OVERVIEW

- Synthetic dataset audit: Employment prediction (income > R50,000/year).
- Focus: Gender (Female/Male) and race (Non-White/White) biases in SA.
- Simulated IBM AI Fairness 360 for metrics and mitigations (McKinsey Global Institute, 2025).
- SA Context: 33.2% unemployment (Statistics SA, 2025), Gini 0.63 (Coetzer & van Zyl, 2024), BEE goals (Govender, 2024).

KEY BIAS PATTERNS DISCOVERED
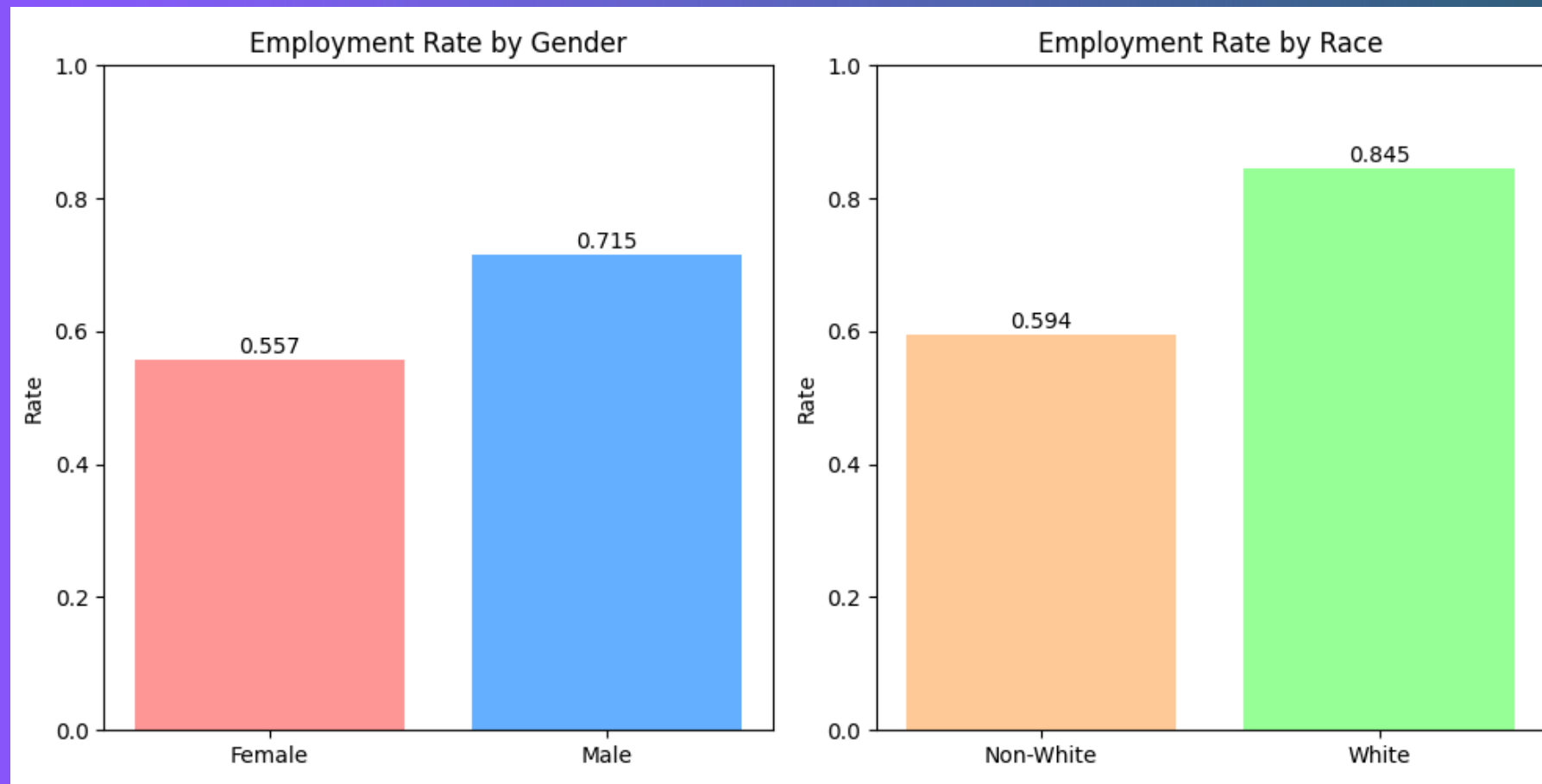
- Dataset: 2,000 samples (80% non-White, 51% female, Stats SA-based) (Coetzer & van Zyl, 2024).
- Gender: Females 16% lower employment rate (0.557 vs. 0.715) (Makhafola, 2024).
- Race: Non-Whites 25% lower (0.594 vs. 0.845) (Makhafola, 2024).
- Validated: Chi-squared ($p < 3.69e-13$ gender, $p < 1.17e-19$ race).
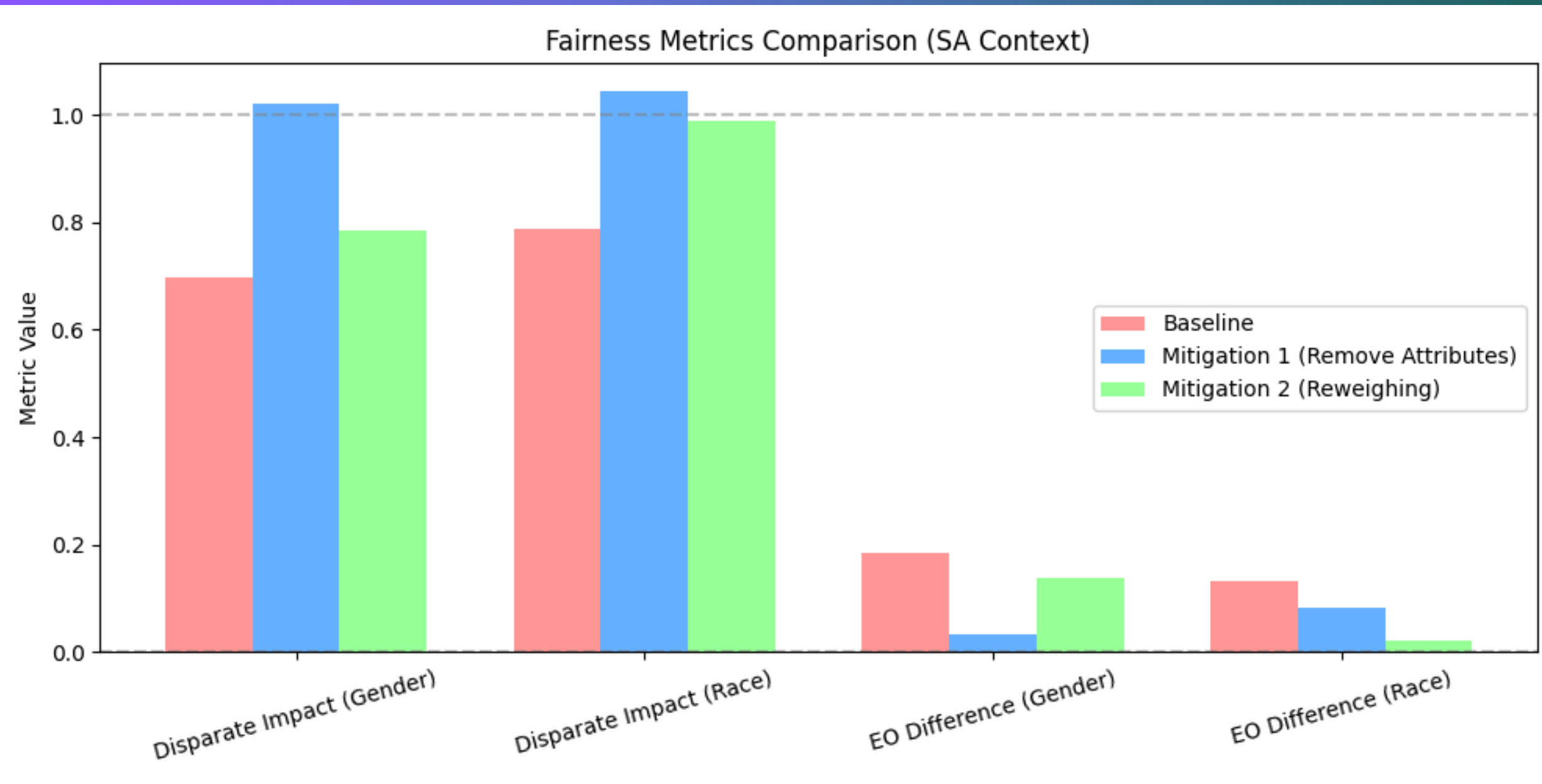
BASELINE FAIRNESS METRICS

- Accuracy: 67.75% (logistic regression) (McKinsey Global Institute, 2025).
- Disparate Impact (DP, ideal=1): Gender 0.696, Race 0.788.
- Equal Opportunity Difference (EO, ideal=0): Gender 0.185, Race 0.132.
- Equalized Odds: TPR/FPR diffs show bias (Makhafola, 2024).

Baseline Accuracy: 0.6775

Baseline Fairness Metrics:

	Metric	Gender	Race
0	Disparate Impact (DP)	0.696	0.788
1	EO Difference	0.185	0.132
2	EOdds TPR Diff	0.185	0.132
3	EOdds FPR Diff	0.416	0.319

MITIGATION STRATEGIES IMPLEMENTED

Fairness Metrics Comparison (SA Context)

Legend:
- Baseline
- Mitigation 1 (Remove Attributes)
- Mitigation 2 (Reweighing)

X-axis categories: Disparate Impact (Gender), Disparate Impact (Race), EO Difference (Gender), EO Difference (Race)

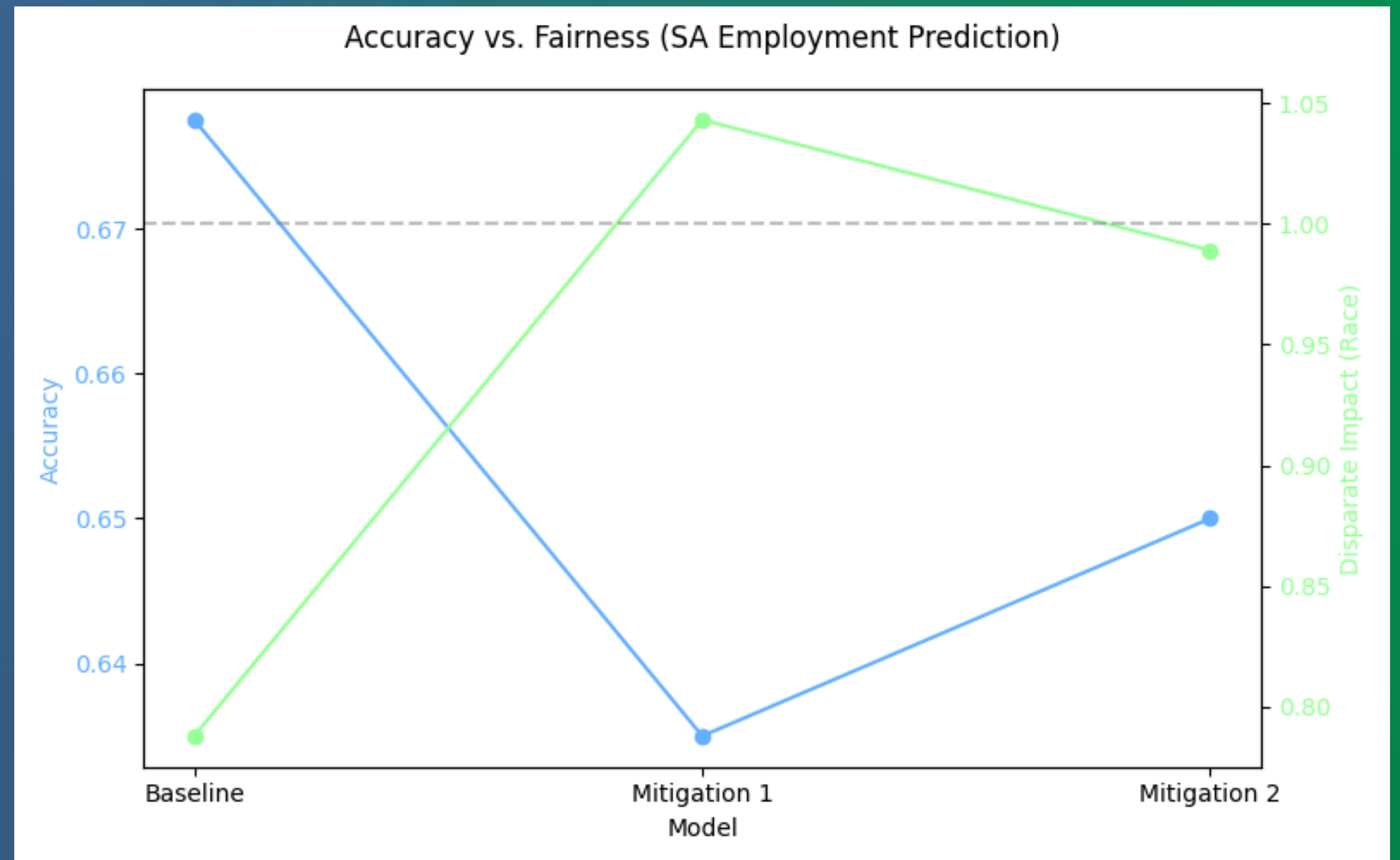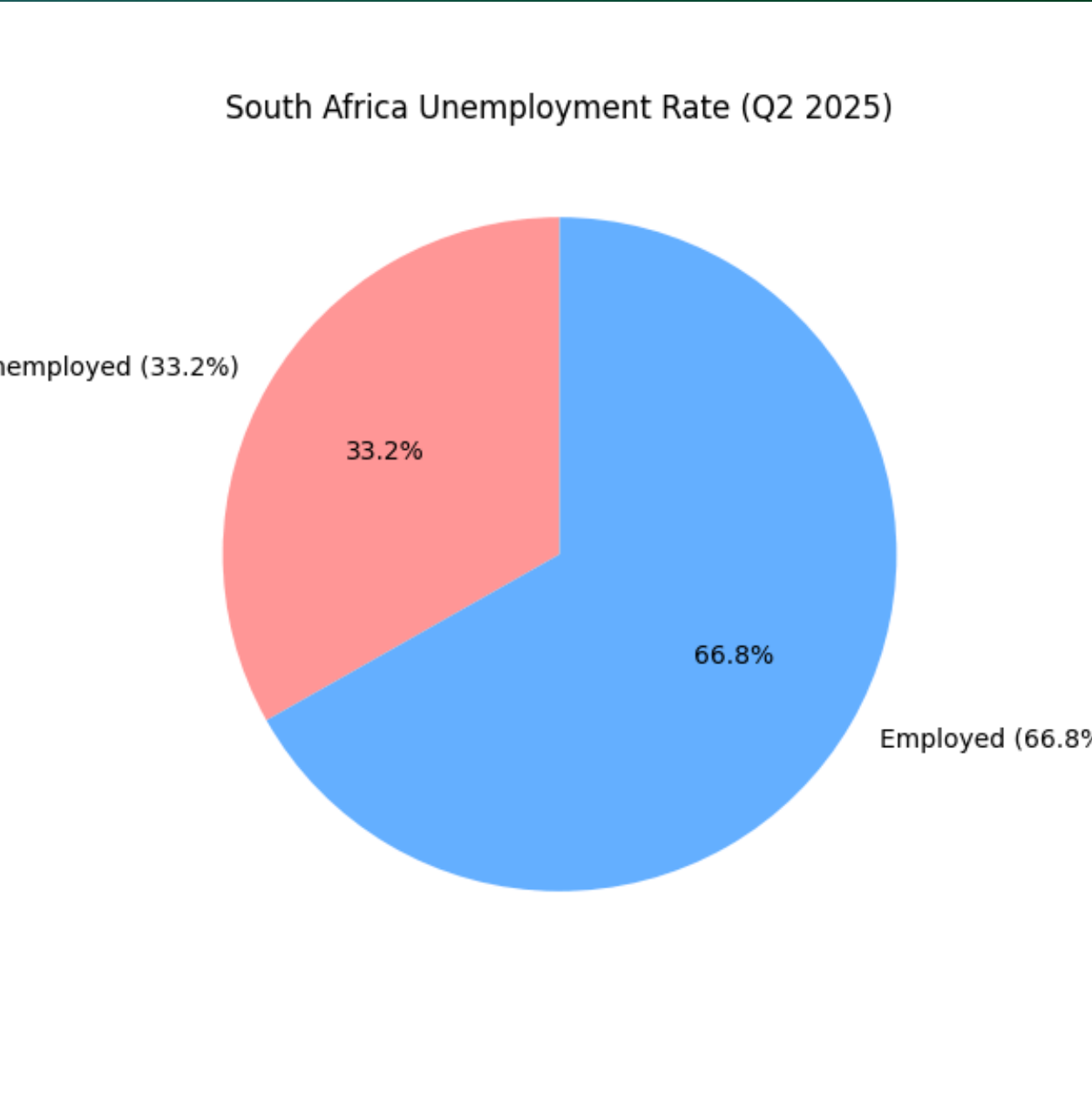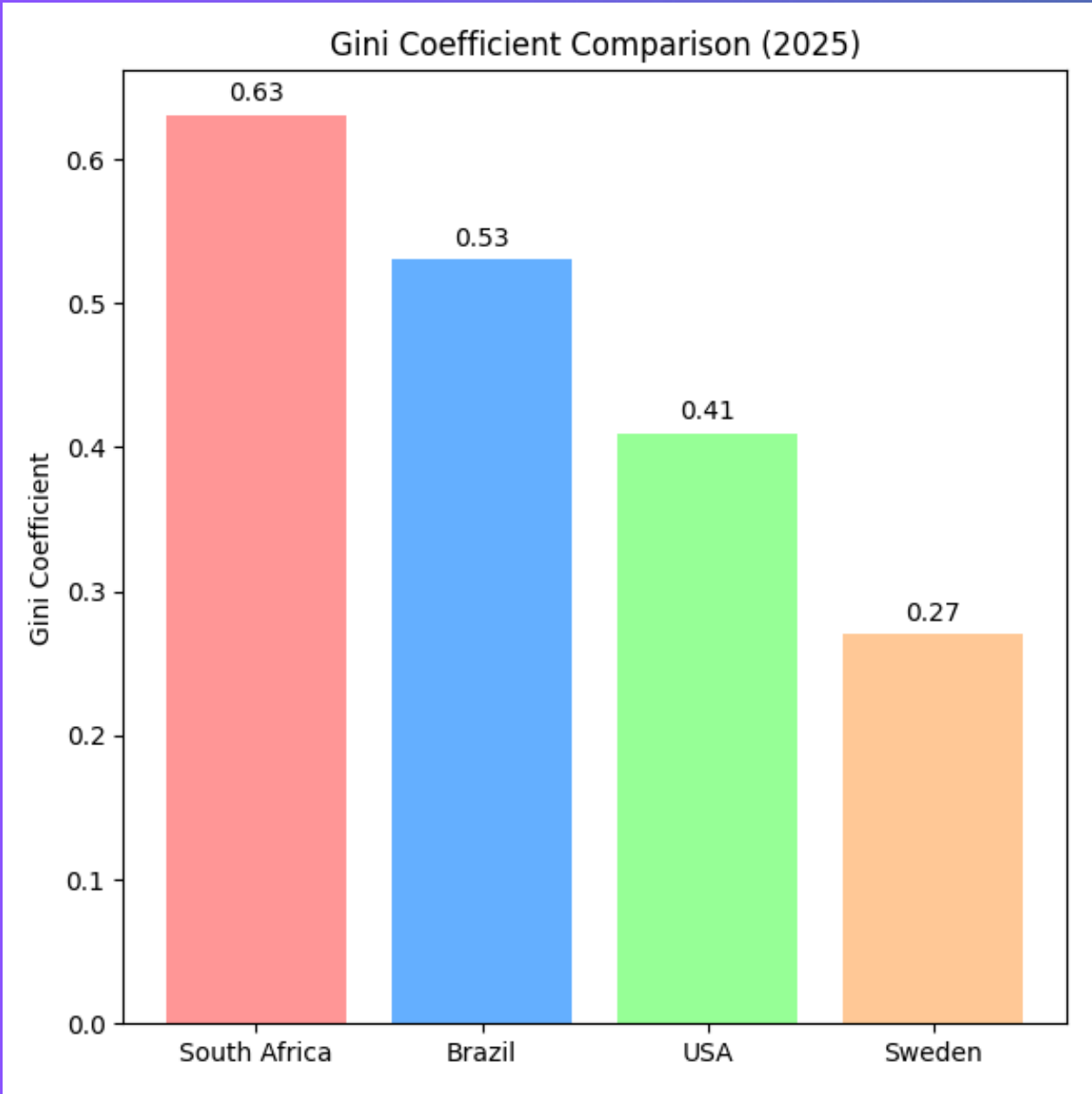Y-axis: Metric Value

- Simulated IBM AI Fairness 360 (McKinsey Global Institute, 2025).
- Preprocessing: Removed gender/race (DP: 1.020 Gender, 1.043 Race).
- Reweighing: Adjusted race weights (DP: 0.784 Gender, 0.989 Race).
- Accuracy: 63.50-65.00%, fairness improved (Makhafola, 2024).

ACCURACY VS. FAIRNESS TRADE-OFF

- Baseline: Accuracy 67.75%, DP (Race) 0.788.
- Mitigation 1: Accuracy 63.50% (-4.25%), DP 1.043.
- Mitigation 2: Accuracy 65.00% (-2.75%), DP 0.989.
- Fairness prioritized for SA equity (Govender, 2024).

Accuracy vs. Fairness (SA Employment Prediction)

REAL-WORLD IMPLICATIONS

South Africa Unemployment Rate (Q2 2025)

Unemployed (33.2%) — 33.2%
Employed (66.8%) — 66.8%

Gini Coefficient Comparison (2025)

	Gini Coefficient
South Africa	0.63
Brazil	0.53
USA	0.41
Sweden	0.27

- Risks job denials, worsening 33.2% unemployment (Statistics SA, 2025).
- Violates Employment Equity Act, widens Gini (0.63) (Coetzer & van Zyl, 2024).
- Undermines BEE, perpetuates apartheid inequities (PwC South Africa, 2023).
- Harms: Poverty, social unrest (Ndung'u & Signé, 2024).

RECOMMENDATIONS & ETHICS

- Developers: Balance data, use skills-based features (McKinsey Global Institute, 2025).
- Policymakers: Mandate audits, align with SA AI Policy (UNESCO, 2024).
- Ethics: Accountability, inclusivity, fairness (see Ethics_Statement.docx) (Ndung'u & Signé, 2024).
- Q&A: How to ensure fair AI in SA?

Accountability

Inclusivity

Fairness

Monitoring

THANK YOU!

FOR YOUR ATTENTION