Crafting Disability Fairness Learning in Data Science: A Student-Centric Pedagogical Approach

Pax Newman
Western Washington University
newmanp@wwu.edu

Tyanin Opdahl
Western Washington University
opdahlt@wwu.edu

Yudong Liu
Western Washington University
liuy2@wwu.edu

Scott Wehrwein
Western Washington University
wehrwes@wwu.edu

Yasmine N. Elglaly
Western Washington University
elglaly@wwu.edu

ABSTRACT

Ensuring the fairness of machine learning (ML) systems for individuals with disabilities is crucial. Proactive measures are required to identify and mitigate biases in data and models, thereby preventing potential harm or bias against people with disabilities. While previous research on ML fairness education primarily concentrated on gender and race fairness, the domain of disability fairness has received comparatively little attention. Addressing this gap, we adopted a student-centric approach to craft a disability fairness teaching intervention. A focus group of students experienced in ML and accessible computing underscored the significance of engagement and scaffolding strategies for effectively learning intricate topics. Consequently, we crafted a disability fairness hands-on programming assignment that delves into uncovering disability bias with a lens that takes intersectionality into account. The assignment was tailored for an introductory undergraduate data science (DS) course. We employed reflective questions and surveys to gauge the effectiveness of our approach. The findings indicate the success of our approach in promoting a deeper understanding of disability fairness within the context of DS education.

CCS CONCEPTS

• Social and professional topics → Computer science education; • Human-centered computing → Accessibility.

KEYWORDS

Fairness, disability, data science, machine learning, CS education

ACM Reference Format:
Pax Newman, Tyanin Opdahl, Yudong Liu, Scott Wehrwein, and Yasmine N. Elglaly. 2024. Crafting Disability Fairness Learning in Data Science: A Student-Centric Pedagogical Approach. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024), March 20–23, 2024, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3626252.3630815

SIGCSE 2024, March 20–23, 2024, Portland, OR, USA

ACM ISBN 979-8-4007-0423-9/24/03…$15.00
https://doi.org/10.1145/3626252.3630815

1 INTRODUCTION

As Machine Learning (ML) and Data Science (DS) systems became more prevalent, the discussion around the fairness of these systems to individuals with disabilities gained significant traction [22, 38, 39]. This is due to the revelation of disability bias within smart systems and their potential to adversely affect individuals with disabilities [14]. Therefore, it is imperative for computing students to grasp the principles of fairness in smart systems, such as identifying and mitigating bias, concerning individuals with disabilities, a concept termed “disability fairness” in this paper. Recent work in CS education has addressed the lack of fairness teaching in ML courses using projects, lectures, and hands-on activities [12, 40]. However, prior research highlighted that more engaging pedagogy that focuses specifically on disability fairness is needed [12]. In line with this, we adopted a student-centered approach to design and evaluate a pedagogical path for disability fairness education.

We conducted a focus group of CS students who completed ML or accessibility courses. The focus group examined a sample disability fairness assignment and brainstormed new assignment ideas. The findings of the focus group study revealed two themes: engagement factors and scaffolding strategies, which students deemed essential for an effective disability fairness assignment.Building on the focus group insights, we developed a disability fairness assignment, and implemented it in an introductory undergraduate-level DS course. The assignment encompassed a programming lab and a set of reflective questions. This assignment centered on uncovering disability bias in BERT [4, 11, 19], compelling students to explore the implications of bias on individuals with disabilities. We qualitatively analyzed the students’ answers to the final set of reflective questions, and ran a pre- and a post-survey. We found that students became more aware of the impacts of disability bias in ML systems and gained better understanding of how to uncover biases in ML systems towards individuals with disabilities. The study was approved by our institution’s IRB office. The contributions of our work are: 1) Presenting a student-centered approach to inform disability fairness pedagogy; 2) Designing a technical assignment on disability fairness; 3) Making the programming assignment and its rubrics publicly available [1]; and 4) Offering insights on pedagogy design factors that can empower educators to expand on ethics teaching in ML courses.

[1] https://github.com/thekindlab/MLFairnessEducation

2 RELATED WORK

2.1 Importance of Fair ML and DS Systems

Fairness, a concept broadly defined as the impartial treatment of individuals and of demographic groups, has gained increasing recognition within the ML and DS communities [6, 8, 9, 25, 27, 34, 41, 43]. Most ML and DS systems rely on vast amounts of data for training their algorithms and frameworks. When training data contains biases, these biases are learnt and subsequently reflected in the algorithm's prediction [28]. Biases may also stem from the design choices of algorithms [16, 21, 28, 42]. Consequently, the outcomes produced by these systems can influence real-world applications and impact users' decisions, leading to a feedback loop of more biased data for training future algorithms. In response, researchers and engineers have diligently worked to mitigate biases by targeting bias concerns throughout the entire data processing pipeline [9].

Addressing fairness issues concerning individuals with disabilities has drawn increasing attention. While creating inclusive and accessible smart systems has become an important goal for the ML and DS communities [2, 7, 18], certain challenges, such as a lack of relevant data and nuanced definitions of disabilities, often result in systems inadequately including disability in data sourcing, model building, and testing processes [38, 39]. In response to these challenges, researchers have started to develop viable solutions, such as envisioning an online platform to facilitate data contributions from disability communities [33], and outlining a disability justice approach, centered on prioritizing the experiences of disabled individuals and addressing the underlying structures and norms that contribute to algorithmic bias [37].

2.2 Fairness Education in ML and DS Courses

As fairness issues become increasingly prominent in the domains of ML and DS, fairness education is receiving more attention and emphasis. It has started to be integrated into general CS curricula, particularly into courses focused on CS ethics [13, 15, 20, 24]. These CS ethics topics are either taught as standalone courses [13] or integrated into ML-related technical courses [5, 12, 15, 24]. CS educators are leaning more towards the latter approach [10, 12]. However, it can be observed that most existing efforts concentrate on ML and AI courses, with very limited resources on fairness in DS courses. This gap serves as one of the motivations for our work to address and bridge this disparity.

When it comes to teaching about fairness for individuals with disabilities, the current focus is largely centered around accessibility [3, 40], including a specific focus on topics like inclusive design [29, 30]. As mentioned earlier, efforts in addressing fairness issues for people with disabilities, particularly within the data processing pipeline, including data sourcing, modeling, and testing aspects, are still in their early stages. In light of this, we believe that incorporating fairness education, particularly for individuals with disabilities, into the DS curriculum is highly relevant and timely.

3 FOCUS GROUP

We conducted a focus group study to learn how to better design an engaging assignment on disability fairness. The study comprised two sessions: a review of a sample assignment and a brainstorming session for new assignment ideas. Each session lasted approximately 55 minutes. The sessions were conducted over Zoom, with one researcher leading the discussion and two others taking notes. The sessions were recorded for transcription and analysis.

3.1 Focus Group Participants

Six CS students from Western Washington University participated in the study, comprising 3 graduate students and 3 undergraduates. Each student had completed at least one course on ML, accessible computing, or both. Three participants were male, 2 were female, and 1 was non-binary. One participant identified as disabled. Only 4 of the participants opted to continue on to the brainstorm session due to time constraints. Each participant received a 25-dollar Amazon e-giftcard.

3.2 Session 1: Discuss Sample Assignment

The participants were given an assignment write-up to discuss and provide feedback on. The assignment highlights how some ML systems incorrectly predict comments containing certain identities (e.g., "disabled") as highly toxic, even when these comments are not inherently toxic (e.g., "I am a disabled woman"). The assignment specifically focuses on educating students about discrimination against people with disabilities. The assignment involves evaluating a pre-trained language model to identify learned biases towards disabled individuals. The assignment also explains the steps in model production that contribute to these biases and explore potential intervention strategies.

The researcher leading the focus group posed a series of questions to the participants. Each participant was asked to share their opinion in a round-robin fashion, and there was dedicated time at the end for free discussion. The questions addressed various aspects of the assignment, including the participants' initial impressions of the topic, the relevance of the topic to their personal interests, their thoughts on the wording and clarity of the assignment, the intellectual stimulation provided by the topic, the best features of the assignment, areas for improvement, the scope of the assignment, and their opinions on the inclusion of intersectionality material.

3.3 Session 2: Brainstorm Assignment Ideas

In the brainstorming session, the participants were given individual time to write down four ideas for an assignment focused on ML and its biases against people with disabilities. Afterward, the participants engaged in a voting process to select their favorite idea from each other participant. The final four ideas were then discussed collectively, with the lead researcher facilitating the discussion and posing relevant questions: What aspects do you appreciate about these top four assignment ideas? Do you believe that the topics of machine learning and fairness for people with disabilities are adequately addressed in these ideas? If not, what additions or modifications are needed to cover these areas? Lastly, how can we further enhance the appeal and engagement of these top four assignment ideas for students? These questions aimed to gather feedback on the strengths of the ideas, ensure comprehensive coverage of relevant subjects, and explore possibilities for making the assignments more captivating and appealing to students.

3.4 Focus Group Analysis and Findings

We used thematic analysis with inductive reasoning to qualitatively analyze the data we collected from the focus group sessions. We identified 2 main themes, which are engagement factors and scaffolding a ML disability fairness assignment.

3.4.1 Engagement Factors in a ML-Disability Fairness Assignment. The participants highlighted several key factors that contribute to the engagement of a disability fairness assignment.

Firstly, the use of modern and industry-relevant technologies such as Python, ML, and BERT was recognized as intellectually stimulating. These tools provide students with practical and up-to-date applications, enhancing their interest and relevance in the assignment. Secondly, the participants emphasized the importance of incorporating real-life examples and connections. They suggested exploring how BERT reacts to cover letters from individuals with and without disabilities, which can shed light on ableism in workplaces and identify factors that contribute to an ableist environment. Additionally, participants proposed an interactive approach where students could send out simulated resumes and analyze the responses, adding a practical dimension to the assignment. They also recommended making connections with recent events to further enhance the intellectual stimulation and relevance of the assignment. Thirdly, the inclusion of ableism and intersectionality was considered valuable by the participants. They recognized that intersectionality helps illustrate how ableism intersects with other forms of discrimination, providing a comprehensive understanding of the topic. Participants found intersectionality engaging because it relates to their own multiple identities and allows them to run the model on themselves, fostering a deeper connection to the subject matter. As one participant said: *"I feel like a lot of students ... have intersectional identities. And I feel like a lot of the students that will be doing [such an] assignment will be able to relate to these identities."*

Lastly, the participants expressed a preference for self-led investigations and personal examples. They found that synthesizing information and relating it to their personal lives enhanced engagement and facilitated self-reflection, making the assignment more impactful and meaningful. Participants recommended to include reflection questions that are specifically aligned with the assignment's goals and expectations. This ensures that students can engage thoughtfully with the topics and integrate their reflections seamlessly into their work.

Overall, the participants emphasized the significance of utilizing modern technology, incorporating real-life examples, exploring intersectionality, and encouraging reflection to create an engaging and intellectually stimulating disability fairness assignment.

3.4.2 Scaffolding a ML-Disability Fairness Assignment . The participants identified key assignment elements that can effectively balance the complexity of an assignment involving two intricate topics, such as ML and disability. They emphasized the importance of showcasing the significance of the assignment, allowing students to understand the purpose and relevance of their work. Furthermore, participants highlighted the role of clearly articulating the assignment *learning objectives* in setting clear expectations for students and providing a well-defined scope for the assignment.

To support students in completing the assignment, the participants emphasized the need for thorough explanations of any programming libraries and the functions of the models employed. As one participant mentioned: "*My first impression was like, Oh, that's a lot of like, technical terms right off the bat.* " To cater to the diverse range of student experience levels, the participants suggested incorporating clickable links, enabling students to access additional resources and information at their own pace.

Additionally, the participants stressed the need for a well-structured assignment. They recognized that a strong and organized structure helps students grasp the tasks and expectations more effectively. One suggestion was to break the assignment into smaller, manageable tasks or to-do items, allowing students to refine their understanding and progress gradually. Furthermore, the participants recommended that this type of advanced assignments should be *preceded by a conceptually easier assignment* that introduces students to the same topic. Lectures were also considered valuable in providing the necessary foundation for students to tackle the assignment successfully. Finally, the participants stressed the need for perfect grammar and spelling to maintain the professionalism and clarity of the work. Furthermore, acronyms used in the assignment should be clearly stated in full before a subsequent use, ensuring that students can fully comprehend the content without confusion.

By considering these identified assignment elements, educators can design assignments that effectively engage students in the complex topics of ML and disability while providing appropriate support and guidance.

4 DS COURSE AND INTERVENTION

The findings of the focus group highlighted the importance of scaffolding and provided engagement factors for a ML disability fairness assignment. We integrated engagement factors into two specific facets of the disability fairness assignment. Firstly, the assignment utilized datasets exemplifying intersectionality [19]. These datasets show, not only disability bias, but also gender and race biases. Secondly, we used reflection questions to engage students with self-reflection. We also adopted a design tool, The Tarot Cards of Tech [36], to further engage students with self-reflection.

We implemented the disability fairness intervention in a 4-credit introductory DS course that is required for DS majors and elective for CS majors. It took place during Winter 2023 and spanned over 10 weeks in a quarter-based system. The course was conducted in-person and covered topics such as quantitative analysis and evaluation of data, fundamentals of ML experimentation, and applications of Natural Language Processing (NLP).

4.1 Scaffolding the Disability Fairness Concept

Before introducing students to the concept of disability fairness, they were first introduced to data ethics using an active learning assignment followed by studying a research paper that describes the impact of a DS application on groups of users.

Introducing Data Ethics. In Week 4 of the course, students were instructed to download their personal data, e.g., from Facebook or Google, and conduct an analysis of the information they discovered. Students were asked to write a one-page paper discussing their perception of the data's accuracy, without any obligation to disclose personal information. The assignment was followed by in-class discussion. An in-class poll was conducted to gauge students' experiences with the personal data they had collected. Additionally, *small group discussions* were facilitated to encourage students to delve into the *ethical considerations surrounding data collection* and reflect on their own reactions to the findings.

Realization of the Ethical Implications of Big Data. In Week 6, students were assigned a news article and an academic paper to read, both focusing on the emotional impact of Meta's news feed on the masses [1, 26]. They were required to write a one-page paper summarizing their evaluation of the paper from an ethics perspective and expressing their personal thoughts on the topic. In the following lecture, the topic of "Generalization within Machine Learning" was discussed. The lecture covered the effects of model bias, variance, and irreducible error, providing insights into the broader aspects of ML and its implications.

4.2 Disability Fairness Assignment

In Weeks 7 and 8, students worked on a hands-on programming assignment to measure disability bias within BERT. BERT is a masked language model architecture widely used in NLP applications like search engines, text summarization, and translation [11]. This assignment required students to engage in practical coding and apply their knowledge to assess bias in the model. The assignment followed the process of BERT evaluation and the datasets provided by Hassan et al [19]. Students used 5 datasets to investigate biases in BERT by having BERT predict words in sentences. They hypothesized that sentences with disability, gender, and/or race-related references might have more negative sentiment. After collecting and filtering the data, students performed sentiment analysis using VADER via the nltk Python package [23], assigning each sentence a polarity score ranging from [-1.0, 1.0]. They created a plot to illustrate the polarity scores and provided their interpretations.

4.3 Disability Fairness Assignment's Reflection

The disability fairness assignment contained 5 reflection questions that aimed to give students a chance to think critically about the assignment's results as well as the origins and impacts of biases. To encourage the students to think deeply about the impacts of models on people, we used a design tool called *The Tarot Cards of Tech* [36] in the first reflection question. This tool offers various "tarot cards" containing prompts which ask the reader to consider how their technology could affect society. For example, the "Smash Hit" card asks readers "What happens when 100 million people use your product?" It also gives follow-up questions asking what mass usage of their product might affect, how a community might change if 80% of it used their product, and how habits or norms might change. Similarly, "The Service Dog" prompts readers to consider what sort of impact their product could make if it was entirely dedicated towards empowering an under-served population, alongside similarly themed follow-up questions.

Q1. Please look through the cards on Tarot Cards of Tech. Pick any two (such as "The Smash Hit" and "The Service Dog") and write about how they each might apply to BERT.

Q2. With the work we've done now, where do you think the biases in BERT come from? What caused these biases to form?

Q3. Now that you've seen examples of bias in an NLP model, what kind of biases or ethical problems do you think other ML models or AI applications could have? For example other language models such as the one used in ChatGPT [31], or other models entirely such as those relating to image recognition/generation, social media analysis, speech recognition, etc.

Q4. Based on your answer from Q2, how might you show that these biases exist in the model/application?

Q5. Write a 1-2 paragraph reflection on what you've learned. Has your view on the ethics of ML models changed? What technical knowledge have you gained?

We followed inductive reasoning to analyze student reflections. Three researchers iteratively created a 12-code codebook. Two researchers employed this codebook to qualitatively code responses, reviewed by the third researcher. The three researchers then grouped codes into three main themes.

5 SURVEY INSTRUMENT

Students were given a pre- and post- survey about disability awareness. The survey consisted of demographics questions and 4 questions relating to the students knowledge and beliefs about disability bias (see Table1). Additionally, the survey contained an open-ended question on the definition of disability bias. *"What do you think "disability bias" means for machine learning systems? If you do not know, it is ok to say so."* We analyzed the aggregated Likert survey data by coding the students' survey responses from 1 to 5 and applying the Wilcoxon Signed-Rank Test at a significance level of $\alpha = 0.05$. Additionally, we conducted qualitative coding on the responses to the definition of "disability bias" question.

5.1 Participants

Out of the 33 students taking the class, 18 students completed the pre-survey (10 male, 6 female, 2 prefer not to say) and 26 students completed the post-survey (16 male, 8 female, 2 prefer not to say). The majority of participants had a major in CS or DS. One student had a major in Marketing, and another in Applied Mathematics. Most students had been in their respective majors for a year or less. The mean of participants' age was 22.8, and standard deviation was 4.6. Three students self-identified as disabled.

6 FINDINGS

6.1 Reflection Questions Findings

Thirty students answered the first, second, and fifth questions, and 28 students answered the third and fourth questions. In total, we collected 146 reflection question responses. Due to a misunderstanding of the first question however, we had to remove one response to that question, leaving 145 valid responses.

6.1.1 *Students' Perception of BERT Bias Root Causes.*

During the analysis of the origin of bias in BERT, students frequently attributed it to human bias. Their overarching conclusion suggested that BERT predominantly mirrors pre-existing human biases. According to the students, BERT's language usage is notably influenced by current events and media trends. The content highlighted in the media tends to be more frequently incorporated, resulting in the adaptation of NLP models to better align with the provided dataset. A student

Disability Bias Question	**Pre-survey Mean and SD**	**Post-survey Mean and SD**	**p-Value**
I am aware of disability biases in machine learning (ML) systems.	Mean=2.4, SD=1.2	Mean=3.4, SD=1.17	**0.02***
I understand how to uncover biases in ML systems towards people with disabilities.	Mean=1.8, SD=0.7	Mean=2.9, SD=1.09	**0.008***
This course helped me realize the implications of ML systems with biases towards people with disabilities.	Mean=2.16, SD=1.15	Mean=3.46, SD=1.06	**0.002***
As a data scientist, it is my professional responsibility to create ML systems without biases.	Mean=4.16, SD=0.98	Mean=4.4, SD=0.75	0.09

Table 1: Survey questions on disability bias and their Wilcoxon Signed-Rank Test results, with * indicating statistical significance.

explained: "*BERT is biased whenever race, gender, or disability is referenced ... BERT was trained on the English Wikipedia which is written by humans. All humans have bias in one way or another. Since BERT learned from biased information, BERT predict[s] biased words.*" (Student 16). Additionally, some students emphasized that biases in ML models and AI applications stem from the inherent biases of their creators. The prevailing sentiment was that AI reflects the biases conveyed by humans, whether these biases originate from dataset bias, dataset collection, or the creators.

6.1.2 Impact of ML Bias on People with Disabilities. The general consensus students came to was that NLP has a negative bias towards people with disabilities and other minority groups. Throughout their work in the class, students recognized how extensive usage of large amounts of data sourced from specific or limited demographics can inadvertently introduce bias into NLP systems.

"*We saw this [bias] as we looked at the sentiments of the words generated. In sentences regarding race, gender, or disabilities, we saw a negative sentiment. This demonstrates the biases that humans tend to have towards marginalized groups.*" (Student 12)

Moreover, the students speculated that if these latent biases remain unresolved within NLP systems, they could contribute to a perpetuation of negative biases in society. They elaborated on how these seemingly unintentional biases could potentially lead to tangible ramifications for minority groups, specifically marginalizing their voices within the larger society. This notion was often exemplified by instances such as image recognition models failing to identify non-white individuals, or NLP systems failing to recognize names with cultural significance.

Students have recognized both positive and negative impacts of NLP systems on individuals with disabilities. The predominant focus of the students' discussions revolved around the notable negative bias exhibited by NLP systems towards individuals with disabilities. As one student remarked, "*Words like 'guilty' and 'uncomfortable' pop up a considerable number of times and I believe this stems from the negative stigma that disabilities have, and I think that is amplified in American subconsciousness.*" (Student 29)

Nevertheless, a few students acknowledged some beneficial aspects for individuals with disabilities, such as the use of AI predictive language to compose emails. Another positive aspect highlighted was the potential for the general public to gain a heightened awareness of societal biases towards minority groups through direct observation of biases present within NLP systems.

6.1.3 Involving People with Disabilities in Creating and Reviewing Datasets. As students reflected on the multifaceted issue of ML bias and considered potential strategies for its prevention and detection, they came up with two primary suggestions that involve individuals with disabilities and those with underrepresented identities. One significant emphasis that emerged from students' reflections was the imperative of constructing inclusive datasets. The students underscored the critical importance of acquiring data from individuals with disabilities and underrepresented groups, thereby ensuring that a diverse array of demographics is adequately and equitably represented within these datasets. Their suggestions align with the notion that the composition of training data has a direct impact on the bias that ML models may subsequently exhibit.

"*First, a dataset can be obtained that includes information regarding underrepresented people from different backgrounds, perhaps representing different ethnic groups, gender identities, sexual orientations, and disabilities, and running those datasets against a model, in order to see what it might predict.*" (Student 6)

The second approach proposed by the students aims to mitigate bias through the active participation of individuals with disabilities and other underrepresented identities in the review of datasets. As noted by one student: "*While it can be hard to detect, the first step to take is to have multiple people review the training set and the application, this is best if there is a diverse set of people to review it.*" (Student 12). Students emphasized the inherent complexity associated with bias detection. They highlighted the crucial step of subjecting both the training dataset and the application to the scrutiny of multiple individuals with diverse identities. By inviting these individuals to directly assess datasets, the approach taps into their nuanced insights, enabling the identification of potential biases that might otherwise go unnoticed.

Overall, the students collectively demonstrated a good understanding of biases within ML models and datasets. The students also recognized both the potential benefits and potential harm of ML for people with disabilities. Students suggested effective approaches to mitigate bias and involve people with disabilities in the development of ML systems. It is worth noting that not every student incorporated all these facets in their reflections. Therefore, we recommend instructors to conduct follow-up discussions in class, allowing students to draw insights from one another's perspectives.

6.2 Survey Results

Using the Wilcoxon Signed-Rank Test, we found that 3 of the disability bias questions (see Table 1) had a statistically significant difference between the students' answer before and after taking the class. The questions with a statistically significant difference highlight that students became more aware of disability biases in ML, methods to uncover these biases, and that the course helped them understand the implications of ML for people with disabilities. The only question that did not show a statistically significant shift pertained to students' belief they had a responsibility to create ML systems without bias. This lack of statistically significant change can be attributed to the high scores for both the pre- and post-survey question.

In the pre-survey, when asked about disability bias in the open-ended question, nearly all participants indicated their lack of familiarity with the concept within ML. Only a few students attempted to provide explanations, although these often missed the essence of the concept. For instance, one student stated, "*I think disability bias is when someone puts in the option of nothing and then the machine has to figure out it's function without that required data.*" In contrast, the post-survey results showed a notable improvement in students' comprehension of disability bias. Only 3 out out the 26 participants failed to answer the question while the remaining students provided an adequate explanation for the term, such as: "*This means that machine learning and AI systems are created and written without including people who have a disability. This makes AI and ML have negative bias against people with disabilities.*"

This shift from minimal awareness to a more informed understanding is indicative of the educational impact of the intervention.

7 DISCUSSION

7.1 Engagement Factors and Scaffolding Facilitated Disability Fairness Learning

We adopted a focus group-based approach to inform the design of educational interventions aimed at teaching ML fairness for individuals with disabilities. The outcomes of the focus group sessions revealed interesting engagement factors such as the importance of self-reflection and how the individual experience and the intersectionality can increase awareness regarding the harm of bias, including ableism. This stems from the idea that students, upon relating to a marginalized identity, can better grasp the harm perpetuated by such biases. Another important outcome of the focus group discussions pertains to the significance of scaffolding when implementing educational interventions that target intricate subjects, such as fairness in ML. Prior to introducing the disability fairness assignment, students were familiarized with data ethics and its potential impact on user groups. They also gained practical DS skills including pre-processing, model execution, and NLP concepts like sentiment analysis and tokenizing.

The combined analysis of students' responses to reflection questions and survey outcomes showcases an enhanced understanding of the disability bias concept. Moreover, students' fervent advocacy for the active engagement of individuals with disabilities and underrepresented identities in the formulation and revision of datasets and their corresponding applications strongly implies an appreciation for the significance of ML fairness. This recommendation resonates with the broader ethical discourse in the field, emphasizing the responsibility to create ML systems that not only perform efficiently but also uphold fairness and inclusivity [17, 33].

7.2 Instructor's Perspective on the Feasibility of ML Fairness In-Class Discussions

Through the interventions and classroom discussions, the instructor observed that the majority of students acquired practical skills in applying text normalization and utilizing black box NLP models. They also comprehended the inherent limitations of such models and acknowledged the potential biases stemming from their training data. The instructor acted as a facilitator and played the devil's advocate to encourage critical thinking. No significant conflicts or instances of offense among students were noted. The instructor attributed the ease of such discussions to the prevailing homogeneity of viewpoints within the class. Engaging in conversations about disability bias did not elicit discomfort for the instructor; however, solutions for addressing these biases were acknowledged as a more intricate challenge. The instructor also highlighted that, while certain examples exhibit clear-cut biases, subtler issues pose greater complexity, demanding nuanced resolutions. Occasionally, some students questioned the relevance of these discussions, yet, on the whole, the instructor received no negative feedback about the assignments, with most students expressing their positive sentiments.

7.3 Covering Fairness Topics in Technical Courses Calls for Creative Pedagogy

Integrating accessibility and disability education into ML courses is valued by the CS Education community [40]. However, designing effective teaching interventions has proven challenging [12]. This study leveraged focus groups to shape the design of the interventions presented. Findings demonstrate the effectiveness of this approach in achieving learning outcomes, particularly those concerning disability fairness. However, the focus group discussions unveiled additional engagement factors that were not implemented in this study such as immersing learners in understanding the bias's impact on their personal data. Additionally, real-world text-based accessibility and disability datasets have become available [32, 35]. In the future, we plan to design novel teaching interventions that incorporate these captivating elements, aiming for a more impactful learning experience.

8 CONCLUSION

We introduced a student-centered pedagogical approach for effective disability fairness learning, informed by focus groups. Through the focus groups, students provided concrete feedback and insightful recommendations on creating a pathway to unpack the complexity of simultaneously learning the conceptual and technical skills related to disability fairness. Our approach, characterized by its scaffolding and engagement strategies, yielded positive outcomes in a DS course. The insights gained from this study serve as a starting point for further research, offering pathways to expand on ethics topics education in technical courses.

9 ACKNOWLEDGMENT

This material is based upon work supported by the SIGCSE Special Projects Grant award # 56004A.

REFERENCES

[1] 2014. *"Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry"*. Retrieved August 1, 2023 from https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html

[2] Patricia Acosta-Vargas, Luis Antonio Salvador-Ullauri, and Sergio Luján-Mora. 2019. A heuristic method to evaluate web accessibility for users with low vision. *IEEE Access* 7 (2019), 125634–125648.

[3] Catherine Baker, Yasmine Elglaly, and Kristen Shinohara. 2020. A Systematic Analysis of Accessibility in Computing Education Research. 107–113. https://doi.org/10.1145/3328778.3366843

[4] Marion Bartl, Malvina Nissim, and Albert Gatt. 2020. Unmasking contextual stereotypes: Measuring and mitigating BERT's gender bias. *arXiv preprint arXiv:2010.14534* (2020).

[5] Benjamin S. Baumer, Randi L. Garcia, Albert Y. Kim, Katherine M. Kinnaird, and Miles Q. Ott. 2022. Integrating Data Science Ethics Into an Undergraduate Major: A Case Study. *Journal of Statistics and Data Science Education* 30, 1 (2022), 15–28. https://doi.org/10.1080/26939169.2022.2038041 arXiv:https://doi.org/10.1080/26939169.2022.2038041

[6] Mariano G Beiró and Kyriaki Kalimeri. 2022. Fairness in vulnerable attribute prediction on social media. *Data Mining and Knowledge Discovery* 36, 6 (2022), 2194–2213.

[7] John Bricout, Paul MA Baker, Nathan W Moon, and Bonita Sharma. 2021. Exploring the smart future of participation: Community, inclusivity, and people with disabilities. *International Journal of E-Planning Research (IJEPR)* 10, 2 (2021), 94–108.

[8] Francois Buet-Golfouse and Islam Utyagulov. 2022. Towards fair unsupervised learning. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1399–1409.

[9] Barbara Catania, Giovanna Guerrini, and Chiara Accinelli. 2023. Fairness & friends in the data science era. *AI & SOCIETY* 38, 2 (2023), 721–731.

[10] Amanda Coston, Alan Mishler, Edward H. Kennedy, and Alexandra Chouldechova. 2020. Counterfactual Risk Assessments, Evaluation, and Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 582–593. https://doi.org/10.1145/3351095.3372851

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://doi.org/10.48550/ARXIV.1810.04805

[12] Samantha Jane Dobesh, Tyler Miller, Pax Newman, Yudong Liu, and Yasmine N. Elglaly. 2023. Towards Machine Learning Fairness Education in a Natural Language Processing Course. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto ON, Canada) *(SIGCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 312–318. https://doi.org/10.1145/3545945.3569802

[13] Casey Fiesler, Natalie Garrett, and Nathan Beard. 2020. What do we teach when we teach tech ethics? a syllabi analysis. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 289–295.

[14] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. "I Wouldn't Say Offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 205–216. https://doi.org/10.1145/3593013.3593989

[15] Natalie Garrett, Nathan Beard, and Casey Fiesler. 2020. More Than" If Time Allows" The Role of Ethics in AI Education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 272–278.

[16] Benedetta Giovanola and Simona Tiribelli. 2023. Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & society* 38, 2 (2023), 549–563.

[17] Alexandra Reeve Givens and Meredith Ringel Morris. 2020. Centering Disability Perspectives in Algorithmic Fairness, Accountability, Transparency. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 684. https://doi.org/10.1145/3351095.3375686

[18] Paula Hall and Debbie Ellis. 2023. A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review* (2023).

[19] Saad Hassan, Matt Huenerfauth, and Cecilia Ovesdotter Alm. 2021. Unpacking the Interdependent Systems of Discrimination: Ableist Bias in NLP Systems through an Intersectional Lens. *CoRR* abs/2110.00521 (2021). arXiv:2110.00521 https://arxiv.org/abs/2110.00521

[20] Diane Horton, Sheila A. McIlraith, Nina Wang, Maryam Majedi, Emma McClure, and Benjamin Wald. 2022. Embedding Ethics in Computer Science Courses: Does It Work?. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 481–487. https://doi.org/10.1145/3478431.3499407

[21] Ayanna Howard, Cha Zhang, and Eric Horvitz. 2017. Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems. In *2017 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. IEEE, 1–7.

[22] Ben Hutchinson, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyl. 2020. Social Biases in NLP Models as Barriers for Persons with Disabilities. (2020). https://doi.org/10.48550/ARXIV.2005.00813

[23] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

[24] Sheikh Rabiul Islam, Ingrid Russell, William Eberle, and Darina Dicheva. 2022. Incorporating the Concepts of Fairness and Bias into an Undergraduate Computer Science Course to Promote Fair Automated Decision Systems. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 1075. https://doi.org/10.1145/3478432.3499043

[25] Brittany Johnson and Yuriy Brun. 2022. Fairkit-learn: a fairness evaluation and comparison toolkit. In *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Companion Proceedings*. 70–74.

[26] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National academy of Sciences of the United States of America* 111, 24 (2014), 8788.

[27] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. 2022. Fairness in recommendation: A survey. *arXiv preprint arXiv:2205.13619* (2022).

[28] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–35.

[29] Alannah Oleson, Christopher Mendez, Zoe Steine-Hanson, Claudia Hilderbrand, Christopher Perdriau, Margaret Burnett, and Amy J. Ko. 2018. Pedagogical Content Knowledge for Teaching Inclusive Design. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. Association for Computing Machinery, New York, NY, USA, 69–77. https://doi.org/10.1145/3230977.3230998

[30] Alannah Oleson, Meron Solomon, Christopher Perdriau, and Amy Ko. 2023. Teaching Inclusive Design Skills with the CIDER Assumption Elicitation Technique. *ACM Trans. Comput.-Hum. Interact.* 30, 1, Article 6 (mar 2023), 49 pages. https://doi.org/10.1145/3549074

[31] OpenAI. 2020. Language Models are Few-Shot Learners. https://cdn.openai.com/better-language-models/language_models_are_few_shot_learners.pdf.

[32] Brandon Palonis, Samantha Jane Dobesh, Selah Bellscheidt, Mohamed Wiem Mkaouer, Yudong Liu, and Yasmine N Elglaly. 2023. Large-Scale Anonymized Text-based Disability Discourse Dataset. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–5.

[33] Joon Sung Park, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. 2021. Designing an Online Infrastructure for Collecting AI Data From People With Disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 52–63. https://doi.org/10.1145/3442188.3445870

[34] Drago Plecko and Elias Bareinboim. 2022. Causal fairness analysis. *arXiv preprint arXiv:2207.11385* (2022).

[35] Jose E Reyes Arias, Kale Kurtzhall, Di Pham, Mohamed Wiem Mkaouer, and Yasmine N Elglaly. 2022. Accessibility Feedback in Mobile Application Reviews: A Dataset of Reviews and Accessibility Guidelines. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[36] tarot 2022. *Tarot Cards of Tech*. Retrieved August 11, 2023 from https://tarotcardsoftech.artefactgroup.com/

[37] Nicholas Tilmes. 2022. Disability, fairness, and algorithmic bias in AI recruitment. *Ethics and Information Technology* 24, 2 (2022), 21.

[38] Shari Trewin. 2018. AI fairness for people with disabilities: Point of view. *arXiv preprint arXiv:1811.10670* (2018).

[39] Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for AI fairness for people with disabilities. *AI Matters* 5, 3 (2019), 40–63.

[40] Chia-En Tseng, Seoung Ho Jung, Yasmine N Elglaly, Yudong Liu, and Stephanie Ludi. 2022. Exploration on Integrating Accessibility into an AI Course. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1*. 864–870.

[41] Zhibo Wang, Xiaowei Dong, Henry Xue, Zhifei Zhang, Weifeng Chiu, Tao Wei, and Kui Ren. 2022. Fairness-aware adversarial perturbation towards bias mitigation for deployed deep models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10379–10388.

[42] Alice C Yu and John Eng. 2020. One algorithm may not fit all: how selection bias affects machine learning performance. *Radiographics* 40, 7 (2020), 1932–1937.

[43] Wenbin Zhang, Tina Hernandez-Boussard, and Jeremy Weiss. 2023. Censored fairness through awareness. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 14611–14619.