Module 8 Signature Assignment

Biased data exists and informs business outcomes – decisions made from faulty or biased data can have devastating consequences on reputations, society, and the bottom line. Identify an example of how bias has affected business analytics in the real-world (i.e. tweet analysis during hurricane Sandy, racial profiling in facial recognition software, pothole tracking in Boston, recidivism prediction tool, etc.).

In your research essay, respond to the following:

- Describe the organization you selected
- Provide detailed background of the situation, describe the type(s) of bias in the data.
- Analyze the failures in light of the data lifecycle – which part (or parts) of the lifecycle were compromised and how?
- What policies and procedures should have been in place to protect against the scenario?
- What ethical violations occurred in reporting or communication?
- Propose a detailed solution that takes into account urgent legal and other strategies.

For this essay, I will discuss COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) as illustrative.

1. Organization Description

The organization involved in the biased data incident is Northpointe, Inc., now known as Equivant (Angwin et al, 2016). This company developed the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool. COMPAS is a risk assessment software used by the criminal justice system to predict the likelihood of a defendant reoffending. It has been widely adopted across various states in the United States to assist judges in making decisions about bail, sentencing, and parole. The tool's algorithm analyzes a range of factors, including criminal history and personal information, to generate risk scores. However, it has faced criticism for exhibiting racial bias, as studies have shown it disproportionately labels African American defendants as higher risk compared to white defendants with similar backgrounds. This has raised significant ethical concerns about the fairness and accuracy of using such tools in judicial decision-making processes.

2. Background of the Situation

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) tool is a risk assessment algorithm used in the U.S. criminal justice system to predict the likelihood of defendants reoffending. The tool has been widely adopted to assist judges in making decisions regarding bail, sentencing, and parole. However, in 2016, ProPublica published an investigative report revealing significant racial bias in the algorithm's predictions. The study found that COMPAS disproportionately labeled black defendants as high risk for recidivism compared to their white counterparts, even when controlling for other factors such as prior crimes, age, and gender (Neosmart, 2024), (Angwin et al, 2016).

ProPublica's analysis showed that black defendants were nearly twice as likely to be incorrectly classified as high risk compared to white defendants. Conversely, white defendants were more often misclassified as low risk, yet went on to reoffend at higher rates than black defendants who were similarly categorized (Angwin et al, 2016), (Mesa, 2021). This disparity highlighted the algorithm's failure to provide an impartial assessment, raising concerns about its fairness and accuracy in judicial decision-making.

The bias in COMPAS stems from the historical data used to train the algorithm, which reflects systemic racial discrimination and inequalities in the American justice system. This biased data perpetuates the very disparities the tool was intended to mitigate (Neosmart, 2024). Despite the controversy, Northpointe defended the algorithm, arguing that it was functioning as intended and that the observed disparities were due to inherent differences in baseline risk between racial groups (Mesa, 2021).

The revelations about COMPAS sparked a broader debate about the use of algorithms in the justice system and the need for transparency and reform. Critics have called for independent reviews of such tools and the development of clear regulations to ensure that AI-driven decisions are fair and unbiased (Neosmart, 2024), (Mesa, 2021).
One of the problems with assessing COMPAS from a technical point of view is that it is a closed system for which details are not widely available.

3. Types of Bias in COMPAS

The COMPAS algorithm is a notable example of how bias can manifest in predictive tools used in the criminal justice system. Several types of bias have been identified in the COMPAS algorithm, contributing to its controversial nature.

A. Racial Bias

Racial bias is the most prominent issue associated with COMPAS. Studies, such as the one conducted by ProPublica, have shown that the algorithm tends to overestimate the risk of recidivism for Black defendants while underestimating it for white defendants (Neosmart, 2024), (Mesa, 2021). This bias arises from the historical data used to train the algorithm, which reflects systemic racial disparities and discrimination present in the American justice system (Neosmart, 2024). Consequently, the algorithm perpetuates these inequalities instead of mitigating them.

B. Prejudice Bias

Prejudice bias occurs when the training data reflects existing prejudices and societal stereotypes, which become embedded in the algorithm's predictions. In the case of COMPAS, the data used contained inherent racial prejudices, which were then learned by the algorithm, leading to biased outcomes against black defendants (DataCamp, 2022). It's likely this extends to other minorities.

C. Sample Selection Bias

Sample selection bias can occur when the training data is not representative of the entire population. For COMPAS, the data was skewed towards certain demographics, failing to accurately represent the broader population. This can result in an algorithm that performs well for some groups but poorly for others, exacerbating racial disparities in risk assessment (DataCamp, 2022).

D. Measurement Bias

Measurement bias arises from errors in data collection or the metrics used to train the model. In COMPAS, the factors considered for risk assessment, such as prior arrests and socio-economic status, might not accurately capture an individual's true risk of reoffending, especially if these factors are influenced by systemic racial biases (DataCamp, 2022).

These biases highlight the challenges of ensuring fairness and accuracy in algorithmic decision-making, particularly in sensitive areas like criminal justice. Addressing these biases requires a comprehensive approach that includes revising data collection methods, ensuring diverse and representative training datasets, and implementing rigorous fairness checks throughout the algorithm's lifecycle.

4. Data Lifecycle Analysis

The data lifecycle of the COMPAS algorithm involves several stages, each of which presents opportunities for bias to infiltrate the system. Understanding these stages is crucial for identifying where and how biases occur.

A. Data Collection

The initial stage of the data lifecycle involves data collection, where historical data about offenders is gathered. This data is inherently biased due to systemic racial inequalities present in the criminal justice system. For instance, black individuals are disproportionately arrested and convicted, leading to skewed data that reflects these disparities (Neosmart, 2024). The reliance on such biased historical data sets the foundation for the algorithm's biased predictions.

B. Data Processing

During the data processing phase, the collected data is prepared for analysis. This involves cleaning, transforming, and structuring the data. If the processing methods do not account for existing biases, they can perpetuate or even exacerbate these biases. For COMPAS, the processing stage did not adequately address the racial bias present in the raw data, allowing it to persist into the analysis phase (Neosmart, 2024).

C. Data Analysis

In the analysis stage, the processed data is used to train the algorithm. The COMPAS tool uses machine learning techniques to identify patterns and make predictions about recidivism risk. However, because the training data is biased, the algorithm learns and replicates these biases in its predictions. ProPublica's investigation highlighted that Black defendants were more frequently mislabeled as high risk compared to white defendants, indicating a failure in the analysis phase to mitigate bias (Rudra, 2020), (Neosmart, 2024).

D. Data Storage

Data storage involves maintaining the data and the algorithm's outputs. If the storage systems lack transparency and accessibility, it becomes challenging to audit the data for biases. COMPAS is a closed system, meaning its scores and underlying data are not publicly available, which limits the ability to identify and correct biases (Neosmart, 2024).

E. Data Utilization

Finally, the data and predictions are utilized in decision-making processes within the criminal justice system. The biased outputs from COMPAS influence judicial decisions, such as sentencing and parole, perpetuating racial disparities in the justice system (Neosmart, 2024).

Addressing bias in the COMPAS data lifecycle requires interventions at each stage, including collecting more representative data, implementing bias-correction techniques during processing and analysis, and ensuring transparency in data storage and utilization.

5. Policies and Procedures

To address the biases identified in the COMPAS algorithm, several policies and procedures should have been implemented to ensure fair and ethical use of data in the criminal justice system.

A. Data Governance and Ethical Guidelines

Implementing robust data governance frameworks is crucial to prevent biased data from influencing algorithmic outcomes. This includes establishing clear ethical guidelines for data collection, processing, and analysis. Organizations must ensure that the data used to train algorithms like COMPAS is representative and free from systemic biases. This can be achieved by diversifying data sources and including input from communities affected by the algorithm's decisions (Neosmart, 2024), (DataCamp, 2022).

B. Transparency and Accountability

Transparency in algorithmic decision-making is essential for accountability. COMPAS operates as a closed system, meaning its inner workings and data are not publicly accessible. This lack of transparency hinders independent audits and bias detection. Policies should mandate that algorithms used in public sectors, such as the justice system, are open to external review and validation by independent experts. This would allow for continuous monitoring and correction of biases (Neosmart, 2024).

C. Bias Mitigation Techniques

Incorporating bias mitigation techniques during the data processing and analysis phases is vital. These techniques can involve re-weighting training data to counteract imbalances or using fairness constraints in the algorithm's design to ensure equitable outcomes across different demographic groups. Regular audits and updates to the algorithm should be conducted to adapt to changing societal contexts and data distributions (Neosmart, 2024), (DataCamp, 2022).

D. Training and Awareness

Providing training for stakeholders involved in the development and deployment of algorithms is crucial. This includes educating developers, policymakers, and end-users about the potential biases in AI systems and the importance of ethical data use. Raising awareness about the implications of biased algorithms can foster a culture of responsibility and vigilance (DataCamp, 2022).

E. Legal and Regulatory Frameworks

Establishing legal and regulatory frameworks to govern the use of AI in sensitive areas like criminal justice is necessary. These frameworks should outline the standards for fairness, accountability, and transparency that algorithms must meet. Additionally, they should include provisions for redress and correction in cases where biased algorithmic decisions adversely affect individuals (Neosmart, 2024).

By implementing these policies and procedures, organizations can better ensure that tools like COMPAS are used ethically and effectively, minimizing the risk of perpetuating existing biases in the justice system. It is apparent that these were lacking when COMPAS was implemented.

6. Ethical Violations

The use of the COMPAS algorithm in the criminal justice system has raised significant ethical concerns, particularly regarding racial bias and transparency.

One of the primary ethical violations associated with COMPAS is its perpetuation of racial bias. Studies, such as the one conducted by ProPublica, have shown that the algorithm disproportionately labels Black defendants as high risk for recidivism compared to white defendants, even when controlling for other factors. This bias stems from the historical data used to train the algorithm, which reflects systemic racial disparities in the justice system (Neosmart, 2024), (Angwin et al, 2016).

Another ethical issue is the lack of transparency in the COMPAS algorithm. The algorithm operates as a closed system, meaning its methodology and data are not publicly accessible. This lack of transparency prevents independent audits and makes it difficult to assess and correct biases within the algorithm. The proprietary nature of COMPAS also raises concerns about accountability, as the company behind the algorithm, Equivant (formerly Northpointe), has not disclosed how the scores are calculated (Hu, n.d.), (Rudin et al, 2020).

Moreover, the use of COMPAS scores in judicial decisions, such as sentencing and parole, raises ethical questions about due process and fairness. Judges often rely on these scores without fully understanding their limitations, leading to potential automation bias, where machine-generated information is perceived as inherently trustworthy (Hu, n.d.). This can result in harsher sentences for individuals incorrectly classified as high risk, disproportionately affecting marginalized communities (Neosmart, 2024).

To address these ethical violations, it is crucial to implement policies that ensure transparency and accountability in the use of algorithms like COMPAS. This includes making the algorithm's workings publicly available for independent review and incorporating bias mitigation techniques during its development and deployment. Furthermore, there should be legal and regulatory frameworks to govern the use of AI in the justice system, ensuring that decisions are fair and unbiased (Neosmart, 2024), (Rudin et al, 2020).

7. Proposed Solutions

To address the biases and ethical concerns associated with the COMPAS algorithm, several solutions have been proposed to improve its fairness, accuracy, and transparency.

A. Increasing Transparency

One of the primary issues with COMPAS is its lack of transparency. The algorithm operates as a closed system, making it difficult for independent experts to review and understand its decision-making process. To remedy this, there should be a mandate for transparency in algorithmic systems used in the justice system. This would involve making the algorithm's methodology and data public, allowing for independent audits and assessments to identify and correct biases (Neosmart, 2024), (Engel et al, 2024), (Rudin et al, 2020).

B. Implementing Bias Mitigation Techniques

Bias mitigation techniques can be employed during the data processing and analysis phases to reduce racial bias in COMPAS. This includes re-weighting training data to counteract imbalances and applying fairness constraints in the algorithm's design to ensure equitable outcomes across different demographic groups. Regular audits and updates to the algorithm are necessary to adapt to changing societal contexts and data distributions (Engel et al, 2024).

C. Legal and Regulatory Frameworks

Establishing legal and regulatory frameworks is crucial to govern the use of AI in sensitive areas like criminal justice. These frameworks should outline standards for fairness, accountability, and transparency that algorithms must meet. They should also include provisions for redress and correction in cases where biased algorithmic decisions adversely affect individuals. This would ensure that algorithms like COMPAS are used ethically and effectively (Neosmart, 2024), (Park, 2019).

D. Training and Awareness

Providing training for stakeholders involved in the development and deployment of algorithms is essential. This includes educating developers, policymakers, and end-users about the potential biases in AI systems and the importance of ethical data use. Raising awareness about the implications of biased algorithms can foster a culture of responsibility and vigilance (Yong, 2019).

E. Developing Alternative Models

Research has shown that simple machine learning models can achieve similar accuracy to COMPAS while potentially reducing bias. Developing alternative models that prioritize fairness and accuracy could provide more equitable outcomes in the justice system. These models should be rigorously tested and validated to ensure they meet ethical and legal standards (Engel et al, 2024), (Yong, 2019).

By implementing these solutions, the justice system can better ensure that tools like COMPAS are used in a manner that is fair, transparent, and free from bias.

8. Conclusion

The deployment of COMPAS in the criminal justice system highlights significant ethical and fairness challenges associated with algorithmic decision-making. While COMPAS was designed to enhance judicial efficiency and objectivity, its implementation has revealed deep-seated biases and a lack of transparency that compromise these goals. The biases identified in COMPAS—ranging from racial disparities to issues of transparency and accountability—reflect broader societal inequalities that are magnified within algorithmic frameworks.

To address these challenges, a multifaceted approach involving legislative action, technological advancements, and ethical oversight is essential. Enhancing algorithmic transparency, conducting independent audits, implementing robust data governance, and integrating bias mitigation strategies are critical steps toward mitigating the adverse effects of such predictive tools. Furthermore, legal reforms and comprehensive training for all stakeholders in the judicial system are necessary to ensure that these tools are used responsibly and ethically.

Ultimately, the goal should be to foster a criminal justice system that uses technology to enhance fairness and justice, rather than perpetuate existing disparities. This requires continuous evaluation and adaptation of the tools we deploy, guided by a commitment to fairness, accountability, and transparency in algorithmic decision-making.

References:

Angwin, Julia, Larson, Jeff, Mattu, Surya Kirchner, Lauren (2016); *Machine Bias*; ProPublica https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Neosmart (unattributed, 2024); *Bias in the COMPAS algorithm: A threat to justice in the United States*; https://neosmart.ai/bias-in-the-compas-algorithm-a-threat-to-justice-in-the-united-states/

Mesa; Natalia (2021); *Can the criminal justice system's artificial intelligence ever be truly fair?*; https://massivesci.com/articles/machine-learning-compas-racism-policing-fairness/

DataCamp (unattributed, 2022); *Data Demystified: The Different Types of AI Bias*; https://www.datacamp.com/blog/data-demystified-the-different-types-of-ai-bias

Rudra, Atri (2020); *Algorithms and Society Notes: COMPAS dataset*; https://www-student.cse.buffalo.edu/~atri/algo-and-society/support/notes/compas/index.html

Hu, Xinlan Emily (no date); *137 Questions: Criminal Risk Assessment Algorithms as a Case Study for Data Ethics*; https://stanfordrewired.com/post/137-questions

Rudin, Cynthia, Wang, Caroline, Coker, Beau (2020); *The Age of Secrecy and Unfairness in Recidivism Prediction*; https://hdsr.mitpress.mit.edu/pub/7z10o269/release/7

Engel, Christoph, Linhardt, Lorenz, Schubert, Marcel (2024); *Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism*; https://link.springer.com/article/10.1007/s10506-024-09389-8

Park, Andrew Lee (2019); *Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing*; UCLA Law Review https://www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/

Yong, Ed (2018); *A Popular Algorithm Is No Better at Predicting Crimes Than Random People*; The Atlantic Magazine https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/