Educating Computer Science Students about Algorithmic Fairness, Accountability, Transparency and Ethics

Maria Kasinidou
Cyprus Center for Algorithmic Transparency, Open University of Cyprus
Nicosia, Cyprus
maria.kasinidou@ouc.ac.cy

Styliani Kleanthous
Cyprus Center for Algorithmic Transparency, Open University of Cyprus & CYENS Centre of Excellence
Nicosia, Cyprus
styliani.kleanthous@ouc.ac.cy

Kalia Orphanou
Cyprus Center for Algorithmic Transparency, Open University of Cyprus
Nicosia, Cyprus
kalia.orphanou@ouc.ac.cy

Jahna Otterbacher
Cyprus Center for Algorithmic Transparency, Open University of Cyprus & CYENS Centre of Excellence
Nicosia, Cyprus
jahna.otterbacher@ouc.ac.cy

ABSTRACT

Professionals are increasingly relying on algorithmic systems for decision making however, algorithmic decisions occasionally perceived as biased or not just. Prior work has provided evidences that education can make a difference on the perception of young developers on algorithmic fairness. In this paper, we investigate computer science students' perception of FATE in algorithmic decision-making and whether their views on FATE can be changed by attending a seminar on FATE topics. Participants attended a seminar on FATE in algorithmic decision-making and they were asked to respond to two online questionnaires to measure their pre- and post-seminar perception on FATE. Results show that a short seminar can make a difference in understanding and perception as well as the attitude of the students towards FATE in algorithmic decision support. CS curricula need to be updated and include FATE topics if we want algorithmic decision support systems to be just for all.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

algorithmic fairness, algorithmic transparency, algorithmic accountability, algorithmic decision making

ACM Reference Format:
Maria Kasinidou, Styliani Kleanthous, Kalia Orphanou, and Jahna Otterbacher. 2021. Educating Computer Science Students about Algorithmic Fairness, Accountability, Transparency and Ethics. In *26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2021), June 26–July 1, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3430665.3456311

ITiCSE 2021, June 26–July 1, 2021, Virtual Event, Germany

ACM ISBN 978-1-4503-8214-4/21/06.
https://doi.org/10.1145/3430665.3456311

1 INTRODUCTION

The use of algorithmic systems for taking important decisions on humans' behalf has rapidly grown and has produced many ethical dilemmas and little consensus about how to resolve them. Algorithmic systems are used for deciding which posts and news we will see on social media [34], prison releases[1][1], ranking job applicants [26], deciding who is entering university [2], recommending courses [4], who will be getting a loan [18] and many more. However, such systems do not always behave as they should thus, reproducing and/or amplifying social stereotypes and inequalities [5]. There are many examples that witness the misbehavior of these systems. Gender discrimination has been detected in resume search engines [7]; auto-complete search terms showed that suggested terms could be viewed as racist, sexist, or homophobic [2]; image search results are gender-biased depending on the search term used [27] and racially-biased towards black teenagers [20].

While much effort has been devoted for developing frameworks of algorithmic fairness [8] and algorithmic models to alleviate biases [21], it will be hard to achieve consensus due to the complexity of 'Fairness' as a concept unless we understand how people perceive concepts such as algorithmic fairness (e.g. [3, 15, 17, 39]), accountability [25, 37] and transparency [12, 29, 35, 36].

Previous work has looked into how the end-users and/or the general public perceive elements of Fairness, Accountability, Transparency and Ethics (FATE) however, it is important to understand how the people who are soon to be involved in developing these algorithmic systems - Computer Science (CS) students- perceive the above concepts. It is also important to understand whether their perception can change after attending related courses and/or training on FATE. To our knowledge, their has not been prior work that looked into the perception of CS students on FATE in algorithmic decision-making (DM) systems and how this changes after they attended seminars on these topics.

We investigate how perception on algorithmic FATE is affected by related course/training, by surveying students in two CS classes, both before and after a seminar on the above topics. We presented

[1] www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[2] https://www.theguardian.com/education/2020/aug/15/controversial-exams-algorithm-to-set-97-of-gcse-results

participants with two scenarios of algorithmic DM systems describing different contexts and asked them to indicate their agreement regarding six statements related to the fairness and justice constructs extracted from related work on fairness and justice [9, 22]. We also asked participants whether they would consider dimensions of fairness, transparency and accountability in a system they develop, and to state their knowledge on Fairness, Accountability, Transparency and Ethics. In the post-questionnaire participants were also asked to ponder *which part of the process would you think it could possibly cause unfairness*. Finally, they were asked to state who would be *held accountable* if the system behaved unfairly to some parts of the population?

2 BACKGROUND

Fairness, Accountability, Transparency and Ethics (FATE) in algorithmic DM systems have been a central of interest within the field of HCI, mathematics, science and related disciplines. The use of algorithmic DM systems is increasing and the perception of people about these systems affects their adoption. Most recent work that has explored the perceived fairness of algorithmic DM [3, 22, 23] and trust from the end-user point of view [40]. However, it is equally important to investigate how developers of such algorithms perceive FATE concepts in algorithmic systems and whether their perception can change. Lee and Baykal [23] investigated the perceived fairness of DM algorithms using a website that helps groups of people divide up things like rent, credit, goods, and tasks. They found that algorithms and systems should consider social and altruistic behavior that may be difficult to incorporate in mathematical modelling. In a different work, Lee [22] conducted an online experiment using managerial decisions that required either mechanical or human skills to compare how people perceive algorithmic DM as compared to human DM. Their results revealed that the public perceives algorithmic DM as less fair and less trustworthy even when the decision requires 'human skills' [22].

Other studies have also looked into how the use of different factors and attributes in algorithmic DM might influence the public's [14, 15, 38, 39] perception of fairness. Grgić-Hlača et al. [15] focused on how the use of particular features influences public perceptions of unfairness in the context of criminal justice. They concluded that people's unfairness perception is multi-dimensional and there is a lack of consensus on which features perceived as unfair by different people. Similarly, Saxena et al. [32] found that certain attributes are not considered as fair when used in defining the outcome of a system in a certain context, suggesting that the use of features and attributes upon which decisions are made are context-dependent and can be perceived as fair or unfair accordingly [13, 22].

Researchers have also looked into related concepts, such as opacity [12] of the DM algorithmic systems and whether different transparency approaches might enhance the perceptions of fairness of those systems [10]. Transparency refers to methods for understanding of a complex model and can act as a mechanism that helps accountability [24] and appears to increase perceptions of fairness [38], although it does not make the system more desirable if that system is demonstrating non-just behavior. Explanations appear to be a popular approach primarily to support the system's decisions that are presented to the user (e.g. [29, 35, 36]). Explanations have also been shown to change how users interact with the system [12] and to help users understand the algorithm behind the system better [29]. Rader et al. [29] found that explanations, in any form, help to create awareness of how the system works and understand potential bias in the system's output. Explanations in group recommendations have been proven to improve the perception of fairness when all or the majority of group members' preferences are taken into account [35], emphasizing how fairness is subjective to each individual person. In contrast to the above, Binns et al. [3] showed that explanations might not be the best approach to help people evaluate a system's fairness, with Edwards and Veale also stressing the fact that the "right to explanation" might not be the best approach towards transparency [11].

The more the users are becoming aware of the concept of algorithmic fairness, the more they worry about potential biases in the decision as well as in the data or the algorithm interaction [6]. Wang [38] found that computer literacy has also been proven to correlate with the perception of algorithmic fairness, while Wang et al. [39] showed that the education level and gender of a person affect their perception of algorithmic fairness. Pierson also showed that students' perception of fairness changed after an hour-long lecture and discussion on algorithmic fairness [28], while Saltz et al. [31] in a study with students, where they integrated ethics in ML courses, showed that students were able to identify and articulate key ethical considerations within their ML projects. Holstein et al. [17] looked into practitioners in machine learning development and found that context-dependent educational resources, metrics, processes and tools are needed, as well as auditing tools [30], methods and procedures would help in understanding, and reducing bias in algorithmic systems and improve fairness.

Thus, in order for future algorithmic systems to become more fair, we need to ensure that future developers of such systems need to be educated and become aware about potential biases and discrimination their system may promote and the negative impact these may potentially have on the society. Hence, it is important to understand how young developers perceive FATE concepts and whether training on these topics can affect their perception. More specifically, we look into CS students' perception on FATE and whether their perception changes after attending a seminar on FATE in algorithmic systems.

3 METHODOLOGY

In order to understand how the perception of CS students on algorithmic FATE changes, we ran seminars and measured their perceptions before and after the seminar.

3.1 Seminar

The seminar was conducted in the "Software Engineering", which is a mandatory course offered to third-year undergraduate students of the CS Degree at the CS Department, University of Cyprus (N=25). The seminar was also run in the "Advanced Software Engineering", which is an elective course offered to postgraduate students of master degree in CS or Advanced Information Technologies at the same institution (N=6). Due to the COVID-19 pandemic, both courses were being offered in an online format for Fall 2020. It is important to mention that the persons who provided the seminar,

were not the regular lecturers of the students, nor lecturers at the University that the students undertake their degrees.

In this seminar participants: i) became aware of FATE issues in the development of (algorithmic) process/systems; ii) learned the core FATE concepts related to software development; iii) developed appreciation for the role that developers play in mitigating algorithmic bias and in promoting ethical practices; iv) became aware of techniques for auditing services/modules used in development. The seminar began with asking the students to fill in the pre-seminar questionnaire. Then a lecture-style introduction and basic definitions of the concepts that were going to be discussed during the course were provided. Examples from real life systems that the students were familiar with (e.g. Google Search Engine, Facebook etc.) and have exhibited behaviour that was not fair or just to some parts of the population, were brought in with the purpose of motivating a discussion between the students and the moderators. Research results were used to explain to the students the methods and approaches followed for uncovering and mitigating bias in such systems and the main stakeholders who are involved i.e. developers, users. Examples of such approaches include Auditing, Fairness Management and Explainability. Moving on, the students became aware of relevant policies - national and international - that attempt to regulate issues related to algorithms FATE e.g. GDPR, ACM Principles of Algorithmic Transparency and Accountability and National Strategies on those topics.

3.2 Materials

3.2.1 ***Scenarios.*** Participants were presented with two different scenarios in the pre-questionnaire and two different scenarios in the post-questionnaire, where algorithms made decisions that influenced humans. For these scenarios we selected contexts that our target population might be familiar with [3]. We used the same context but different story line for Scenario A in pre-questionnaire and Scenario B in post-questionnaire; respectively Scenario B in pre-questionnaire had the same context but different story line to Scenario-A in post-questionnaire. The use of corresponding scenarios in the pre- and post- questionnaires aimed to examine whether the perception of the students changes after attending a short seminar on FATE. The following scenarios were used to trigger the participants' judgement on six fairness constructs extracted from the literature on FATE [9].

- **Scenario A:** A car insurance company's premiums dynamically-priced, based on personal details and driving behaviour. This scenario was adopted from Binns et al. [3].
- **Scenario B:** A system is used to filter and rank CVs for the hiring manager, in order to assist on shortlisting the best candidates.

For each scenario, participants were asked to rate their agreement in five statements according to [9] in addition to 'Trust'. A 5-point Likert scale, ranging from '1 - Strongly Disagree' to '5 - Strongly Agree', was employed for each of the six statements:

S1 Agreement: "I agree with the decision"
S2 Understanding: "I understand the process by which the decision was made"
S3 Appropriateness of factors: "The factors considered in the decision were appropriate"
S4 Fair process: "The decision-making process was fair"
S5 Deserved outcome: "The individual deserved this outcome given their circumstances or behaviour"
S6 Trust: "I would trust this system's decision more than a human's decision"

[3]More information on the scenarios can also be found in Kasinidou et al. [19]

3.2.2 ***Pre-seminar Questionnaire.*** Participants self-assessed their knowledge on Fairness, Accountability, Transparency and Ethics in algorithmic DM systems using a 5-point Likert scale (1, Not at all - 5, Very Knowledgeable) and self-reported (*Yes/No/Other* write-in) whether they have taken *"any training/course on Fairness, Accountability and Transparency in algorithmic systems"*.

Then, they were asked, whether they would *"consider dimensions of fairness in [their] system"* and whether they would *"need to do [their] work a certain way to make a system (more) fair."* (5-point Likert scale: 1, Strongly Disagree - 5, Strongly Agree). Next participants were asked whether they would *"consider possible solutions for making a system more transparent to the user"* (5-point Likert scale: 1, Strongly Disagree - 5, Strongly Agree). Then, we presented the participants with three statements about who should be held accountable in case a system behave unfairly and asked them to indicate whether they agree with each on a Likert scale (1, Strongly Disagree - 5, Strongly Agree): *My team would be held accountable*; *The system would be held accountable*; *Neither the system nor my team would be held accountable.*

3.2.3 ***Post-seminar Questionnaire.*** After attending the seminar, participants were asked to assess their knowledge on FATE in algorithmic DM systems using a 5-point Likert scale (1, Not at all - 5, Very Knowledgeable). Participants were asked, whether they would *"consider dimensions of fairness in [their] system"* and whether they would *"need to do [their] work a certain way to make a system (more) fair."* (5-point Likert scale: 1, Strongly Disagree - 5, Strongly Agree). Then, they were asked to in case that a system behaves unfairly to indicate *"on which part of the process [they] would focus"* from the following options that were explained during the seminars: *Input*, *Output*,*Algorithm*, *Training Data*, *Third Party Constraints*,*Fairness Constraints*, *User*.

Participants then were asked whether they would *"consider possible solutions for making a system more transparent to the user"* (5-point Likert scale: 1, Strongly Disagree - 5, Strongly Agree). Then participants were asked to *"Please explain [their] answers"* to the above statement with free text. Then, we presented participants with the same three statements on accountability as in the pre-seminar questionnaire and asked them to indicate whether they agree with each on a Likert scale (1, Strongly Disagree - 5, Strongly Agree) and asked them to *"explain [their] answers"* to the above statements with free text.

3.3 Participants

54 undergraduate and postgraduate students replied to the questionnaires. Twenty-three participants did not answer both questionnaires, thus 31 respondents were considered in the analysis. Participation was voluntary and all participants provided us with written, informed consent for their data to be used. The study has

received ethical clearance by the Cyprus National Bioethics Committee. 77.4% of our respondents were male, with 77.4% in the age group of 18-24 and the rest were between 25-32. The majority of the participants (80.6%) identified themselves as undergraduates, and 88% of that group were in their third or fourth year of studies.

4 FINDINGS

4.0.1 ***Knowledge and Formal Training on FATE.*** In the pre-seminar questionnaire, we asked participants to self-report whether they *...have taken any kind of training/course on Fairness, Accountability, Transparency issues in Algorithmic Systems.* Understanding the participants' responses required us first to appreciate their previous experience with, and perceived knowledge in topics related to algorithmic fairness. 12.9% of our participants had taken some kind of training on the above topics, while the majority (77.4%) had not and the rest of the participants answered "Other".

We also asked participants to state their knowledge on the above topics before and after the seminar using a Likert-scale (1, Not at All – 5, Very Knowledgeable). Interestingly, Wilcoxon signed ranked test shows significant differences between the pre-seminar and post-seminar questionnaire replies with replies prior to the seminar being significantly lower compared to their replies after the seminar, in the above questions ($z=-3.947$, $p<0.001$); ($z=-4.008$, $p<0.001$); ($z=-3.857$, $p<0.001$) respectively. These results show that students felt more knowledgeable on FATE topics after they have attended the seminar.

4.1 Perception on FATE

4.1.1 ***Algorithmic Fairness.*** When asked whether they would consider fairness in their system most of the participants (80.7% in pre-questionnaire, 93.2% in post-seminar questionnaire) responded affirmatively (4-5), 16.1% in pre-seminar questionnaire and 6.5% in post-seminar questionnaire seemed undecided (3), and 3.2% in the pre-seminar questionnaire and none in the post-seminar questionnaire indicated that they would not consider fairness (1-2). It is important that the percentage of students who appear undecided in the pre-seminar questionnaire moved into options 4 and 5 in the after the seminar indicating that they would consider dimensions of fairness in their systems. When asked whether they would work in a certain way to make a system (more) fair the majority of the participants (80.6% in pre-seminar questionnaire, 87.1% in post-seminar questionnaire) responded affirmatively (4-5), 3.2% in both pre-seminar and post-seminar questionnaire indicated that they would not consider fairness (1-2).

Participants were asked to choose the parts of the process they think it could possibly cause unfairness in the system and explain their choices. The majority of the participants indicated that the **Algorithm** and the **Training Data** (25 out of 31) are most possible to cause unfairness in a system. Most of the participants shared the opinion that *"unfairness can be caused due to the fact that we do not have enough data for all cases of the system, the way of operation and classification of the elements by the algorithm may favor some specific cases"* (p13). Some participants also discussed that *"developers with their own bias can affect the system, the data may not have been chosen to be representative for all sectors, just as developers and users influence the system with their biases"* (p16). Other participants specifically mentioned the *"Biased dataset of training data"*(p4) and that *"it is not the way of it's implemented that is responsible but the way of data entry and the way of its training"* (p11). On the other hand, some participants discussed that *"[t]he system and its developers are responsible for the proper functioning"* (p8).

Often participants referred to the **Input** (18 out of 31) of the system as a possible cause of unfairness. Participants referred to unfairness *"[d]ue to incorrect entry, for example with the Microsoft bot, where users were responsible for logging in and learning the model"*(p21). 16 responses discussed the **User** as a possible cause of unfairness. Participant 14 pointed out that *"[u]ser's biases often get in the system"* (p14), and "[i]f the system learns from the users, then the system may learn based on wrong data that are given from the user causing wrong results" (p4). 14 participants chose **Third Party Constraint** and **Fairness Constraints**. Only few participants explained why they chose these parts. Participant 4 mentioned for the Fairness Constraints that *"[t]he operator of a system may have biased perceptions in a specific topic and set the system based on his beliefs"*. Participant 28 explained for choosing Third Party Constraints that *"third parties can with their own views indirectly influence even the writing of the algorithm"*. Finally, only 11 out of the 31 said that the **Output** could cause unfairness in a system but none specifically explained their choice.

4.1.2 ***Algorithmic Transparency.*** Participants' view on considering possible solutions for making the system more transparent did not significantly differ between the pre-seminar and the post-seminar questionnaire. The majority of the participants, in the pre-seminar questionnaire (87.1%) and post-seminar questionnaire (83.9%) agreed that (4-5 on the Likert scale) they would consider possible solutions for making the system more transparent, compared to 6.5% for pre-seminar questionnaire and 3.2% post-seminar questionnaire who indicated that they would not (1-2 on the Likert scale). We can also observe here that some students moved to the positive part of the scale after the seminar.

In the free-text explanations, participants were asked to explain ways the would use to make the system more transparent. Participants free-text responses were coded and thematically analysed [33]. Two researchers analyzed the participants' free-text responses independently to define emerging categories. We allowed multiple categories per answer and calculated the co-occurrence of themes in responses in an attempt to capture the interplay of different themes in participants' perceptions.

Six themes emerged (see Table 1). The majority of the participants (13 out of 31) suggested that they would explain to the user how the **System/Algorithm** works and other (5 out 31) they would explain the **Output**. Some participants mentioned that the *"user must know how the system works"* (p4), and that *"every user has the right to observe how they interact with the system"* (p8). Other participants specifically discussed that they would make the system more transparent by letting the user know how the data are used (p13, p23) and the procedures followed by the system (p14, p15). Two participants though mentioned that the user should not know how the algorithm works (p18, p30). The participants who chose to explain the output to promote transparency often mentioned that the user should know how the system concluded to the specific output (p12, p14, p15).

Table 1: Themes emerged from Transparency Strategies question: name, description and frequency

Category	Description	#
Explaining the System/Algorithm	*explaining the process followed by the system*	13
Explaining the Output	*explaining the output to the user; why a specific decision was made*	5
Training Data	*the dataset/information used for training the algorithm*	4
Unbiased	*algorithm/outcomes without social biases or discrimination*	4
Third party	*the impact of third parties on the system*	1
Other	*[falls outside of the established themes]*	10

Four participants suggested that they would focus on **Training Data**. They discussed that they would *"re-examine the training data and the algorithm"* (p1) and others briefly mentioned that they would explain to the users how they collected the data used to train the algorithm (p10, p13). One participant stated the use of *"more accurate and complete training data"*(p20) for training the system. Four participants discussed the development of **Unbiased** algorithms. They mentioned that *"[they] would try to limit the unfairness as much as possible"* p(2) and "avoiding injustice such as gender, skin color" (p22). Other briefly mentioned they would develop fair, ethical, transparent and without biases systems (p26, p27). The least common theme, **Third Party** received only one response and discussed the need to *"check for 3rd party influence"*(p1). Ten responses fell under the catch-all other category, which includes thoughtful responses where the participant indicated they would make the system more transparent because this is important but they do not specify how (p8, p11) or did not give any response (p21,p25).

4.1.3 ***Algorithmic Accountability.*** The last part of the study examined the concept of accountability and how the participants perceive it. Participants' view on accountability did not significantly differ between the pre-seminar and the post-seminar questionnaire. Before the seminar, most of the participants (70%) agreed with (4-5 on the Likert scale) the statement that "their team" would be held accountable, compared to 41.9% who agreed that "the system" would be held accountable and 16.1% who agreed that "neither the system nor my team" would be held accountable. The majority of the participants after the seminar (87.1%) agreed with (4-5 on the Likert scale) the statement that "their team" would be held accountable, compared to 48.9% who agreed that "the system" would be held accountable and 9.7% who agreed that "neither the system nor my team" would be held accountable. Indicating that the seminar had an impact in their perception of algorithmic Accountability.

In the post-seminar questionnaire participants were asked to explain their responses using free-text. In the free-text explanations of their choices, participants remarked that *"We implement the system so we are responsible for the system"* (p6) and that their team should be held accountable since *"[our] team may have made [the] mistake on the algorithm or choose the wrong data set"* (p1). Some participants justified their responses that the team would be held accountable with the fact that the system is not autonomous, and instead a human chooses the factors that the system uses to make decisions and the data that they use to train the system. For instance, participant 12 noted that *"The system works based on how it is programmed to do and the data which was given to it for training"*. Participants sharing this opinion felt that the humans that developed the system should be held accountable for the unfairness of the system. *"The system cannot be held accountable in any case, if someone is responsible is the development team, unless there was a wrong or malicious use of the system, wherein this case the user is responsible"* (p18). On the other hand, some participants felt that both the team that developed the system and the system itself should take the responsibility: *"I believe that both the team and the system itself will be held responsible because the team in part allowed discrimination to occur and the system can also learn in this way from the users who use it"* (p26).

4.2 Can views on FATE be changed?

4.2.1 ***Perception and Interplay of Fairness Constructs.*** Quantitative analysis was employed in order to explore whether participants' perception of each individual construct for Scenarios changed after the FATE seminar. To examine whether participants' perception changed we conducted a Wilcoxon Signed Ranks Test. Descriptive statistics for all of the variables used in the Pearson correlations are available in Table 2. The comparison of the results between Scenario A in pre-seminar questionnaire with the corresponding Scenario B in post-seminar questionnaire, indicate significant statistical differences in the responses of the students for Agreement, Understanding, Fairness of the DM process and Trust. With selections after the seminar being considerably lower compared to prior. This shows that students' perception on those issues changed after they were educated on the FATE concepts. More specifically, more students selected lower scores in the Likert scale for Agreement with the decision of the system after the seminar (z=-2.511, p=0.012), as well as Understanding of the process by which the decision was made (z=-2.941, p=0.003). Similarly, students' responses on the Fairness of the DM process show that they perceived the decision making as less fair (z=-2.424, p=0.015) and their Trust to the system's decision compared to a human also (z=-2.064, p=0.039). We did not have any statistical significant differences in the students' responses regarding the other two scenarios.

5 DISCUSSION

In this work, we seek to understand CS students' perception of FATE in algorithmic DM systems.

Perception on FATE. Consistent with [17], who looked into developers in the industry, students in our sample selected the Training Data and the Algorithm as the components they are most likely to cause unfairness in a system. They emphasised that potential biases and discrimination can exist in the training data and consequently will be learnt by the system. This shows that students

Table 2: Descriptive Statistics for the constructs

	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Pre Scenario A		Post Scenario B		Pre Scenario B		Post Scenario A	
Agreement	2.90	1.012	2.32	.909	2.61	.955	2.39	1.022
Understanding	4.23	.669	3.61	1.086	3.81	1.046	3.97	.912
Appropriateness	2.81	.873	2.48	.890	2.74	.773	2.77	1.023
Fair	2.81	.980	2.23	1.055	2.29	1.101	2.29	1.006
Deserved	2.58	1.025	2.23	.884	2.16	1.068	2.16	1.003
Trust	2.65	1.050	2.10	.870	2.16	.860	1.97	1.048

understand the need for creating more diverse data sets to be used for training machine learning algorithms integrated in DM systems. Students also mentioned that the developers of a system might unintentionally be promoting their own biases, indicating that they are aware of the human influence in the process. However, CS degrees in their majority do not provide training that can educate students on how to remain neutral towards the algorithms they are developing.

When asked about ways to make the system more transparent, it comes with no surprise that the majority of responses discussed providing explanations to the user. Explaining the system/algorithm was the most preferred strategy, although some responses opted (also) for explaining the output. This finding aligns with Rader et al.'s work [29], who found that adding explanations helped users to become more aware of how the system works and determine whether the system is biased. However, studies are still inconclusive [3, 11, 16] as to which types of explanations are suitable for which type of user (e.g. general public Vs experts). More research is needed in order to understand how and in what way explanations can be used for providing transparency to the users.

Finally, when participants asked who need to be held accountable, in a case a system they develop behaves unfairly, the majority agreed that their team should be held accountable. This may be an indication that future developers understand their responsibility of delivering "fair behaving algorithms" to their users and the possible consequences in case the system they develop is behaving unfairly to some parts of the population. Since students in our sample indicated that they lack relevant training on these topics, we understand there is a need to provide training and resources that CS students will attend when they need to.

Changing views on FATE. Our finding that CS students lack knowledge on topics related to FATE, builds on previous work [17] and reflects the need for incorporating modules and training courses in the computing-related degrees. Our findings are aligned with previous work [28], which also reported evidences of statistically significant changes in perception and attitudes of students towards algorithmic fairness and transparency just after an hour of lecture and discussion. It is important for CS students – who are likely to develop such system in the future– to ensure they are aware of concepts related to FATE in algorithmic systems. They also need to be aware that the systems they are developing have an impact (positive or negative) to the society.

Since we are expecting algorithmic systems to behave in fair and just manner, we need to educate CS student on algorithmic FATE. They need to be aware of the possible ways that biases can be introduced in a system, ways of auditing their systems prior to release, and ways of making their systems overall more transparent to their users. In addition, CS students and future developers need to develop a sense of responsibility to the users of the systems they are developing and in the society in general. CS degrees should be rationalized into incorporating algorithmic FATE related courses. Although, there are standalone seminars on this topics, courses such as Software Engineering could include modules that will provide students with the necessary knowledge on the above topics.

Limitations. While this work has provided some very interesting insights on the topics under investigation, there are some limitations that may have impacted the results. Our participants were a rather limited number (N=31), homogeneous group (young people, same race, studying at the same institution), CS students. Future work will focus on collecting, further data in both quantitative and qualitative form that may help to develop a better understanding of the perception on FATE topic and how training on these topics affects individuals' perception.

6 CONCLUDING REMARKS

We examined how computer science students perceive algorithmic FATE and whether their views and attitudes towards FATE can change after attending a relevant seminar. Our findings suggest that our participants identified the training data and the algorithm as the most likely causes of unfairness. We find that adding 'explanations of the process and output' is the most preferred strategy to make a system more transparent. After attending the seminar, participants felt more knowledgeable on FATE topics; they became more likely to consider elements of fairness in their system and believed the team developing a system should be held accountable in case the system behaves unfairly. Finally, this work showed that short seminars can make a difference in the attitude of students towards FATE in algorithmic decision making, however, a more universal top down approach is needed for educating Computer Science students on algorithmic FATE.

7 ACKNOWLEDGMENTS

This project is partially funded by the Cyprus Research and Innovation Foundation under grant EXCELLENCE/0918/0086 (DESCANT) and by the European Union's Horizon 2020 Research and Innovation Programme under agreements No. 739578 (RISE) and 810105 (CyCAT).

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. ProPublica. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, 23 Mai 2016.

[2] Paul Baker and Amanda Potts. 2013. 'Why do white people have thin lips?'Google and the perpetuation of stereotypes via auto-complete search forms. *Critical discourse studies* 10, 2 (2013), 187–204.

[3] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. "It's Reducing a Human Being to a Percentage": Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[4] Ludovico Boratto, Gianni Fenu, and Mirko Marras. 2019. The Effect of Algorithmic Bias on Recommender Systems for Massive Open Online Courses. In *Advances in Information Retrieval*, Leif Azzopardi, Benno Stein, Norbert Fuhr, Philipp Mayr, Claudia Hauff, and Djoerd Hiemstra (Eds.). Springer International Publishing, Cham, 457–472.

[5] Engin Bozdag. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology* 15, 3 (2013), 209–227.

[6] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12.

[7] Irene Y. Chen, Fredrik D. Johansson, and David Sontag. 2018. Why is My Classifier Discriminatory?. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 3543–3554.

[8] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data* 5, 2 (2017), 153–163.

[9] Jason A Colquitt and Jessica B Rodell. 2015. Measuring justice and fairness. (2015).

[10] Nicholas Diakopoulos. 2016. Accountability in algorithmic decision making. *Commun. ACM* 59, 2 (2016), 56–62.

[11] Lilian Edwards and Michael Veale. 2017. Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.* 16 (2017), 18.

[12] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[13] Ben Green and Lily Hu. 2018. The myth in the methodology: Towards a recontextualization of fairness in machine learning. In *Proceedings of the machine learning: the debates workshop*. Stockholm, Sweden.

[14] Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. 2018. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*. CA: AAAI, Palo Alto, 51–60.

[15] Nina Grgić-Hlača, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. 2018. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 903–912.

[16] Leif Hancox-Li. 2020. Robustness in Machine Learning Explanations: Does It Matter?. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 640–647.

[17] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–16.

[18] Nicolas Huck. 2019. Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research* 278, 1 (2019), 330–342.

[19] Maria Kasinidou, Styliani Kleanthous, Pınar Barlas, and Jahna Otterbacher. 2021. I agree with the decision, but they didn't deserve this: Future Developers' Perception of Fairness in Algorithmic Decisions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 690–700.

[20] Kyriakos Kyriakou, Styliani Kleanthous, Jahna Otterbacher, and George A Papadopoulos. 2020. Emotion-based Stereotypes in Image Analysis Services. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 252–259.

[21] Preethi Lahoti, Krishna P Gummadi, and Gerhard Weikum. 2019. ifair: Learning individually fair data representations for algorithmic decision making. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 1334–1345.

[22] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 2053951718756684.

[23] Min Kyung Lee and Su Baykal. 2017. Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1035–1048.

[24] Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology* 31, 4 (2018), 611–627.

[25] Frank Marcinkowski, Kimon Kieslich, Christopher Starke, and Marco Lünich. 2020. Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 122–130.

[26] Dena F Mujtaba and Nihar R Mahapatra. 2019. Ethical considerations in ai-based recruitment. In *2019 IEEE International Symposium on Technology and Society (ISTAS)*. IEEE, 1–7.

[27] Jahna Otterbacher, Jo Bates, and Paul Clough. 2017. Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6620–6631.

[28] Emma Pierson. 2017. Demographics and discussion influence views on algorithmic fairness. arXiv:1712.09124 [cs.CY]

[29] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13.

[30] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 33–44.

[31] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating Ethics within Machine Learning Courses. *ACM Trans. Comput. Educ.* 19, 4, Article 32 (Aug. 2019), 26 pages.

[32] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. 2019. How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 99–106.

[33] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.

[34] Kjerstin Thorson, Kelley Cotter, Mel Medeiros, and Chankyung Pak. 2019. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society* 0, 0 (2019), 1–18.

[35] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, Viet Man Le, Ralph Samer, and Martin Stettinger. 2019. Towards Social Choice-Based Explanations in Group Recommender Systems. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP '19)*. Association for Computing Machinery, New York, NY, USA, 13–21.

[36] Chun-Hua Tsai and Peter Brusilovsky. 2019. Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP '19)*. Association for Computing Machinery, New York, NY, USA, 22–30.

[37] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[38] AJ Wang. 2018. Procedural Justice and Risk-Assessment Algorithms. *Available at SSRN 3170136* (2018).

[39] Ruotong Wang, F. Maxwell Harper, and Haiyi Zhu. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[40] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14.