PART 3: PRACTICAL AUDIT - COMPAS RECIDIVISM BIAS ANALYSIS

300-Word Bias Audit Report

EXECUTIVE SUMMARY

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) recidivism prediction system exhibits significant racial bias that systematically disadvantages Black defendants. This audit quantifies disparities in false positive rates and recommends remediation strategies to improve fairness without sacrificing predictive accuracy.

KEY FINDINGS

False Positive Rate Disparity (Primary Finding):

The system incorrectly classifies non-reoffending defendants as high-risk at substantially different rates across racial groups:

- Black defendants: ~45% false positive rate
- White defendants: ~23% false positive rate
- Disparate impact ratio: 1.96 (nearly 2x higher false positive rate for Black defendants)

This disparity means Black individuals are almost twice as likely to be labeled "high-risk" despite having similar actual recidivism outcomes as white counterparts. In practical terms, among 100 Black defendants who do not reoffend, 45 are incorrectly flagged as dangerous; for white defendants, only 23 are misclassified.

False Negative Rate Concerns:

Conversely, the system fails to identify reoffending risk in white defendants at higher rates (~20% FNR) compared to Black defendants (~15% FNR). This asymmetry indicates the model systematically overestimates white defendant safety and underestimates Black defendant safety.

Compounding Effects:

These disparities create cascading harms: Black defendants receive harsher sentences, higher bail amounts, and reduced parole eligibility based on algorithmically inflated risk assessments. Over time, these decisions feed into retraining data, reinforcing biases in future model iterations.

ROOT CAUSES

1. **Training Data Bias**: Historical overpolicing of Black communities resulted in overrepresentation in criminal databases
2. **Proxy Variables**: The model uses features (neighborhood, prior arrests, education) correlated with race
3. **Model Optimization**: System optimized for aggregate accuracy, not fairness across demographic groups

---

REMEDIATION STEPS

Immediate (0-3 months):

1. Discontinue use of COMPAS as sole decision-making tool; require human review for all high-risk classifications
2. Audit all prior sentences influenced by disparate COMPAS scores; identify cases for appeal
3. Implement bias monitoring dashboard tracking false positive rates by race

Short-term (3-6 months):

1. Retrain model with fairness constraints (equalized odds objective)
2. Remove correlated proxy variables (zip code, prior arrests that reflect systemic bias)
3. Achieve disparate impact ratio ≥0.85 before any future deployment

Long-term (6+ months):

1. Develop alternative risk assessment methods not dependent on historical criminal data
2. Involve affected communities and legal advocacy in system redesign
3. Establish independent algorithmic audit board with judicial oversight

CONCLUSION

COMPAS demonstrates how AI amplifies systemic biases at scale. Technical fixes alone are insufficient; remediation requires legal reform, community involvement, and a fundamental commitment to equitable criminal justice. Until fairness metrics are achieved and human oversight guaranteed, the system should not influence sentencing decisions.