

Article

Not peer-reviewed version

Mitigating Age-related Bias in Predictive Policing Algorithms

[Ahmed Saud Almasoud](#)*

Posted Date: 24 November 2023

doi: 10.20944/preprints202311.1534.v1

Keywords: predictive policing; algorithms; fairness; age bias; strategic subject list



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Mitigating Age-Related Bias in Predictive Policing Algorithms

Ahmed S. Almasoud

Department of Information Systems, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia; Almasoud@psu.edu.sa

Abstract: This study addressed algorithmic bias in predictive policing, focusing on the Chicago Police Department's Strategic Subject List (SSL) dataset. We specifically focused on identifying and mitigating age-related biases, a notably underexplored area in prior research. Our research introduced Conditional Score Recalibration as a bias mitigation strategy alongside the well-established Class Balancing technique. Conditional Score Recalibration involved reassessing and adjusting risk scores for individuals initially assigned moderately high-risk scores in the dataset. This recalibration marked such individuals as low risk if they met three conditions, namely: no prior arrests for violent offenses, no previous arrests for narcotic offenses, and having never been involved in shooting incidents. These fairness strategies were implemented on the Random Forest model, and the fairness metrics employed included Equality of Opportunity Difference, Average Odds Difference, and Demographic Parity. The results showed a significant improvement in model fairness, particularly for age biases, without compromising the model's accuracy. These findings challenged the often-assumed trade-off between fairness and accuracy, underscoring the feasibility of achieving fairness without compromising accuracy.

Keywords: predictive policing; algorithms; fairness; age bias; strategic subject list

1. Introduction

In the evolving landscape of law enforcement and public safety, the use of predictive algorithms has gained significant traction. These algorithms aim to anticipate potential criminal activities and identify individuals at higher risk of involvement in crimes. While the strategic application of such algorithms holds promise for enhancing public safety, it raises critical questions about fairness and bias, especially when considering sensitive demographic features like age, race, and gender (Asif *et al.*, 2018; Saleem & Malik, 2022; El Baroudy, 2023).

Meanwhile, the issue of fairness in policing algorithms is not just a technical challenge but also a societal concern, as biases in these algorithms can have far-reaching consequences (Kyriakakis, 2021; Idowu & Almasoud, 2023). Therefore, our research is motivated by the need to ensure that the deployment of such predictive tools does not inadvertently perpetuate or exacerbate existing societal biases. This study investigates bias in the Chicago Police Department's Strategic Subject List (SSL) dataset and then mitigates the bias in algorithms built on the dataset. The specific research objectives are as follows:

1. To analyze the bias in the Strategic Subject List dataset with respect to age and race.
2. To implement and evaluate the effectiveness of bias mitigation strategies in enhancing the fairness of predictive policing algorithms.

2. Related Works

2.1. Fairness Strategies in Policing Algorithms

Recent studies have made significant contributions to addressing fairness concerns in predictive policing algorithms. Downey et al. (2023) explored the integration of domain knowledge into predictive policing datasets. This approach involved augmenting datasets with socio-economic indicators such as education levels, employment rates, and poverty statistics from the 2014 census. While this integration showed limited effects on traditional classification metrics, it yielded notable improvements in fairness across various protected classes, particularly evident in the improvements seen in the Theil index related to racial fairness.

Berk et al. (2021) focused on enhancing fairness in risk assessment algorithms through counterfactual reasoning. The methodology involved initially training a stochastic gradient boosting model on data from a more privileged group (White offenders). Then, it transports the joint predictor distribution from the less privileged class (Black offenders) to the more privileged class with the aim of treating the disadvantaged class members as if they are members of an advantaged class. This effectively reduced the bias against the disadvantaged class.

Furthermore, Hung & Yen (2023) conducted a causal analysis of predictive policing algorithms (PPAs) and identified police actions as a primary contributor to model discrimination. The study revealed a self-perpetuating cycle where increased police deployment leads to higher arrest rates, subsequently inflating reported crime rates and justifying further police deployment.

Jain et al. (2019) introduced Singular Race Models, a method designed to mitigate race-based biases in recidivism prediction. By segregating datasets along racial lines and training race-specific models, this approach succeeded in improving prediction accuracy and reducing racial discrimination within these models. Similarly, Rodolfa et al. (2020) and Montana et al. (2023) proposed the implementation of individually tailored interventions based on specific demography groups and cohorts.

2.2. Sensitive Features in Policing Algorithms

In Table 1, we present a comparison between the sensitive features investigated in previous studies on fairness of policing and recidivism algorithms. The comparison shows that most studies focused on race (12 out of 13 studies). Meanwhile, only two of the 13 papers examined bias with respect to socio-economic status. Also, only three of the 13 papers each investigated bias with respect to gender and age. In this study, we will address part of this research gap by thoroughly investigating bias related to age.

Table 1. Sensitive features investigated by previous studies on policing algorithms.

Paper	Sensitive Features			
	Sex	Race	Status*	Age
Downey et al. (2023)	✓	✓		✓
Berk et al. (2021)		✓		
Pastaltzidis et al. (2022)		✓		
Udoh (2020)	✓	✓		✓
Montana et al. (2023)				✓
Mohler et al. (2018)		✓		
Somalwar et al. (2021)		✓		
Urcuqui et al. (2020)			✓	
Van Berkel et al. (2019)	✓	✓	✓	
Jain et al. (2019)		✓		
Khademi & Honavar (2020)		✓		
Ingram et al. (2022)	✓	✓		
Rodolfa et al. (2020)		✓		

3. Methods

3.1. Dataset

This research used the Chicago Police Department's Strategic Subject List dataset which covers 398,684 people from August 2012 to July 2016. The Strategic Subject List (SSL) score is a risk assessment score that reflects an individual's probability of being involved in a shooting incident either as a victim or an offender. The scores range from 0 (extremely low risk) to 500 (extremely high risk). According to Posada (2017), scores above 250 are deemed to be high risk.

Initially, the dataset has 398,684 data instances and 48 features. We applied preliminary pre-processing steps which involved dropping columns with significant missing values, dropping rows with missing values for longitude and latitude columns, among others. Afterwards, we followed up with additional pre-processing steps highlighted below:

1. The feature 'AGE AT LATEST ARREST' originally contained age ranges. We simplified this by categorizing individuals into two groups: those under 30 years old ('less than 20' and '20-30') were encoded as 1 and those 30 years or older were encoded as 0.
2. For the 'RACE CODE CD' feature, we retained only two categories for simplicity and relevance to the study's focus: 'BLK' (Black) and 'WHI' (White), encoded as 1 and 0, respectively. Rows of data that did not fall into these two categories were dropped.
3. The 'WEAPON I' and 'DRUG I' features were binary encoded 1 (True) and 0 (False).
4. For the target variable, SSL SCORE, scores above 250 were encoded as 1 indicating high risk, while scores of 250 and below were encoded as 0, indicating low risk aligning with the provided by Posada (2017).

In the end, we have 170,694 data instances and 12 features. The features are listed in Table 2.

Table 2. Features selected as predictors.

Feature	Description
AGE AT LATEST ARREST	The individual's age at the time of their most recent arrest.
VICTIM SHOOTING INCIDENTS	The number of times an individual has been the victim of a shooting.
VICTIM BATTERY OR ASSAULT	The number of times an individual has been the victim of an aggravated battery or aggravated assault;
ARRESTS VIOLENT OFFENSES	The number of times the individual has been arrested for a violent offense.
GANG AFFILIATION	Indicator if an individual has been confirmed to be a member of a criminal street gang.
NARCOTIC ARRESTS	The number of times the individual has been arrested for a narcotics offense.
TREND IN CRIMINAL ACTIVITY	The trend of an individual's recent criminal activity.
UUW ARRESTS	The number of times the individual has been arrested for Unlawful Use of Weapons.
WEAPON I	Is 'Yes' if at least one Weapon (UUW) Arrest in past 10 years.
DRUG I	Is 'Yes' if at least one Drug Arrest in past 10 years.
LATITUDE	Latitude of Centroid of Census Tract of Arrest for the Subject's latest arrest record.
LONGITUDE	Longitude of Centroid of Census Tract of Arrest for the Subject's latest arrest record.

Additional feature engineering: After getting initial results for our models, we obtained feature importance and based on the values for each feature, we decided to drop WEAPON I, DRUG I,

LATITUDE I, AND LONGITUDE I from the features as they have very little contribution to the model.

Sensitive Features: The sensitive features in this research are AGE and RACE. However, it should be noted that RACE is only used for bias assessment. It is not included as a predictor in this task just like the dataset provider (Chicago Police Department) excluded it.

3.2. Fairness Metrics

The fairness metrics used for evaluation include Equality of Opportunity Difference, Average Odds Difference and Demographic Parity.

1. **Demographic Parity:** This metric measures whether the probability of a positive outcome (e.g., being classified as high risk) is the same across different groups. In other words, demographic parity is achieved if each group has an equal chance of receiving the positive outcome, regardless of their actual proportion in the population.
2. **Equality of Opportunity:** This metric specifically focuses on the true positive rate, ensuring that all groups have an equal chance of being correctly identified for the positive outcome when they qualify for it. It is a more nuanced metric that considers the accuracy of the positive predictions for each group, thereby ensuring fairness in the model's sensitivity.
3. **Average Odds Difference:** This metric evaluates the average difference in the false positive rates and true positive rates between groups. It combines aspects of both false positives (instances wrongly classified as positive) and true positives (correctly classified positive instances), providing a comprehensive view of the algorithm's performance across different groups. A lower value in this metric indicates a more fair algorithm, as it suggests minimal disparity in both false and true positive rates across groups.

3.3. Research Objective 1: Bias Analysis

According to Posadas (2017) and Downey (2023), the Strategic Subject List dataset is highly skewed and has bias with respect to age, gender, and race. People over the age of 40 tend to have a low SSL score. In addition, while it is unlikely that all the people under 20 years old would be involved in a shooting, the dataset showed all of them as high risk (Downey, 2023). Furthermore, there are 127,513 individuals on the list who have never been arrested or shot, but around 90,000 of them are deemed to be at high risk (Posadas, 2017). Therefore, it's critical to analyze the bias in the SSL dataset before building the algorithms.

1. The core of our analysis involves assessing the distribution of SSL scores across the demographic categories i.e., age, gender, and race. We did this analysis using histogram plots as shown in Figures 3–5.
2. Second, we examined a specific subset of dataset capturing individuals with SSL scores over 250 who are classified as "High Risk" despite having no prior arrest for violent offenses, no prior arrest for narcotic offenses and never been a victim of shooting incidents.

3.4. Research Objective 2: Bias Mitigation

Our preliminary analysis of the dataset shows that all individuals below the age of 30 have scores higher than 250, that is, they are all marked as High Risk. This shows that the dataset is significantly skewed against young people. However, rather than implement a mitigation strategy that will focus on young people alone (as that would likely introduce bias against old people), we introduced a more nuanced mitigation strategy, **Conditional Score Recalibration** which was applied to the entire dataset, young and old. This strategy is designed to reassess and adjust the risk scores assigned to individuals within a specific subset of the dataset. Under the Conditional Score Recalibration framework, individuals originally assigned a moderately high-risk score, ranging from 250 to 350, are re-evaluated and marked as low risk if they meet **ALL** the following conditions:

- a. No prior arrest for violent offenses
- b. No prior arrest for narcotic offenses
- c. Never been a victim of shooting incidents.

Class Balancing: Post-recalibration, the dataset comprised 111,117 individuals classified as low risk and 59,577 individuals classified as high risk. To address this imbalance, we employed a class balancing technique. This technique involved under-sampling the majority class, which in this case was the low-risk group. By randomly selecting a subset of the low-risk individuals equivalent in size to the high-risk group, we achieved a balanced dataset.

4. Results and Discussion

4.1. Research Objective 1: Bias Analysis

Figure 1 illustrates a clear racial disparity within the dataset, with a significantly higher number of black individuals compared to white. This discrepancy suggests that Black individuals are more frequently involved in arrests or encounters with the police.

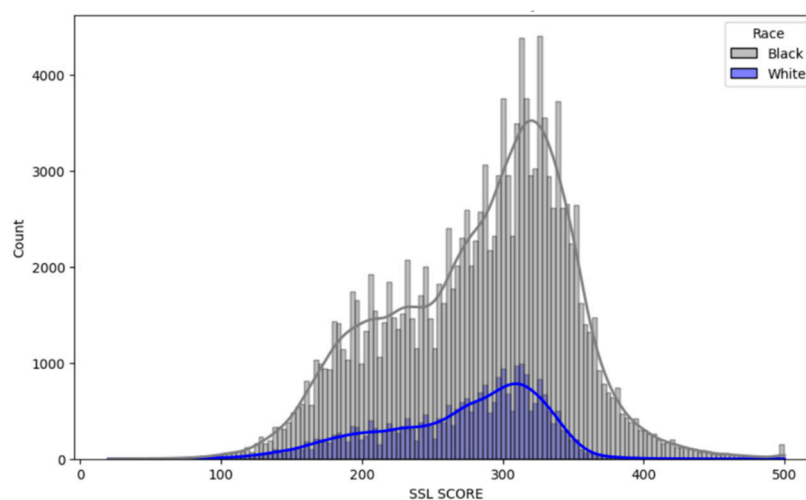


Figure 1. Distribution of SSL Scores by Race.

Figure 2 reveals a notable difference in the age demographic within the dataset compared to the trends seen in race. First, the dataset comprises a higher proportion of younger individuals (under 30 years) compared to older individuals (30 years and above). Furthermore, a striking aspect revealed by the analysis is that all individuals in the younger age bracket are assigned SSL scores above 250, categorizing them as high risk. In contrast, the older age group shows a markedly different distribution as only a small fraction of individuals in this demographic are classified as high risk, with the majority being deemed low risk.

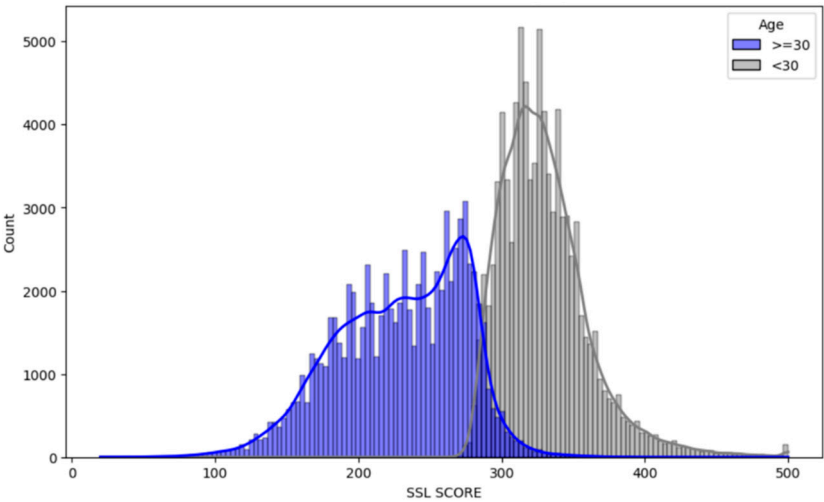


Figure 2. Distribution of SSL Scores by Age.

This disparity in risk classification between younger and older individuals raises critical considerations about the underlying factors driving these assessments. The 100% classification of younger individuals as high risk, regardless of other factors, may point to age-related biases in the Chicago Police department assessment process. This finding aligns with Posadas (2017) which stated that the SSL scores assigned by the Chicago Police Department do little more than reinforce the “age out of crime” theory – a theory that suggests individuals “grow out” of crime in their 30s. No doubt, this shows that the dataset is highly biased against young people. It also aligns with findings by Hung & Yen (2023) which noted that police actions constitute the primary contributor to model discrimination.

Figure 3 from our analysis further highlights a significant aspect of the dataset. It shows a considerable number of individuals who have not had any prior arrests for violent or narcotic offenses and have never been victims of shooting incidents, yet they have been assigned SSL scores above 250, categorizing them as high risk. This finding raises important questions about the criteria used for determining SSL scores and suggests a potential disconnect between an individual’s documented criminal history and their assessed risk level.

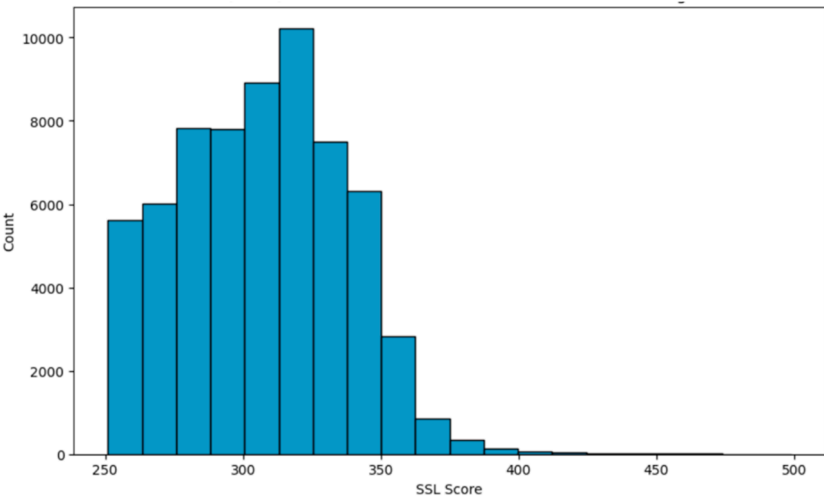


Figure 3. SSL Scores (>250) for those with no violent & narcotic arrests and shooting incidents.

4.2. Research Objective 2: Bias Mitigation

To determine the most suitable model for our analysis, we initially developed three different models: Logistic Regression, Random Forest, and Gradient Boosting, without incorporating any bias

mitigation strategies. The performance of all three models was comparable; however, we ultimately selected the Random Forest model due to its robustness against overfitting and its ability to handle non-linear relationships in large datasets more effectively.

The performance and fairness result of the Random Forest model are in Table 3.

Table 3. Results obtained for Random Forest Model. **Accuracy = 0.8314; F1 = 0.83.**

Sensitive Features	Demographic Parity	Equality of Opportunity	Average Odds Difference
Race	0.09923	0.08356	0.07679
Age	0.8517	0.7616	0.3349

These results reveal that the model performs fairly with respect to race, as all fairness metrics for race are within the generally accepted threshold of -0.1 to 0.1. However, there is a significant bias regarding age indicating a pronounced disparity in the model’s predictions. This high model bias against young people is not a surprise given the distribution of SSL Scores by Age in Figure 3.

After implementing the Conditional Score Recalibration and Class balancing strategies to mitigate the bias, we obtained the following results (Table 4):

Table 4. Results obtained after applying bias mitigation strategies. **Accuracy = 0.9014; F1 = 0.90.**

Sensitive Features	Demographic Parity	Equality of Opportunity	Average Odds Difference
Race	0.1170	0.08768	0.03523
Age	0.3128	0.1521	0.02024

The post-mitigation results show a significant improvement in the model’s fairness with respect to age. The Demographic Parity metric for age decreased significantly from 0.8517 to 0.3128, and the Equality of Opportunity metric saw a reduction from 0.7616 to 0.1521. The Average Odds Difference for age also improved from 0.3349 to a much lower value of 0.02024, indicating a substantial reduction in bias. For race, the results remained largely within the acceptable threshold, with the Equality of Opportunity and Average Odds Difference metrics well within the -0.1 to 0.1 range. The Demographic Parity metric, while slightly exceeding this range at 0.117, is still relatively close, indicating that the model maintains a fair degree of parity across racial groups.

The implementation of bias mitigation strategies not only improved fairness but also significantly improved the model’s overall performance. The accuracy of the model increased from 0.8314 to 0.9014, and the F1 score increased from 0.83 to 0.90. This improvement effectively dispels the commonly held notion that efforts to increase fairness invariably compromise performance. In fact, our findings are in line with recent research on policing and recidivism algorithms, such as Urcuqui et al. (2020) and Pastaltzidis et al. (2022), which assert that there is no strict trade-off between fairness and accuracy.

5. Conclusion

Our study on the fairness and accuracy of predictive policing algorithms, specifically focusing on the Chicago Police Department’s Strategic Subject List (SSL) dataset, reveals significant insights into the balance between predictive accuracy and fairness in law enforcement algorithms. Our analysis initially identified a notable age-related bias in the SSL dataset, particularly against younger individuals. In response, we implemented two bias mitigation strategies: Conditional Score Recalibration and Class Balancing.

We observed that the application of these bias mitigation strategies led to a significant reduction in age-related biases. Remarkably, this improvement in fairness did not come at the cost of accuracy. Instead, we witnessed an increase in the model’s accuracy from 0.8314 to 0.9014, and an improvement in the F1 score from 0.83 to 0.90. These results challenge the prevalent notion that fairness and

accuracy are mutually exclusive. Also, the study contributes to the ongoing discourse on the responsible use of AI in law enforcement. It underscores the importance of continuously scrutinizing and refining predictive policing tools to ensure they serve the public equitably.

Funding: This research was funded by Prince Sultan University, Saudi Arabia. The APC was funded by the same institution.

Data Availability Statement: The data supporting the findings of this study are from the publicly available Strategic Subject List - Historical dataset, released by the Chicago Police Department. The dataset can be accessed at <https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List-Historical/4aki-r3np>.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Asif, M., Shahzad, M., Awan, M.U. and Akdogan, H. (2018), "Developing a structured framework for measuring police efficiency", *International Journal of Quality & Reliability Management*, Vol. 35 No. 10, pp. 2119-2135. <https://doi.org/10.1108/IJQRM-04-2017-0067>
2. Berk, R. A., Kuchibhotla, A. K., & Tchetgen Tchetgen, E. (2021). Improving fairness in criminal justice algorithmic risk assessments using optimal transport and conformal prediction sets. *Sociological Methods & Research*, 00491241231155883.
3. Downey, A., Islam, S. R., & Sarker, M. K. (2023, May). Evaluating Fairness in Predictive Policing Using Domain Knowledge. In *The International FLAIRS Conference Proceedings* (Vol. 36).
4. El Baroudy, J. (2023). Autonomous Weapon Systems: Attributing the Corporate Accountability. *Access to Just. E. Eur.*, 222.
5. Hung, T. W., & Yen, C. P. (2023). Predictive policing and algorithmic fairness. *Synthese*, 201(6), 206.
6. Idowu, J., & Almasoud, A. (2023). Uncertainty in AI: Evaluating Deep Neural Networks on Out-of-Distribution Images. *arXiv preprint arXiv:2309.01850*.
7. Ingram, E., Gursoy, F., & Kakadiaris, I. A. (2022, December). Accuracy, Fairness, and Interpretability of Machine Learning Criminal Recidivism Models. In *2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)* (pp. 233-241). IEEE.
8. Jain, B., Huber, M., Fegaras, L., & Elmasri, R. A. (2019, June). Singular race models: addressing bias and accuracy in predicting prisoner recidivism. In *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 599-607).
9. Khademi, A., & Honavar, V. (2020, April). Algorithmic bias in recidivism prediction: A causal perspective (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 10, pp. 13839-13840).
10. Kyriakakis, J. (2021). *Corporations, Accountability and International Criminal Law: Industry and Atrocity*. Edward Elgar Publishing.
11. Mohler, G., Raje, R., Carter, J., Valasik, M., & Brantingham, J. (2018, October). A penalized likelihood method for balancing accuracy and fairness in predictive policing. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 2454-2459). IEEE.
12. Montana, E., Nagin, D. S., Neil, R., & Sampson, R. J. (2023). Cohort bias in predictive risk assessments of future criminal justice system involvement. *Proceedings of the National Academy of Sciences*, 120(23), e2301990120.
13. Pastaltzidis, I., Dimitriou, N., Quezada-Tavarez, K., Aidinlis, S., Marquenie, T., Gurzawska, A., & Tzovaras, D. (2022, June). Data augmentation for fairness-aware machine learning: Preventing algorithmic bias in law enforcement systems. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2302-2314).
14. Posadas, B. (2017, June 26). *How strategic is Chicago's "Strategic subjects list"? upturn investigates*. Medium. Retrieved October 6, 2022, from <https://medium.com/equal-future/how-strategic-is-chicagos-strategic-subjects-list-upturn-investigates-9e5b4b235a7c>
15. Rodolfa, K. T., Salomon, E., Haynes, L., Mendieta, I. H., Larson, J., & Ghani, R. (2020, January). Case study: predictive fairness to reduce misdemeanor recidivism through social service interventions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 142-153).

16. Saleem, F., & Malik, M. I. (2022). Safety management and safety performance nexus: role of safety consciousness, safety climate, and responsible leadership. *International journal of environmental research and public health*, 19(20), 13686.
17. Somalwar, A., Bansal, C., Lintu, N., Shah, R., & Mui, P. (2021, October). AI For Bias Detection: Investigating the Existence of Racial Bias in Police Killings. In *2021 IEEE MIT Undergraduate Research Technology Conference (URTC)* (pp. 1-5). IEEE.
18. Udoh, E. S. (2020, September). Is the data fair? An assessment of the data quality of algorithmic policing systems. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (pp. 1-7).
19. Urcuqui, C., Moreno, J., Montenegro, C., Riascos, A., & Dulce, M. (2020, November). Accuracy and fairness in a conditional generative adversarial model of crime prediction. In *2020 7th International Conference on Behavioural and Social Computing (BESC)* (pp. 1-6). IEEE.
20. Van Berkel, N., Goncalves, J., Hettiachchi, D., Wijenayake, S., Kelly, R. M., & Kostakos, V. (2019). Crowdsourcing perceptions of fair predictors for machine learning: A recidivism case study. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-21.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.