Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #45685

Case Study: Using Synthetic Datasets to Examine Bias in Machine Learning Algorithms for Resume Screening

Annika Haughey, Duke University

Graduate Student

Dr. Brian P. Mann, Duke University

Dr. Brian Mann is an endowed Professor of Mechanical Engineering at Duke University. He received his BS degree in 1996 from the University of Missouri prior to accepting a position with McDonnell Douglas Corporation. Three years later, he accepted a pos

Dr. Siobhan Oca, Duke University

Siobhan Rigby Oca is the director of master studies and an assistant professor of the practice in the Thomas Lord Department of Mechanical Engineering and Materials Science at Duke University, NC, USA. She received her B.Sc. from Massachusetts Institute of Technology and Master in Translational Medicine from the Universities of California Berkeley and San Francisco. She completed her Ph.D. in Mechanical Engineering in 2022 from Duke University. Her research interests include applied medical robotics, human robot interaction, and robotics education.

Case Study: Using Synthetic Datasets to Examine Bias in Machine Learning Algorithms for Resume Screening

Annika Haughey, Brian Mann, Siobhan Oca
Mechanical Engineering & Materials Science Department
Duke University

Abstract

The increasing use of artificial intelligence (AI) in recruitment, particularly through resume screening algorithms, raises significant ethical concerns due to the potential for biased decision-making. This case study explores these issues by developing a synthetic dataset mimicking the Amazon hiring tool controversy, where biases in training data led to discriminatory outcomes. Using artificial resumes that reflect a diverse applicant pool, students trained and interacted with a machine learning algorithm, which, despite excluding explicit demographic information, exhibited biases against underrepresented groups. This exercise highlights the ethical implications of deploying AI in decision-making processes and equips students with problem-solving techniques for addressing such challenges. Initially introduced in a graduate-level ethics course, this case study serves as a framework for teaching the intersection of technology and ethics, offering valuable lessons in recognizing and mitigating bias in AI systems for undergraduate and graduate students in engineering. [1]

Introduction

Machine learning has become a powerful tool in automating decisions and tasks across various fields, but its application can sometimes result in discriminatory practices. Notably, machine learning models have been increasingly employed in hiring processes. In this context machine learning algorithms have often been employed to screen resumes and identify the best candidates. While this may streamline recruitment, it has also led to instances of bias, where certain demographic groups are unfairly excluded or prioritized. These biases often stem from historical data used to train the models, which may reflect existing inequalities in the workforce. Such outcomes not only raise ethical concerns but also risk violating anti-discrimination laws. Addressing these issues requires developing algorithms that account for fairness and bias mitigation, alongside rigorous testing and transparency in how decisions are made. Without such measures, machine learning risks reinforcing systemic inequalities rather than promoting inclusivity and diversity in the workplace.

[1] The code associated with this study is available at: https://github.com/annikaLindstrom/EthicsInAI.git

Amazon Hiring Tool

A recent notable example of machine learning introducing bias in hiring practices involved Amazon. Since 2014, the company had been developing a machine learning model to assist with recruitment [1]. The tool was designed to automatically sort and score applicant resumes on a scale of one to five, aiming to identify top candidates for open positions [2]. Amazon trained 500 models to recognize approximately 50,000 terms extracted from 10 years of past applicant resumes. The system was configured to assign less weight to skills commonly listed by most applicants, such as proficiency in specific programming languages or technical abilities. Consequently, the model learned to favor candidates who used action verbs like ”executed” or ”captured,” which were more frequently found on resumes from male applicants, leading to a biased evaluation process.

In 2015, Amazon discovered that their recruitment tool was systematically rating female applicants lower for software developer and other technical roles. The model had learned to favor male-dominated keywords and specific gender-associated activities, colleges, or clubs. For instance, it penalized resumes containing the word ”women's” and downgraded graduates from two all-women's colleges [2]. After struggling to find a solution and losing confidence in the algorithm's fairness, Amazon ultimately discontinued the tool in 2017.

AI Use in Recruitment and Hiring

The Amazon case was one of the earliest and most notable examples of AI bias in the hiring process; however, more recent reports have uncovered similar issues. In 2016, LinkedIn's search function was found to exhibit gender bias by suggesting male alternatives for female names [3]. For example, a search for ”Andrea” would prompt the system to ask if the user meant ”Andrew,” while searches for male names did not trigger comparable suggestions. This discrepancy highlighted biases within the platform's algorithm, potentially limiting the visibility of female professionals. More recently, a study demonstrated that AI tools used for resume screening can also exhibit bias based on applicants' names [4]. Researchers trained large language models (LLMs) for resume screening and discovered that these models ranked candidates differently depending on the perceived race or gender associated with their names, resulting in discriminatory hiring outcomes.

Despite the well-documented biases and potential harms introduced by AI, its use in the hiring process continues to grow rapidly. In 2017, a survey by the talent software firm CareerBuilder found that 55% of U.S. human resource managers anticipated AI becoming a regular part of their work. However, actual adoption has far exceeded those predictions. The chair of the Equal Employment Opportunity Commission estimated that 99% of Fortune 500 companies now utilize some form of automated tool in their hiring practices [5]. According to a survey of HR professionals nearly two-thirds of employers now use AI in some aspect of their hiring processes [6]. AI applications range from online assessments that identify capable candidates to chatbots that answer questions about the application process [7].

As AI continues to advance in the hiring field despite its biases, finding solutions to address these issues is critical. A survey by HireVue found that both HR professionals and job applicants expressed hesitations about the use of AI in hiring [8]. The survey emphasized that education is one of the most effective solutions. To safely and effectively implement AI in recruitment, it is essential to raise awareness about the potential harms and biases inherent in these technologies. In response to the growing adoption of AI in hiring, the UK's Department for Science, Innovation, and Technology published a guide titled "Responsible AI in Recruitment." This guide warns employers about the risks of perpetuating historical hiring biases and highlights how the misuse of AI could lead businesses to inadvertently violate UK laws prohibiting discriminatory job advertisements [9].

The use of AI in hiring brings significant challenges, particularly with biases that can result in discriminatory practices. Despite these issues, its widespread adoption highlights its value in streamlining recruitment. Addressing these biases is essential to ensure ethical implementation and compliance with anti-discrimination laws. Education is a crucial step in this process, helping HR professionals and decision-makers understand AI's risks, limitations, and strategies for minimizing bias. This case study aims to highlight the dangers of unregulated AI in hiring and provide actionable solutions for responsible implementation. By promoting awareness and exploring bias mitigating strategies organizations can leverage AI's benefits while upholding fair and equitable hiring practices.

Case Studies in Ethics Education

Case studies are valuable instructional tools for demonstrating concepts to students through practical, real-world examples. By connecting theoretical knowledge to tangible applications, case studies help students understand the relevance and implications of the concepts they learn. In the context of engineering ethics, case studies are particularly effective for illustrating the importance of ethical principles. For instance, a case study by Lin and Greenberg [10] explores the challenges of using artificial intelligence in defense settings, underscoring the critical need for fairness in AI systems, particularly when decisions can have life-or-death consequences. Similarly, Zhang et al. [11] present a case study on the role of fairness in healthcare technology, demonstrating how biases in diagnostic algorithms can lead to inaccurate outcomes for specific population groups. Expanding on the use of case studies in engineering ethics, our study introduces a novel approach by addressing the ethical challenges of AI in hiring practices. By incorporating a synthetic dataset, it offers students hands-on experience, reinforcing ethical concepts and providing practical skills to analyze and mitigate biases in AI systems.

We propose a case study designed for active learning in a classroom which can be used to educate students on the ethics of AI implementation in decision-making roles. We introduce the Amazon resume screening case as an example and build a synthetic dataset designed to mimic the applicant resumes. We then build code to train a model, assess its fairness, and apply bias mitigation techniques. This code base [12] is designed to be implemented in a classroom with input from students to allow for active learning. In the subsequent sections we introduce the synthetic dataset and code and provide learning outcomes for students and ideas for future improvements.

Table 1: Features Contained in Synthetic Resume Dataset

Feature	**Description**	**Options**
Sex	Sex of the applicant	Male, Female
Employment_Gaps	Number of gaps in employment history	0, 1, 2
College_Club	Participation in college clubs	Coding Society, Tech Club, Student Council, Fraternity, Women In Tech, Student Council
Resume_Keywords	Action verbs used in resumes	achieved, innovated, strategic, organized, led, collaborative, support, helped
Education_Level	Highest education level attained	Master's, Bachelor's, PhD
Years_Experience	Years of relevant experience	0.0 - 17.0 years
Skills	Skills listed on the resume	system architecture, software engineering, data analysis, machine learning, project management, customer service
Position_Level	Job position level	Mid, Entry, Senior
Certifications	Professional certifications	Google Analytics Certified, Certified Scrum Master, AWS Certified, PMP, None
Programming_Languages	Programming languages known	Python, C++, Java, JavaScript, SQL, None
Project_Count	Number of projects completed	0 - 22 projects
GPA	Grade Point Average	2.0 - 4.0
Hired	Whether the applicant was hired	1 (Hired), 0 (Not Hired)

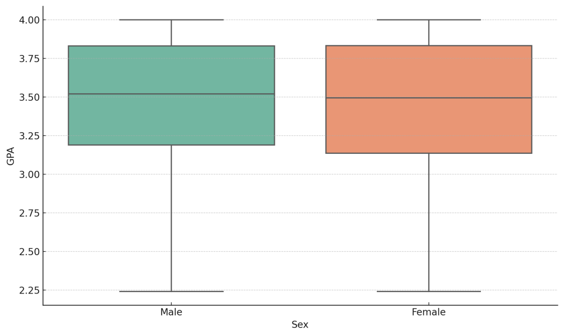
Methods

Feature Definition

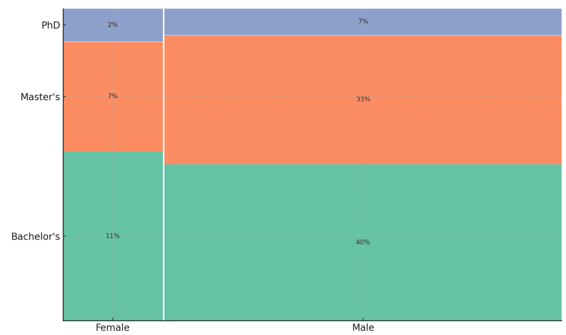
We developed a synthetic dataset of resumes to teach students about the ethical considerations when implementing AI in decision-making processes. The dataset contains data from 1000 resumes with 80% male and 20% female candidates, designed to mimic a standard distribution of employment in tech companies [13]. There are 13 features enumerating the information typically found on resumes, these are outlined in Table 1.

Each feature in the dataset was assigned based on a probability distribution tailored to its type and context. Numerical features, such as grade point average (GPA), project count, and years of experience, were generated using normal distributions centered around specific means.

Non-numerical features, like education level and certifications, were assigned using weighted probabilities, with individual weights defined for each category. Figure 1 outlines the distribution of two sex-agnostic features: GPA and education level. While there are slight variations due to the randomness introduced in creating the dataset, overall the GPA and education level distributions are equal between sexes.



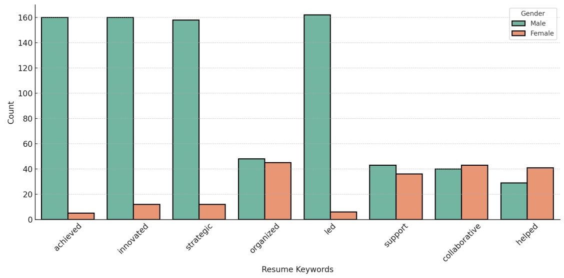
(a) Box plot showing the equal distributions of GPA among female and male applicants. Both sexes have similar distributions of GPA.



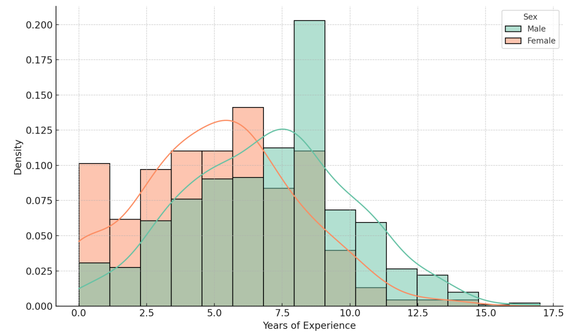
(b) Mosaic plot highlighting the distribution of degrees among male and female applicants. The education breakdown is approximately equal between sexes.

Figure 1: Sex-Agnostic Features

Certain features were influenced by the applicant's sex to reflect observed trends in the workforce. For example, the mean for years of experience was lower for female applicants, reflecting trends in the tech industry where women often younger on average. Features with sex-based definitions included employment gaps (women were more likely to have gaps in their work history), college clubs (male resumes could include fraternities, while female resumes might feature organizations like the Society of Women Engineers), and resume keywords (men were more likely to use stronger action-oriented terms). Two of these features: resume keywords and years experience are displayed in Figure 2, highlighting the difference between the sexes. These distinctions were intentionally introduced to explore the impact of such variations on hiring outcomes. All other features were distributed independently of sex, ensuring a mix of uniform and variable attributes across the dataset.



(a) Bar chart showing the number of resumes of each sex that had certain keywords. Certain keywords are seen significantly more often in male resumes.



(b) Histogram showing the distribution of years experience for each sex. The mean years of experience is notably lower for female applicants.

Figure 2: Sex-Dependent Features

Hiring Algorithm

We developed a scoring algorithm to determine whether a resume was labeled as hired or rejected, which serves as the dependent variable in the dataset. The algorithm assigns weights to various aspects of each resume and calculates a composite score. Resumes with the top 50% of scores were labeled as hired. Table 2 outlines the details of this scoring algorithm, including the features considered, the weights assigned to each feature, and the specific values that received higher scores. Notably, this algorithm is designed to be tunable, enabling adjustments to introduce more or less bias depending on the desired use case.

Table 2: Components of the Hiring Algorithm

Feature	**Weight**	**Condition**	**Notes**
GPA	0.15	–	Continuous variable
Years of Experience	0.15	–	Continuous variable
Education Level	0.2	Master's degree	Binary indicator
Position Level	0.2	Mid-level	Binary indicator
Skills	0.1	Software Engineering	Binary indicator
College Clubs	0.1	Tech Club	Binary indicator
Resume Keywords	0.05	Led	Binary indicator
Position Level	0.05	Senior level	Bonus for senior roles
Certifications	0.05	AWS Certified	Bonus for certification
Project Count	0.05	–	Scaled by max(project count)
Employment Gaps	-0.05	–	Penalty for gaps

This algorithm was designed to emulate a realistic system that a recruiter might use to evaluate candidates for a specific position. Although the algorithm does not explicitly factor in sex, bias is subtly introduced through differences in feature definitions influenced by sex. For example, the algorithm assigns a higher weight to years of experience. Since female applicants in the dataset tend to be younger, their scores are inadvertently lower in this category. Similar biases emerge in other features, reflecting the unequal probabilities embedded in the data generation process.

This scoring algorithm provides a framework for ranking resumes and defining the subset of high-ranking resumes as the hired group. While the machine learning model trained on this dataset is unaware of the scoring algorithm, it learns to approximate the mechanism as it trains on the labeled data. Therefore, the model inherits biases present in the training data, even when the sex feature is excluded, highlighting how bias can persist in machine learning systems through indirect pathways.

Machine Learning Model

To demonstrate how bias manifests in the machine learning model, we trained a model using the synthetic dataset. The model was trained without any knowledge of the algorithm used to generate the dataset, meaning it was ”blind” to the scoring methodology. We implemented a random forest classification model due to its ease of use and interpretability. Following a standard train-test split, the model was trained and its predictions are detailed below. The code for the model, along with interactive Python notebooks that document the process and results, is available on GitHub [12].

Once trained, we used the model to select the top 50 candidates from the test dataset. Without any external interventions, the model's top candidates included only 2 females (4%) and 48 males (96%). This stark disparity highlights how the underlying biases in the training data are absorbed and perpetuated by the model. Since demographic information such as sex is often excluded in real-world hiring practices, we trained a secondary model without this feature. This resulted in only a marginal improvement, with females comprising 6% of the top hires. This minimal change is unsurprising given that the original scoring algorithm did not explicitly factor in sex. Instead, bias is introduced indirectly through other features that correlate with demographic differences.

These initial results underscore the impact of historical bias in training data on machine learning models. Even when sensitive features like sex are excluded, the model implicitly learns the biases present in the data. In the next section we will highlight bias mitigation techniques to improve the fairness of the model.

Techniques to Mitigate Bias

The machine learning model trained on the historical data exhibited significant bias against female resumes. We aim to create a model that hires the most qualified candidates regardless of sex, this requires implementing strategies to mitigate this bias so that highly qualified female candidates are chosen for employment. To attempt to mitigate the bias in our model, we implemented three techniques in the interactive code, highlighting the strengths and limitations of each. The techniques explored were: feature selection, data balancing, and penalized training.

The most straightforward approach, feature selection, involves removing features that are highly correlated with the sensitive attribute, in this case, sex. To identify these features, we calculated the correlation between each feature and the sensitive attribute. The resulting correlations are

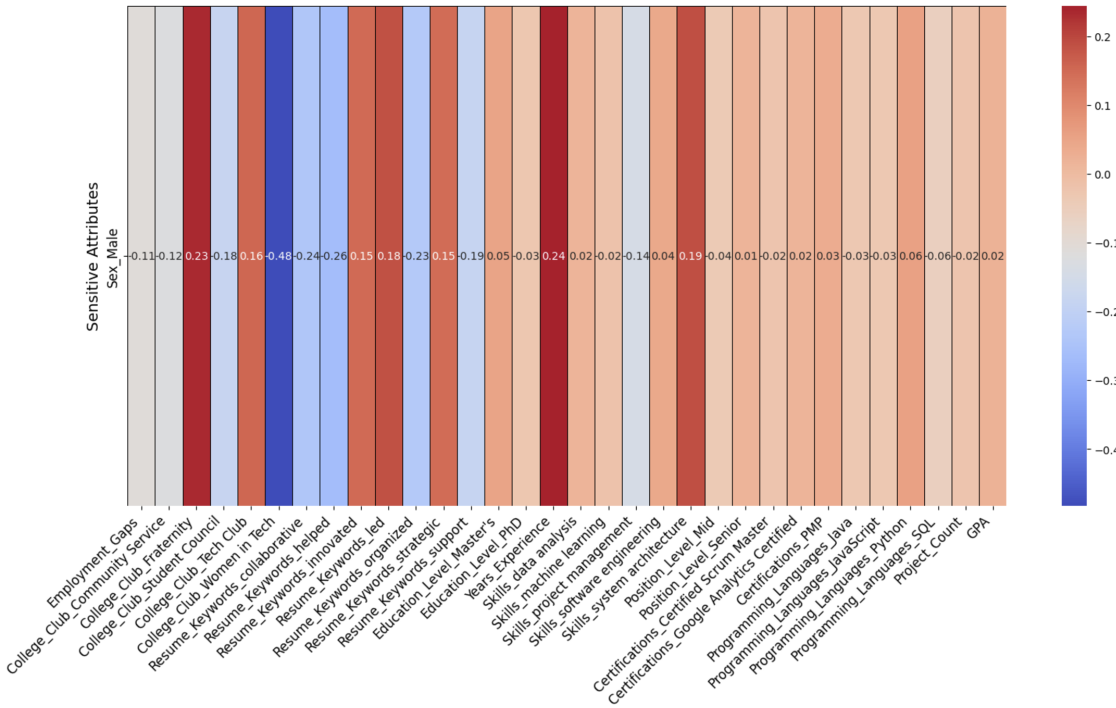


Figure 3: Feature Correlations in Synthetic Dataset

visualized in Figure 3, with deeper shades of red representing strong positive correlations and deeper shades of blue representing strong negative correlations. Features with the highest correlations to sex were identified as certain college clubs and years of experience. By excluding these features, we aimed to reduce the indirect propagation of bias in the model.

Excluding the features most correlated with sex resulted in a modest improvement in fairness. For example, training a model without the ”college clubs” and ”years of experience” features increased the percentage of females hired to 10%. However, this improvement comes at the cost of reduced predictive power, as the excluded features contain valuable information that contributes to the model’s accuracy. Despite the change, fairness remains limited, with 10% female hires still representing a significantly biased outcome.

Additionally we employed artificial data balancing to attempt to remove bias from the model. Since the dataset initially contained 80% male applicants, we reduced the number of male applicants to create a dataset balanced equally between sexes. This approach yielded much greater success, with the model trained on the balanced dataset hiring 34% female candidates. While this demonstrates the potential of balancing techniques to reduce bias, the model is not trained on a large number of applicants which results in a less robust model.

The final method utilized was penalized training through Microsoft’s Fairlearn library [14]. This library incorporates fairness constraints into machine learning models in interpretable ways. We applied a demographic parity constraint, penalizing the model’s loss function when unequal outcomes were predicted between sexes. This resulted in a significant improvement, with 47% of female candidates hired. However, the increase in fairness came at the expense of predictive power, reducing the model’s overall accuracy. As shown in Figure 4, while the true positive rates

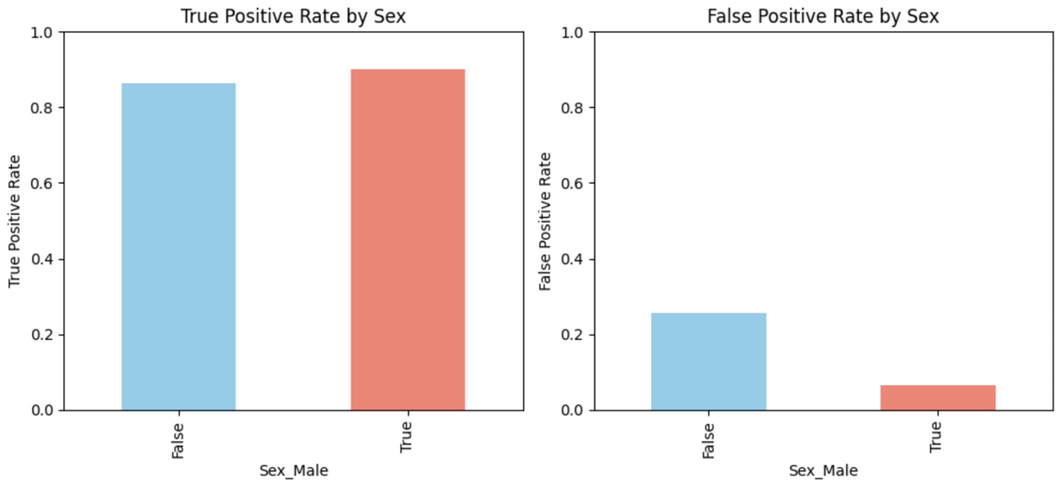


Figure 4: True and False Positive Rates using Fairlearn's Demographic Parity Penalization.

for each sex remain relatively high, the false positive rate for female applicants (blue) is notably significant. While Fairlearn's constraints have created an artificially more equitable model, the high false positive rates for females still indicate that the underlying model holds bias.

The metrics used by the model's algorithm indicate that it is hiring ”unqualified” female candidates. Objectively, this is not the case when comparing metrics that do not have hidden bias such as GPA of the candidates for which the mean is 3.5 for both sexes. The algorithm for hiring however places weight on metrics that have hidden biases such as years experience and resume keywords that often have discrepancies among the sexes. This leads the model to assume certain female candidates are not as qualified as male candidates. Fairlearn aims to counter this by artificially balancing the number of hired candidates. This results in a new model that might look unfair, as it appears to be hiring a higher percentage of ”unqualified” female candidates. The definition of unqualified here however comes from the original model which was shown to have significant bias. This highlights the difficulty in building fair machine learning models and requires the users to fully understand where the biases are introduced.

None of these techniques achieved complete fairness between female and male applicants, and each introduced trade-offs by reducing the model's accuracy to varying degrees. Achieving a perfectly fair and equal machine learning model is a complex challenge, particularly when working with human data that inherently reflects societal biases. This case study underscores this reality for students, emphasizing the critical importance of transparency and interpretability when developing and deploying machine learning models that impact decision-making processes. Understanding these trade-offs is essential for creating responsible and ethical AI systems.

Case Study in the Classroom

This case study emphasizes the challenges of building ethical AI and serves as a valuable educational tool in this field. It allows students to explore how hidden biases in training data can propagate through machine learning models, resulting in unfair and unethical outcomes. By presenting a real-world example of the potential harms caused by biased AI, the case study offers students hands-on experience in attempting to mitigate these biases and their effects.

We introduced the case study in the classroom by starting with a discussion of the Amazon resume screening example and the ethical implications of using AI in decision-making roles. Following this, we presented the synthetic dataset and adopted an active learning approach, allowing students to engage hands-on during the class period. Students implemented portions of the provided code base and explored various bias mitigation strategies, fostering a practical understanding of the challenges and complexities involved in addressing bias in AI systems.

Table 3: Expected Student Learning Outcomes

Outcome #	**Description**
1	Understand how biases in training data can propagate through machine learning models
2	Interpret machine learning model results and communicate the implications of these outcomes in terms of fairness and accuracy
3	Evaluate the strengths and limitations of bias mitigation strategies
4	Reflect on the ethical challenges of deploying machine learning systems in decision-making scenarios, emphasizing the importance of transparency, interpretability, and responsible AI practices

Through using this case study in the classroom, students gain a deeper understanding of the complexities involved in eliminating bias, recognizing that the process is far from straightforward. Additionally, they learn that creating fairer models often comes at the cost of reduced predictive power, highlighting the trade-offs inherent in ethical AI development. The expected student learning outcomes for implementing this case study in the classroom are detailed in Table 3.

Conclusion

We present a case study on bias in machine learning, with a specific focus on resume-screening. This case study draws from a real-world example of a tool developed by Amazon in 2014 to automate resume screening. The tool was later abandoned due to its biases against female applicants and female-coded attributes in resumes. Through this example, we examine the ethical implications of such scenarios and introduce a synthetic dataset designed for active learning settings. This dataset helps to illustrate how bias can manifest in machine learning models and provides a practical framework for exploring these issues.

This case study has been successfully implemented in graduate-level courses, including an ethics in automation course and an introduction to machine learning class, to highlight the ethical implications of AI in real-world applications. We envision its use in both graduate and undergraduate courses across disciplines such as machine learning, ethics, and engineering. Unlike other case studies commonly used in engineering education [10, 11], this approach incorporates a synthetic dataset, enabling active learning opportunities and providing students with hands-on experience in machine learning and bias mitigation techniques. The case study design is adaptable, allowing for varying levels of programming complexity depending on the students' experience.

The synthetic dataset is highly versatile and can be used in courses focused on both bias mitigation strategies and machine learning model development. It offers flexibility to include additional features or to modify the scoring algorithm to introduce more or less bias. Furthermore, it can be expanded to address biases against other demographics, such as age or race, making it a valuable tool for teaching the complexities of fairness in AI systems. This adaptability ensures the case study remains relevant and impactful in diverse educational contexts.

Future work will include assessing the learning impacts of the case study in a classroom setting. We plan to evaluate the effectiveness of this tool as a means of educating students on bias in machine learning by implementing pre- and post-class surveys. These surveys will measure changes in students' understanding of key concepts, such as the origins of bias in training data, the ethical implications of biased AI, and the trade-offs between fairness and predictive accuracy. Additionally, we will gather qualitative feedback on the usability of the synthetic dataset, the relevance of the exercises, and the clarity of the concepts presented.

Beyond surveys, we aim to analyze student performance on assignments or projects related to the case study to assess practical skill development in bias mitigation and fairness-aware model design. This feedback will inform potential improvements to the case study, such as expanding the dataset to include additional demographic biases, incorporating more advanced fairness metrics, or tailoring the exercises for different levels of programming expertise. By refining the case study based on classroom implementation, we hope to create a scalable and versatile educational tool that can be adapted for use across a variety of courses and disciplines.

References

[1] C. M. University. (2018) Amazon scraps secret artificial intelligence recruiting engine that showed biases against women. [Online]. Available: https://www.ml.cmu.edu/news/news-archive/2016-2020/2018/october/amazon-scraps-secret-artificial-intelligence-recruiting-engine-that-showed-biases-against-women.html

[2] J. Dastin, "Amazon scraps secret ai recruiting tool that showed bias against women," *Reuters*, 2018. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

[3] K. Fingas, "How linkedin's search engine may reflect a bias," *The Seattle Times*, 2020. [Online]. Available: https://www.seattletimes.com/business/microsoft/how-linkedins-search-engine-may-reflect-a-bias/

[4] U. of Washington, "Gender, race, and intersectional bias in resume screening via language model retrieval," Tech. Rep., 2024.

[5] J. Hanson, "Ai is replacing humans in the interview process—what you need to know to crush your next video interview," *Forbes*, 2023. [Online]. Available: https://www.forbes.com/sites/janehanson/2023/09/30/ai-is-replacing-humans-in-the-interview-processwhat-you-need-to-know-to-crush-your-next-video-interview/

[6] Workable. (2023) Ai in hiring: Use cases and examples. [Online]. Available: https://resources.workable.com/stories-and-insights/ai-in-hiring-use-cases

[7] T. T. Business and M. C. Summit. (2024) How to avoid the pitfalls when using ai to recruit new employees. [Online]. Available: https://www.thetimes.com/business-money/ceo-summit/article/how-to-avoid-the-pitfalls-when-using-ai-to-recruit-new-employees-ml3vnlpgr

[8] HireVue, "Ai in hiring report," 2024. [Online]. Available: https://www.hirevue.com/resources/report/ai-in-hiring-report

[9] I. Department for Science and Technology, "Responsible ai in recruitment: Guidance," Department for Science, Innovation and Technology, United Kingdom, Tech. Rep., March 2024. [Online]. Available: https://www.gov.uk/government/publications/responsible-ai-in-recruitment-guide

[10] E. Lin-Greenberg, "Wrestling with killer robots: The benefits and challenges of artificial intelligence for national security," 2021. [Online]. Available: https://mit-serc.pubpub.org/pub/wrestling-with-killer-robots/release/2

[11] H. Zhang, T. Hartvigsen, and M. Ghassemi, "Algorithmic fairness in chest x-ray diagnosis: A case study," 2023. [Online]. Available: https://mit-serc.pubpub.org/pub/algorithmic-chest/release/2

[12] "Ethics in ai," 2025. [Online]. Available: https://github.com/annikaLindstrom/EthicsInAI.git

[13] D. D. Investor. (2024) Amazon scraps secret ai recruiting engine that showed biases against women. [Online]. Available: https://medium.datadriveninvestor.com/amazon-scraps-secret-ai-recruiting-engine-that-showed-biases-against-women-995c505f5c6f

[14] S. Bird, M. Dudík, R. Edgar, B. Horn, R. Lutz, V. Milan, M. Sameki, H. Wallach, and K. Walker, "Fairlearn: A toolkit for assessing and improving fairness in ai," 2020, microsoft Research Technical Report MSR-TR-2020-32. [Online]. Available: https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/