require more attention since they are prone to producing unfairer outcomes when using predictive systems.

Our observation reveals that 43 different fairness notions are proposed to enhance and fit well in achieving fairer outcomes. Besides, over 20 studies highlight the shortcomings of existing fairness notions and provide suggestions and guidelines for future direction in defining fairness principles. These facts prove the necessity of settling fairness notions in a particular domain. On the other hand, out of 65 studies, only a few are focused on customer-oriented domains, which suggests the insufficiency of existing notions, and we express our observation that there still needs to be a gap in setting boundaries of fairness in CO domains. There are a few challenging questions, such as whether these proposed notions are practically applicable in real-world situations, whether they work with the dynamic datasets, worked well on the sample datasets in the proposals, etc.

D. RQ 4: What factors should be considered when defining algorithmic fairness in customer-oriented domains?

We observed the essential factors to be considered in defining algorithmic fairness in the profit-oriented business domains.

1) Balancing in profit, accuracy, and algorithmic fairness: Previous literature related algorithmic fairness with actual fairness and how impactful the prediction results on a person or a group of people in real life. Most case studies are based on justice, legal, societal, and preventing discrimination. Regarding business, the primary preference is to gain profit and not lose customers. Therefore, it is equally important to balance gaining profit, getting accurate prediction results, and providing fair treatment to the customers. On the other hand, producing accurate results reduces the unfairness of the respective group of people and increases the profit. Concerning our example, when a classifier predicts the actual churners as the churners correctly, the service provider would appeal to those groups of customers back to their service with the appropriate incentives. Therefore, it would reduce the impact on profitability if the customers feel they have received care from the services they are using. Thus, profitability and accuracy come together. At the same time, the threshold of how much accuracy could be traded off in gains of algorithmic fairness and vice versa needs to be considered.

2) Harmony in choosing the right sensitive attributes and defining algorithmic fairness: There are usually target and predictor variables in supervised classification or prediction. In our example, the target variable is what we will predict whether a customer will be a churner or a non-churner. Predictor variables are those which support to obtain the most accurate results in the prediction process. In terms of measuring algorithmic fairness tasks, selecting the sensitive variables or attributes is required to ensure that using those variables would not alter the prediction results to the discriminated ones towards minority or unprivileged groups.

The nature of sensitivity in the variables could vary due to the different domains. In customer-oriented and profit-centered domains, the sensitive variable could be the income, the package the customers have bought, and the type of the customers (e.g., gold customers or diamond customers, etc.) rather than gender or religion. Therefore, choosing the most suitable sensitive attributes would be very helpful in measuring algorithmic unfairness in any domain.

3) Variations based on the regions, cultures, and boundaries: The availability and reliability of historical data could be different based on the country or region. At a detailed level, numerous cases reveal that utilization of ML for unjust price differentiation is based solely on customers' socio-economic and personal characteristics like age, income, education level, and zip codes. Several American insurance firms have imposed higher health insurance rates on immigrants, ethnic minorities, and marginalized communities, believing these groups are high-risk and may need more healthcare resources. Such restrictions could vary country by country.

In some developing countries or regions, the available data could not be trustworthy due to several factors, including long-term corruption, systematic discrimination against certain people, and lack of adequate data. On the other hand, the opinions on the definition of fair treatment and what is not could be different based on the variations of cultures and beliefs between different regions. For example, zip code can be a sensitive variable in the United States since it reveals which races of people are staying in a particular area. But it would not be the case in another country. In [36], the authors pointed out this factor by comparing the different assumptions of algorithmic fairness with those of Indian and Western cultures. These scenarios show that defining enhanced fairness notions for a specific domain is required rather than using the existing ones to obtain fairer outcomes.

IV. CONCLUSION

To sum up, algorithmic fairness has emerged as a crucial role to consider in the rapidly growing demands of automated decision-making applications. Especially in the domain of the customer-oriented sector, where equal treatment towards the customers is desirable to sustain the businesses. Before solving the problem of unfair outcomes of algorithmic decisions, it is essential to work on defining "fairness" [48], which is also a supportive idea for solving the former problem.

In this SLR, we work on revisiting the definitions of algorithmic fairness and the boundaries and limitations of algorithmic fairness demonstrated in the previous literature. We carefully included a total of 65 articles. We identified that under 10% of existing literature explicitly focuses on the CO domain. Hence, it shows that there is a gap to bridge between the existing fairness definitions and the demands of the CO domain. It is noteworthy to find the three key factors to consider in formulating fairness in the context of CO business applications. Those outlines would significantly contribute to developing fairer CO business applications shortly. In the following sessions, the limitations of the study, research gaps, and implications are discussed briefly.

A. Limitations of the study

In this study, we focused on the existing literature on algorithmic fairness, highlighting fairness notions in different sectors and formulating those definitions into the mathematical fairness metrics. Our focus is to investigate the factors to consider in formulating algorithmic fairness in the business sectors, which are customer-oriented and profit-centered. We searched the most significant literature in the five prominent electronic databases. Even though we tried our best to work on the literature search, there might have been some literature that could have been missed. Moreover, most of the related literature was found in arXiv.org, an archive of the recent articles. Due to its relatedness, we include more literature from arXiv.org than other electronic databases.

The second limitation is that although we created the assessment criteria for inclusion and exclusion of the literature to gain the practical selection of articles, some of the good literature needs to directly discuss which algorithmic fairness definitions they are applying. Instead, they implied and related with the philosophy of fairness with the algorithmic fairness. In that case, we could not remark exactly which definition they were applying, but we included such articles due to the relatedness of the contents.

B. Research gaps and implications

Reviewing the results in RQ1, existing research focuses on general domains more than specific domains. Based on the statistical data, only a few domains are emphasized (see Fig.3), and several sectors are left to be considered for AF practices. These limited empirical studies of AF in various industries would yield restricted benchmarking when the AF problems come into different sectors in the future. On the other hand, combating AF issues will be dragged back due to a lack of focus on perceived fairness. For example, scrutinizing fairness principles in the customer care service and recruitment sector would be very different. The results of RQ 2 and 3 show that the demands of redefining AF are required to produce a fairer and trustworthy decision. This indicates that more specific and personalized AF constraints will be coming up, which will inflate AF definitions.

By considering the identified factors from the discussion of RQ 4, future research directions include redefining AF principles for customer-oriented sectors, which should not violate consumers' rights and should not impact service or product owners. One important thing for future researchers is that a person or a group of customers could be unfairly treated in two or more different groups concurrently based on the involvement of different sensitive attributes in the datasets. For example, a woman (i.e., gender) with low income (i.e., income status) is decided as a churner while she is not churned. Therefore, multiple features should be considered when developing a framework for fairer prediction models. Subsequently, a fair framework that considers the different perspectives, including ethical, legal, and profitability, would be in demand in machine-assisted decision-making systems.

Acknowledgment

We thank TM Research and Development from Telekom Malaysia (TM), Malaysia for supporting fund for this work (Ref: MMUE/190002).

References

[1] M. De-Arteaga, S. Feuerriegel, and M. Saar-Tsechansky, "Algorithmic fairness in business analytics: Directions for research and practice," *Prod. Oper. Manag.*, vol. 31, no. 10, pp. 3749–3770, 2022, doi: 10.1111/poms.13839.

[2] S. Verma and J. Rubin, "Fairness definitions explained," *Proc. - Int. Conf. Softw. Eng.*, pp. 1–7, 2018, doi: 10.1145/3194770.3194776.

[3] S. T. Lim, J. Y. Yuan, K. W. Khaw, and X. Chew, "Predicting Travel Insurance Purchases in an Insurance Firm through Machine Learning Methods after COVID-19," *J. Informatics Web Eng.*, vol. 2, no. 2, pp. 43–58, 2023, doi: 10.33093/jiwe.2023.2.2.4.

[4] C.-C. Wong, L.-Y. Chong, S.-C. Chong, and C.-Y. Law, "QR Food Ordering System with Data Analytics," *J. Informatics Web Eng.*, vol. 2, no. 2, pp. 249–272, 2023, doi: 10.33093/jiwe.2023.2.2.18.

[5] Z. Y. Poo, C. Y. Ting, Y. P. Loh, and K. I. Ghauth, "Multi-Label Classification with Deep Learning for Retail Recommendation," *J. Informatics Web Eng.*, vol. 2, no. 2, pp. 218–232, 2023, doi:10.33093/jiwe.2023.2.2.16.

[6] H. R. Kouchaksaraei and H. Karl, "Service function chaining across openstack and kubernetes domains," *DEBS 2019 - Proc. 13th ACM Int. Conf. Distrib. Event-Based Syst.*, pp. 240–243, 2019, doi:10.1145/3328905.3332505.

[7] C. Hertweck, J. Baumann, M. Loi, E. Viganò, and C. Heitz, "A Justice-Based Framework for the Analysis of Algorithmic Fairness-Utility Trade-Offs," pp. 1–15, 2022.

[8] A. Kasirzadeh, "Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy," *AIES 2022 - Proc. 2022 AAAI/ACM Conf. AI, Ethics, Soc.*, pp. 349–356, 2022, doi:10.1145/3514094.3534188.

[9] E. Black, H. Elzayn, A. Chouldechova, J. Goldin, and D. Ho, "Algorithmic Fairness and Vertical Equity: Income Fairness with IRS Tax Audit Models," *ACM Int. Conf. Proceeding Ser.*, pp. 1479–1503, 2022, doi: 10.1145/3531146.3533204.

[10] T. Y. Sun, S. Bhave, J. Altosaar, and N. Elhadad, "Assessing Phenotype Definitions for Algorithmic Fairness," pp. 1–13, 2022.

[11] M. Schmitz, R. Ahmed, and J. Cao, "Bias and Fairness on Multimodal Emotion Detection Algorithms," 2022, doi:10.13140/RG.2.2.14341.01769.

[12] M. von Zahn, S. Feuerriegel, and N. Kuehl, "The Cost of Fairness in AI: Evidence from E-Commerce," *Bus. Inf. Syst. Eng.*, vol. 64, no. 3, pp. 335–348, 2022, doi: 10.1007/s12599-021-00716-w.

[13] C. Kern, R. L. Bach, H. Mautner, and F. Kreuter, "Fairness in Algorithmic Profiling: A German Case Study," pp. 1–33, 2021.

[14] N. Kozodoi, J. Jacob, and S. Lessmann, "Fairness in credit scoring: Assessment, implementation and profit implications," *Eur. J. Oper. Res.*, vol. 297, no. 3, pp. 1083–1094, 2022, doi:10.1016/j.ejor.2021.06.023.

[15] R. J. Chen *et al.*, "Algorithm Fairness in AI for Medicine and Healthcare," pp. 1–49, 2021.

[16] M. A. Bakker *et al.*, "DADI: Dynamic Discovery of Fair Information with Adversarial Reinforcement Learning," no. NeurIPS, 2019.

[17] J. Sargent and M. Weber, "Identifying biases in legal data: An algorithmic fairness perspective," *ACM Conf. Equity Access Algorithms, Mech. Optim. Oct. 5-9, 2021*, vol. 1, no. 1, pp. 1–14, 2021.

[18] Y. Yang, Y. Wu, X. Chang, and M. Li, "Toward a Fairness-Aware Scoring System for Algorithmic Decision-Making," 2021.

[19] A. Kumar, Y. Vorobeychik, and W. Yeoh, "Using Simple Incentives to Improve Two-Sided Fairness in Ridesharing Systems," *Proc. Int. Conf. Autom. Plan. Sched. ICAPS*, vol. 33, no. 1, pp. 227–235, 2023, doi: 10.1609/icaps.v33i1.27199.

[20] Y. Yu and G. Saint-Jacques, "Choosing an algorithmic fairness metric for an online marketplace: Detecting and quantifying algorithmic bias on LinkedIn," *arXiv Comput. Sci.*, 2022.

[21] A. Noriega-Campero, B. Garcia-Bulle, M. A. Bakker, and A. S. Pentland, "Active fairness in algorithmic decision making," *AIES 2019 - Proc. 2019 AAAI/ACM Conf. AI, Ethics, Soc.*, pp. 77–83, 2019, doi:10.1145/3306618.3314277.

[22] F. Arif Khan, E. Manis, and J. Stoyanovich, "Towards Substantive Conceptions of Algorithmic Fairness: Normative Guidance from Equal Opportunity Doctrines," *ACM Int. Conf. Proceeding Ser.*, vol. 1, no. 1, pp. 1–16, 2022, doi: 10.1145/3551624.3555303.

[23] M. Karimi-Haghighi, C. Castillo, D. Hernandez-Leo, and V. M. Oliver, "Predicting Early Dropout: Calibration and Algorithmic Fairness Considerations," no. Ml, pp. 1–10, 2021.

[24] G. Pleiss, M. Raghavan, F. Wu, J. Kleinberg, and K. Q. Weinberger, "On fairness and calibration," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5681–5690, 2017.

[25] B. Woodworth, S. Gunasekar, M. I. Ohannessian, and N. Srebro, "Learning Non-Discriminatory Predictors," no. 1, 2017.

[26] K. Mohammadi, A. Sivaraman, and G. Farnadi, "FETA: Fairness

Enforced Verifying, Training, and Predicting Algorithms for Neural Networks," 2022.

[27] S. Galhotra, Y. Brun, and A. Meliou, "Fairness testing: Testing software for discrimination," *Proc. ACM SIGSOFT Symp. Found. Softw. Eng.*, vol. Part F1301, pp. 498–510, 2017, doi:10.1145/3106237.3106277.

[28] W. Jiang and Z. A. Pardos, *Towards Equity and Algorithmic Fairness in Student Grade Prediction*, vol. 1, no. 1. Association for Computing Machinery, 2021. doi: 10.1145/3461702.3462623.

[29] J. Baumann, A. Hannák, and C. Heitz, "Enforcing Group Fairness in Algorithmic Decision Making: Utility Maximization Under Sufficiency," *ACM Int. Conf. Proceeding Ser.*, pp. 2315–2326, 2022, doi: 10.1145/3531146.3534645.

[30] C. Bakalar *et al.*, "Fairness On The Ground: Applying Algorithmic Fairness Approaches to Production Systems," 2021.

[31] J. J. Howard, E. J. Laird, Y. B. Sirotin, R. E. Rubin, J. L. Tipton, and A. R. Vemury, "Evaluating Proposed Fairness Models for Face Recognition Algorithms," pp. 1–11, 2022.

[32] Y. Li *et al.*, "Contextualized Fairness for Recommender Systems in Premium Scenarios," *Big Data Res.*, vol. 27, p. 100300, 2022, doi:10.1016/j.bdr.2021.100300.

[33] R. Xu *et al.*, "Algorithmic Decision Making with Conditional Fairness," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 2125–2135, 2020, doi: 10.1145/3394486.3403263.

[34] H. Elzayn *et al.*, "Fair algorithms for learning in allocation problems," *FAT* 2019 - Proc. 2019 Conf. Fairness, Accountability, Transpar.*, pp. 170–179, 2019, doi: 10.1145/3287560.3287571.

[35] N. Grgić-Hlača, M. B. Zafar, K. P. Gummadi, and A. Weller, "On Fairness, Diversity and Randomness in Algorithmic Decision Making," 2017.

[36] N. Sambasivan, E. Arnesen, B. Hutchinson, T. Doshi, and V. Prabhakaran, "Re-imagining algorithmic fairness in India and beyond," *FAccT 2021 - Proc. 2021 ACM Conf. Fairness, Accountability, Transpar.*, pp. 315–328, 2021, doi:10.1145/3442188.3445896.

[37] M. Cheng, M. De-Arteaga, L. Mackey, and A. T. Kalai, *Social Norm Bias: Residual Harms of Fairness-Aware Algorithms*, vol. 1, no. 1. Association for Computing Machinery, 2021.

[38] T. Chakraborti, A. Patra, and J. A. Noble, "Contrastive Fairness in Machine Learning," *IEEE Lett. Comput. Soc.*, vol. 3, no. 2, pp. 38–41, 2020, doi: 10.1109/locs.2020.3007845.

[39] H. Zhao and G. J. Gordon, "Inherent Tradeoffs in Learning Fair Representations," *J. Mach. Learn. Res.*, vol. 23, no. NeurIPS, 2022.

[40] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *Adv. Neural Inf. Process. Syst.*, pp. 3323–3331, 2016.

[41] P. K. Lohia *et al.*, "Bias Mitigation Post-Processing for Individual and Group Fairness IBM Research and IBM Watson AI Platform 1101 Kitchawan Road , Yorktown Heights , NY , USA," pp. 2847–2851, 2019.

[42] S. Barocas, "Big Data ' S Disparate Impact," vol. 671, pp. 671–732, 2014.

[43] N. A. Saxena, E. Defilippis, G. Radanovic, and D. C. Parkes, "How Do Fairness Definitions Fare ? Examining Public Attitudes Towards Algorithmic Definitions of Fairness," 2018.

[44] J. Foulds, R. Islam, K. Keya, and S. Pan, "Bayesian Modeling of Intersectional Fairness: The Variance of Bias," 2018.

[45] J. R. Foulds, R. Islam, K. N. Keya, and S. Pan, "An intersectional definition of fairness," *Proc. - Int. Conf. Data Eng.*, vol. 2020-April, pp. 1918–1921, 2020, doi: 10.1109/ICDE48307.2020.00203.

[46] S. S. Abraham, P. Deepak, and S. S. Sundaram, "Fairness in clustering with multiple sensitive attributes," *Adv. Database Technol. - EDBT*, vol. 2020-March, pp. 287–298, 2020, doi: 10.5441/002/edbt.2020.26.

[47] I. Serna, A. Morales, J. Fierrez, and N. Obradovich, "Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning," *Artif. Intell.*, vol. 305, p. 103682, 2022, doi: 10.1016/j.artint.2022.103682.

[48] P. Wong, "Democratizing Algorithmic Fairness," no. 2010, 2019.

[49] M. Veale, M. Van Kleek, and R. Binns, "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making," pp. 1–14, 2018.

[50] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," *FAT* 2019 - Proc. 2019 Conf. Fairness, Accountability, Transpar.*, pp. 59–68, 2019, doi: 10.1145/3287560.3287598.

[51] A. Parkavi, A. Jawaid, S. Dev, and M. S. Vinutha, "The Patterns that Don't Exist : Study on the effects of psychological human biases in data analysis and decision making," *Proc. 2018 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2018*, pp. 193–197, 2018, doi: 10.1109/CSITSS.2018.8768554.

[52] M. Robles Carrillo, "Artificial intelligence: From ethics to law," *Telecomm. Policy*, vol. 44, no. 6, pp. 1–16, 2020, doi:10.1016/j.telpol.2020.101937.

[53] M. P. Hauer, J. Kevekordes, and M. A. Haeri, "Legal perspective on possible fairness measures – A legal discussion using the example of hiring decisions," *Comput. Law Secur. Rev.*, vol. 42, p. 105583, 2021, doi: 10.1016/j.clsr.2021.105583.

[54] K. Makhlouf, S. Zhioua, and C. Palamidessi, "Machine learning fairness notions: Bridging the gap with real-world applications," *Inf. Process. Manag.*, vol. 58, no. 5, p. 102642, 2021, doi:10.1016/j.ipm.2021.102642.

[55] S. Kleanthous, M. Kasinidou, P. Barlas, and J. Otterbacher, "Perception of fairness in algorithmic decisions: Future developers' perspective," *Patterns*, vol. 3, no. 1, 2022, doi:10.1016/j.patter.2021.100380.

[56] D. C. Parkes and R. V. Vohra, "Algorithmic and economic perspectives on fairness", *arXiv*:1909.05282, 2019.

[57] S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq, "Algorithmic decision making and the cost of fairness," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F1296, pp. 797–806, 2017, doi: 10.1145/3097983.3098095.

[58] B. J. Kleinberg, J. Ludwig, S. Mullainathan, and A. Rambachan, "Algorithmic Fairness †," pp. 22–27, 2018.

[59] H. J. P. Weerts, "An Introduction to Algorithmic Fairness," pp. 1–18, 2021.

[60] N. Zhou, Z. Zhang, V. N. Nair, H. Singhal, J. Chen, and A. Sudjianto, "Bias, Fairness, and Accountability with AI and ML Algorithms," pp. 1–18, 2021.

[61] E. Pierson, "Demographics and discussion influence views on algorithmic fairness," pp. 1–10, 2017.

[62] R. Binns, "Fairness in Machine Learning: Lessons from Political Philosophy," no. 2016, pp. 1–11, 2017.

[63] S. Corbett-Davies and S. Goel, "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning," no. Ec, 2018.

[64] S. Passi and S. Barocas, "Problem formulation and fairness," *FAT 2019 - Proc. 2019 Conf. Fairness, Accountability, Transpar.*, pp. 39–48, 2019, doi: 10.1145/3287560.3287567.

[65] D. Pessach and E. Shmueli, "A Review on Fairness in Machine Learning," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–44, 2023, doi:10.1145/3494672.

[66] I. Ahmed, G. L. Colclough, and D. First, "Building Fair and Transparent Machine Learning via Operationalized Risk Management: Towards an Open-Access Standard Protocol," *Int. Conf. Mach. Learn. AI Soc. Good Work.*, 2019.

[67] N. A. Saxena, W. Zhang, and C. Shahabi, "Missed Opportunities in Fair AI," *Proc. 2023 SIAM Int. Conf. Data Min.*, pp. 961–964, 2023, doi: 10.1137/1.9781611977653.ch110.

[68] C. Dwork, M. Hardt, and R. Zemel, "Fairness Through Awareness," pp. 214–226.

[69] Q. Zhou, J. Mareček, and R. Shorten, "Subgroup fairness in two-sided markets," *PLoS One*, vol. 18, no. 2 February, pp. 1–25, 2023, doi:10.1371/journal.pone.0281443.

[70] M. Veale, M. Van Kleek, and R. Binns, "Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 2018-April, 2018, doi: 10.1145/3173574.3174014.

[71] A. Chouldechova, "Fair prediction with disparate impact : A study of bias in recidivism prediction instruments," pp. 1–17, 2017.

[72] S. Barocas and A. D. Selbst, "Big Data ' s Disparate Impact," vol. 671, pp. 671–732, 2016.