From Predictive Pricing to Fair Pricing: A Causal Fairness Audit Framework for Insurance Models

Manuel Caccone

Senior Actuary | *IAA AI Task Force*

Insurance Data Science Conference
9 - 10 June 2026
House of Insurance, Leibniz Universität Hannover

25 MINUTES AGENDA

⚠️ **The Problem — why fairness in pricing matters now**

⚖️ **Fairness Criteria & Luck Egalitarianism**

🔍 **Phase 1 — Calibration Audit**

🔎 **Phase 2 — SHAP Proxy Detection**

🔄 **Phase 3 — Counterfactual Fairness**

✅ **Audit Verdict & Recommendations**

🇪🇺 **EU AI Act Mapping & Contribution**

THE PROBLEM

THREE TYPES OF DISCRIMINATION IN PRICING MODELS

🚨 **Direct discrimination —** the model explicitly uses protected attributes like age or gender as rating factors. Detectable, but still present in legacy systems.

🕵️ **Indirect (proxy) discrimination —** seemingly neutral features carry protected information. BonusMalus proxies for age, postal code for ethnicity. Much harder to detect.

🔁 **Systemic discrimination —** historical data encodes past societal biases. Even a correctly specified model will reproduce unfair patterns.

⚖️ **EU AI Act (Art. 10) —** now mandates bias examination for high-risk AI. Insurance pricing falls under Annex III.

FAIRNESS CRITERIA

THREE STATISTICAL CRITERIA — AND WHY WE CHOOSE SUFFICIENCY

🎯 **Sufficiency (Calibration) —** same predicted premium = same expected loss, regardless of group. The actuarial gold standard.

⚖️ **Separation (Equalized Odds) —** equal error rates across groups. Common in credit scoring, less natural for pricing.

📊 **Independence (Demographic Parity) —** identical premium distributions across groups. Strongest criterion but conflicts with risk-based pricing.

🔴 **Chouldechova (2017) —** proved these cannot hold simultaneously. We choose sufficiency: it aligns with premium adequacy.

OPTION LUCK VS BRUTE LUCK

DWORKIN (1981) — AN ETHICAL LENS FOR RATING FACTORS

🎲 **Option luck —** arises from deliberate choices: driving behavior, vehicle type, deductible.

✅ **Ethically acceptable to price on these factors.**

🧬 **Brute luck —** arises from circumstances beyond control: age, gender, genetics.

❌ **Pricing on brute luck penalizes people for things they cannot change.**

🧩 **The BonusMalus puzzle**

Looks like option luck (driving record)…

…but correlates r = −0.48 with age

Young drivers get high BM simply because they lack years of experience.

⚠️ **BM converts brute luck into apparent option luck.**

DATASET & MODELS

FREMTPL2FREQ — FRENCH MOTOR THIRD-PARTY LIABILITY

🗃️ **Dataset — 678,013 policies** · CASdatasets R package

🎯 **Target: claim frequency** · 🛡️ Protected attribute: driver age (6 bands: 18-25 → 65+)

📐 **Model 1 — GLM Poisson**

Log-link · 7 covariates · Transparent and interpretable

→ Traditional actuarial benchmark

🌲 **Model 2 — XGBoost Poisson**

500 rounds · max_depth = 6 · Captures non-linear interactions

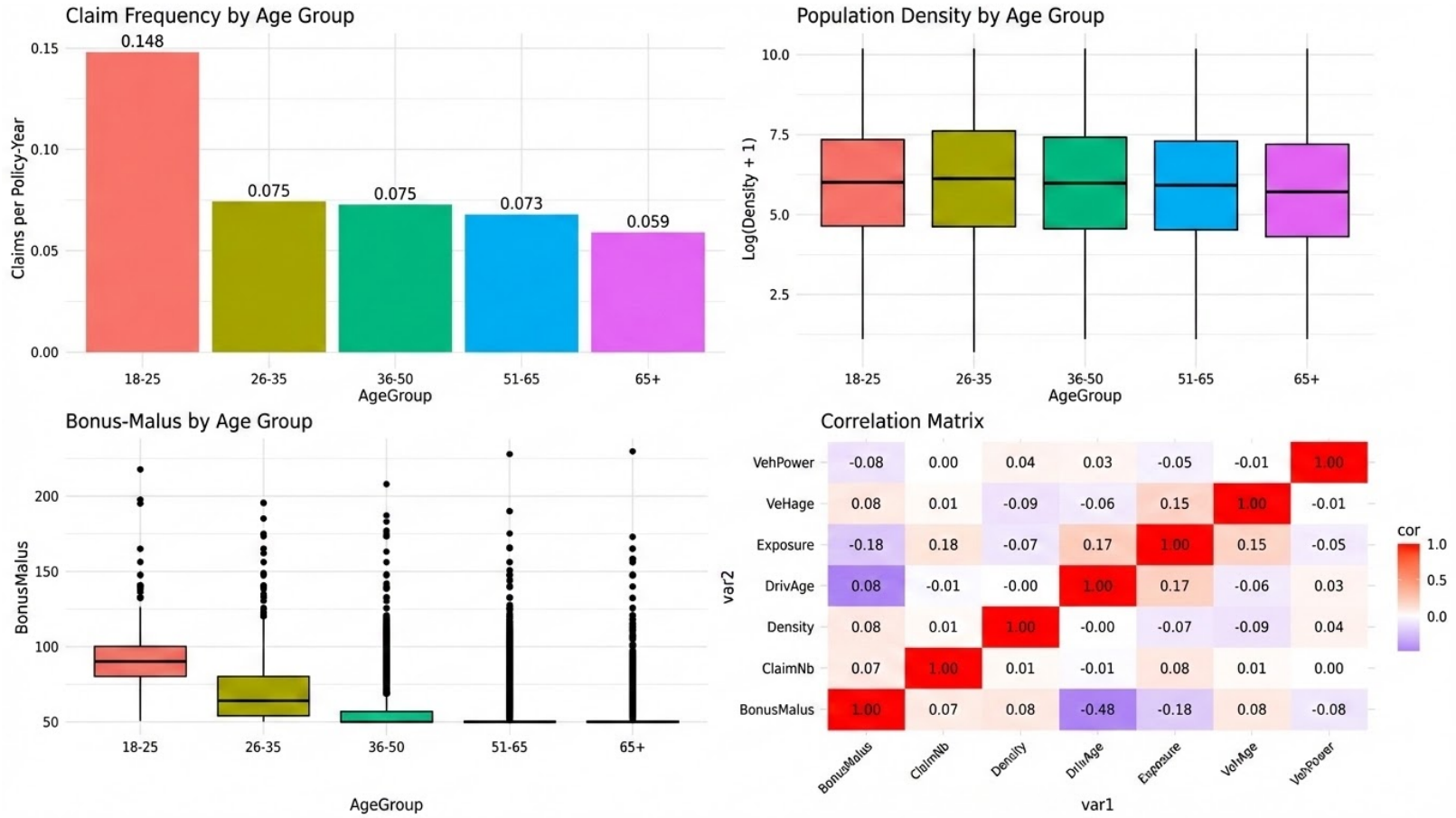
→ Less transparent, higher flexibility

⚠️ **Both models include BonusMalus — deliberately.**

We test whether BM creates a proxy discrimination channel.

EDA OVERVIEW

Age concentrated in the 35-55 band with fewer young and elderly drivers. BonusMalus has a **median of 54** (50 = best), with a heavy right tail driven almost entirely by young drivers. Claim frequency is highest for young drivers and decreases with age.



KEY INSIGHT: BONUSMALUS AS AGE PROXY

📉 **BonusMalus × Age correlation: r = −0.48 — strong negative association.**

🔴 **Young drivers (18-25) → mean BM = 97**

Not due to bad driving — simply no time to accumulate no-claims bonuses.

🟢 **Middle-aged drivers (46-55) → mean BM = 53 (near minimum)**

⚠️ **The hidden channel:**

A model using BonusMalus is implicitly using age — even if age is excluded from features.

🎯 Audit goal: quantify how much of BM's predictive power is

→ ✅ legitimate risk signal

→ ❌ age proxy

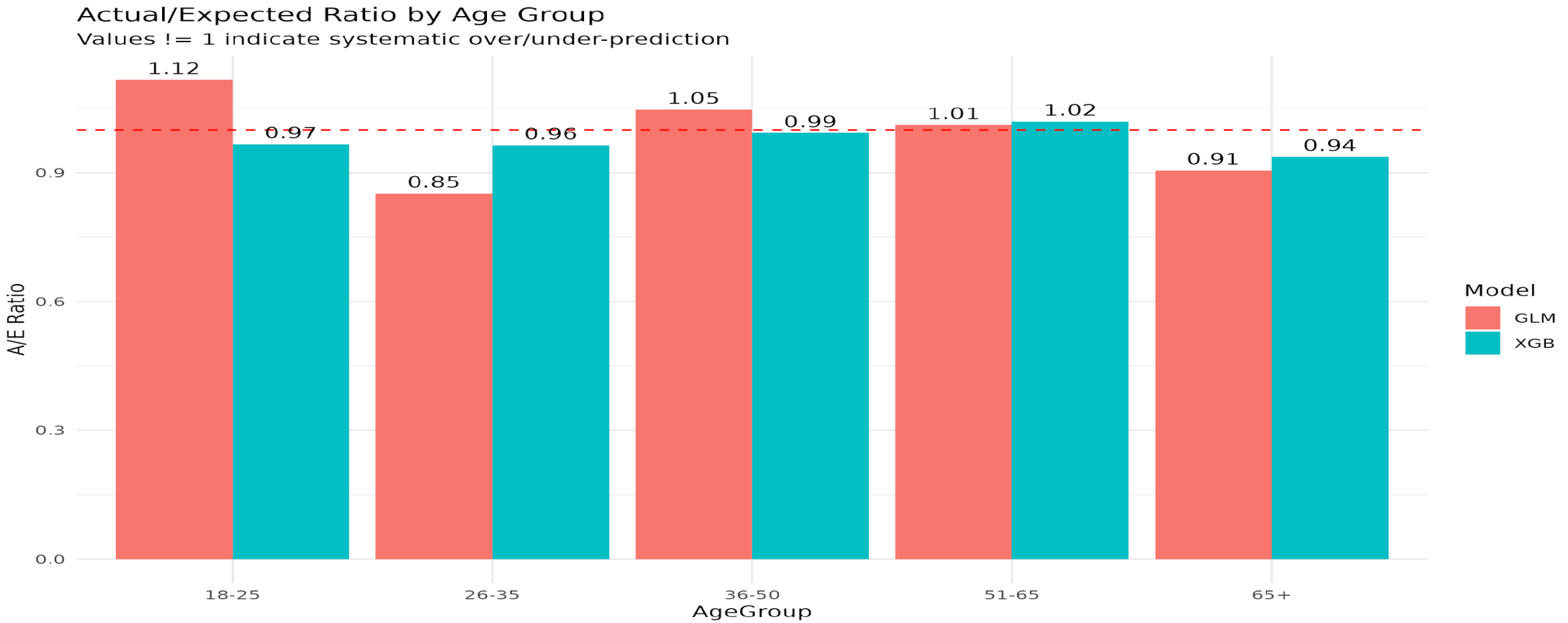
PHASE 1

Calibration Audit

A/E ratios and calibration slopes by age group

A/E RATIOS BY AGE GROUP

The GLM shows **wide deviations (0.85 to 1.12)**, over- and under-predicting across age groups. The XGBoost model achieves much **tighter calibration (0.94 to 1.02)**, capturing non-linear age effects that the GLM cannot represent.

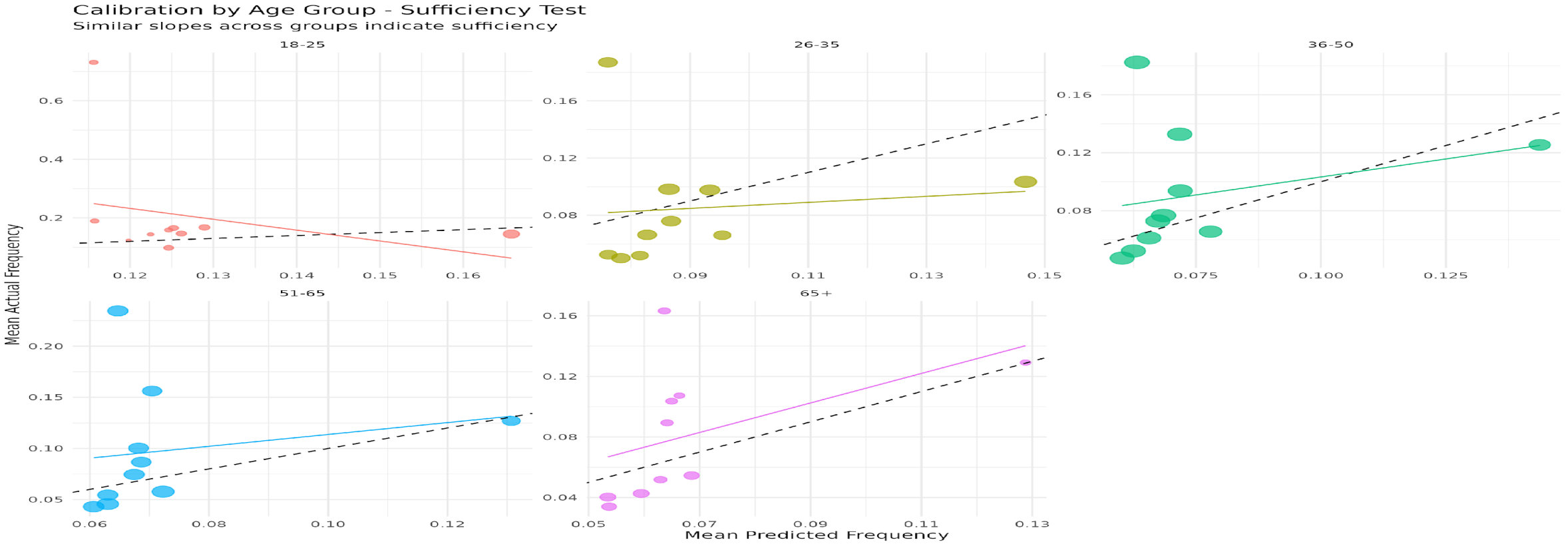


CALIBRATION SLOPE TEST

We regress actual claims on predicted claims within each age group. Perfect calibration requires slope = 1.

The 65+ group has a slope of 0.17 in the GLM — catastrophic miscalibration. Predictions are essentially meaningless for elderly drivers.

All age groups fail the calibration test for the GLM at 95% CI. XGBoost is better but still fails for 65+.



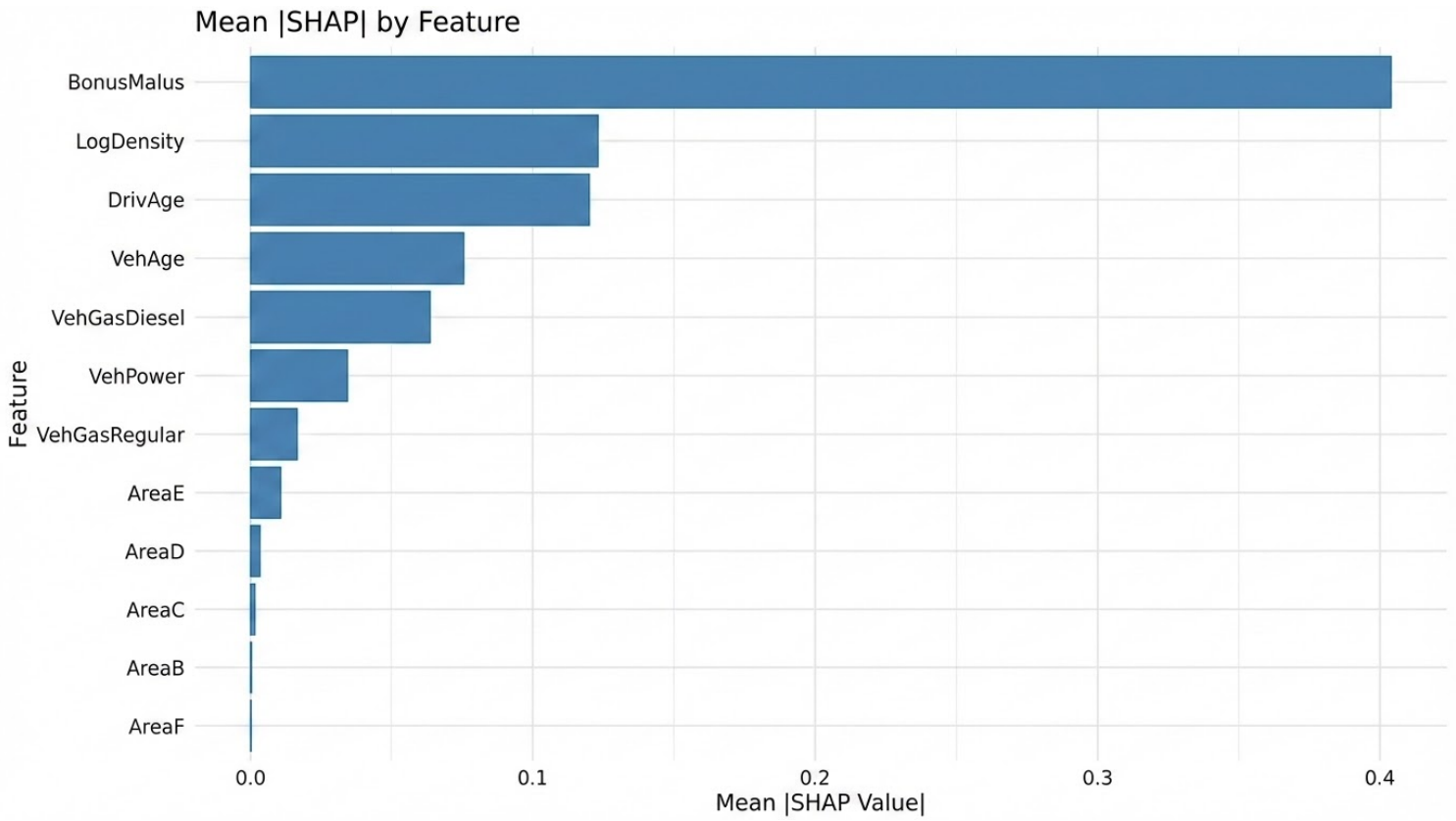
PHASE 2

SHAP Proxy Detection

Identifying hidden discrimination channels

SHAP FEATURE IMPORTANCE

- 🏆 BonusMalus — 47% total SHAP importance · dominant feature by far
- 🚗 Vehicle power — 15%
- 🏙️ Density — 12%
- 🗺️ Region — 10%
- 👤 Driver age (direct) — 8%



SHAP BEESWARM

🔴 High BM (young drivers) →

predictions pushed strongly upward

🔵 Low BM (experienced drivers) →

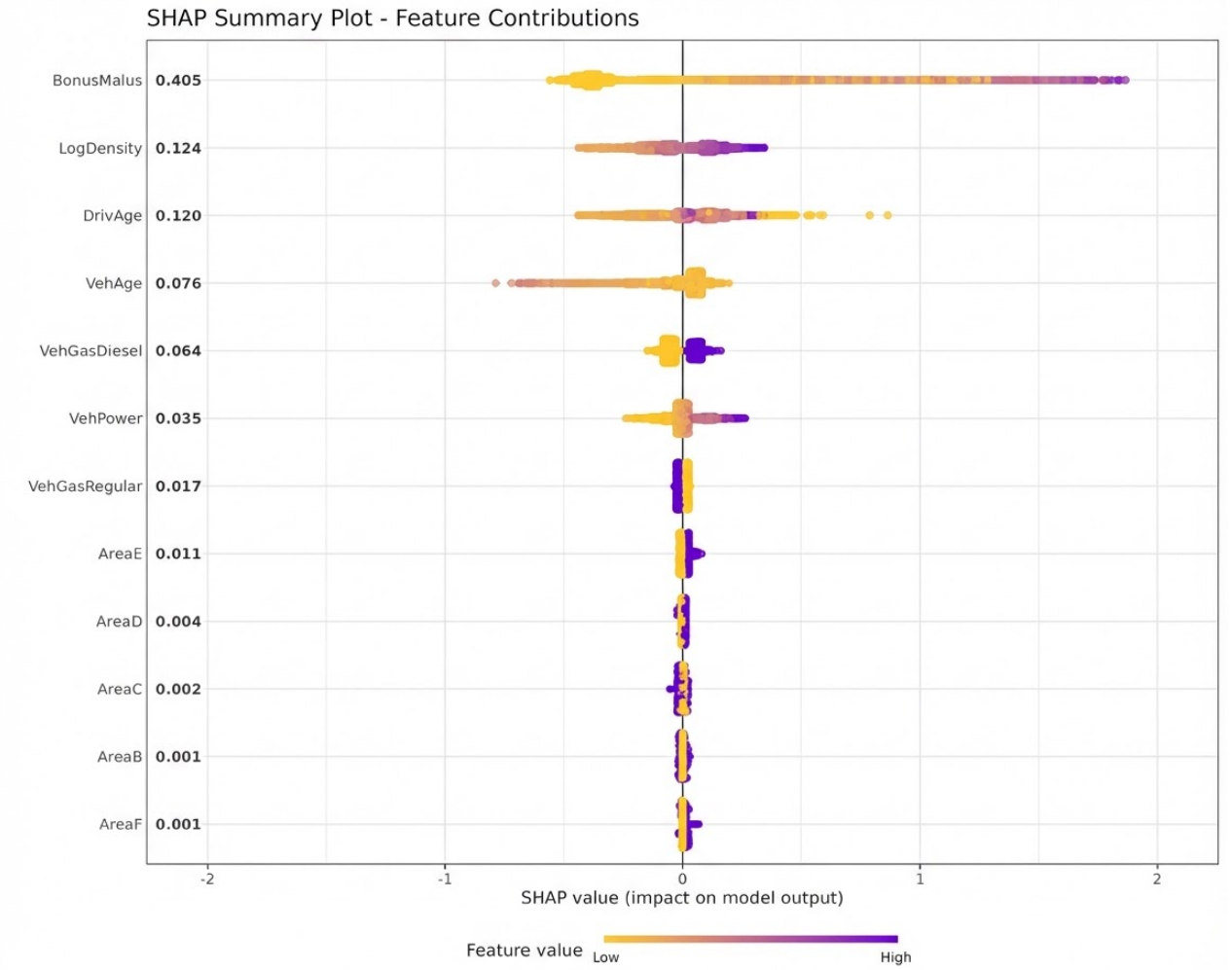
modest downward effect

⚠️ **The effect is asymmetric:**

BM penalizes young drivers more than it rewards experienced ones.

📍 Each dot = one policy ·

Yellow = high feature value
Violet = low



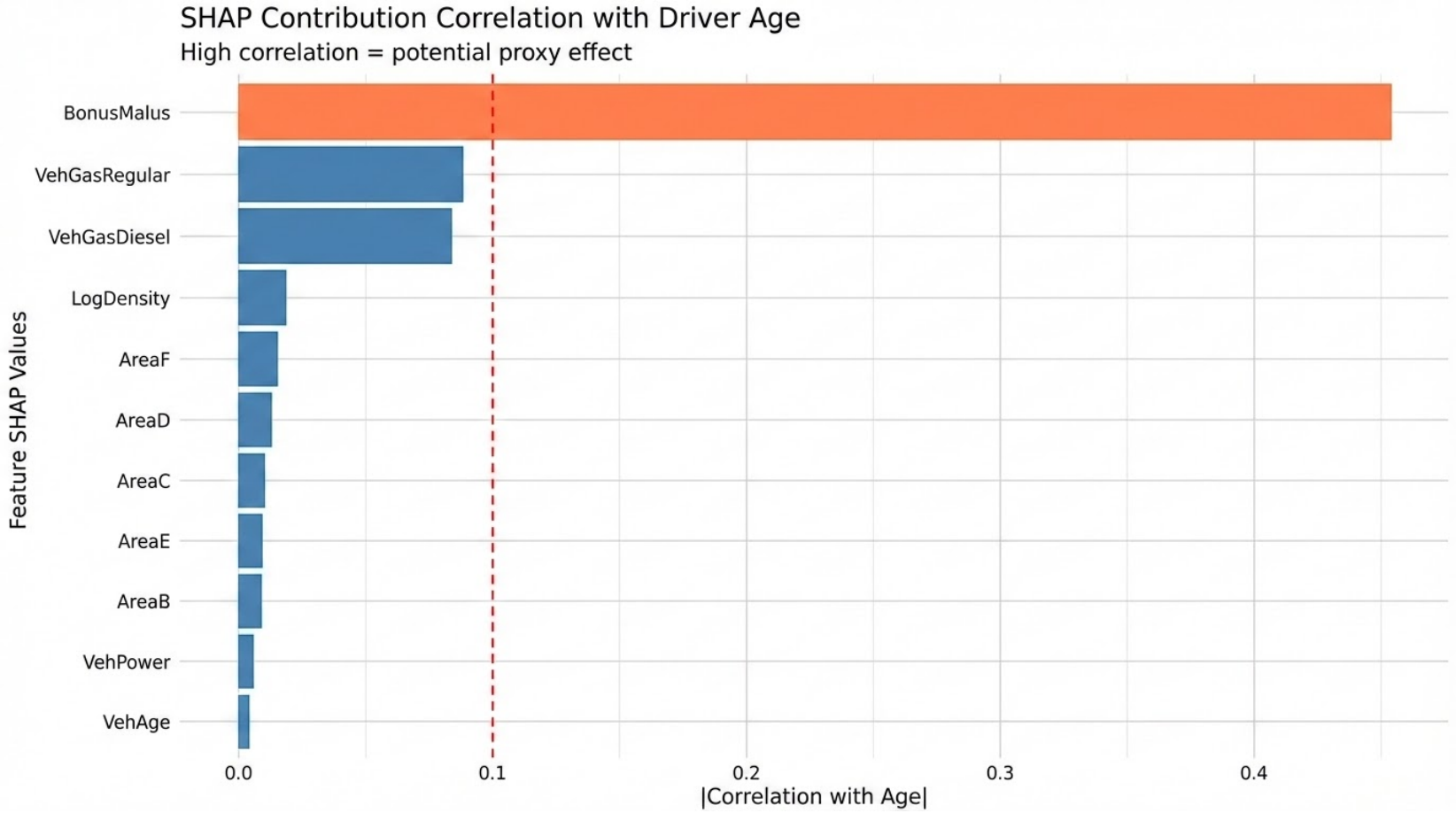
PROXY DETECTION: SHAP-AGE CORRELATION

📊 BM SHAP × Age correlation: r = 0.45

45% of BM's predictive impact covaries with the protected attribute.

- 🔴 BonusMalus — proxy correlation: 0.45 ← dominant channel
- 🟡 VehPower — 0.12
- ⚪ Region / Density — ≈ 0.00

⚠️ **BM is the primary indirect path through which age enters the model.**



SHAP BY LUCK TYPE

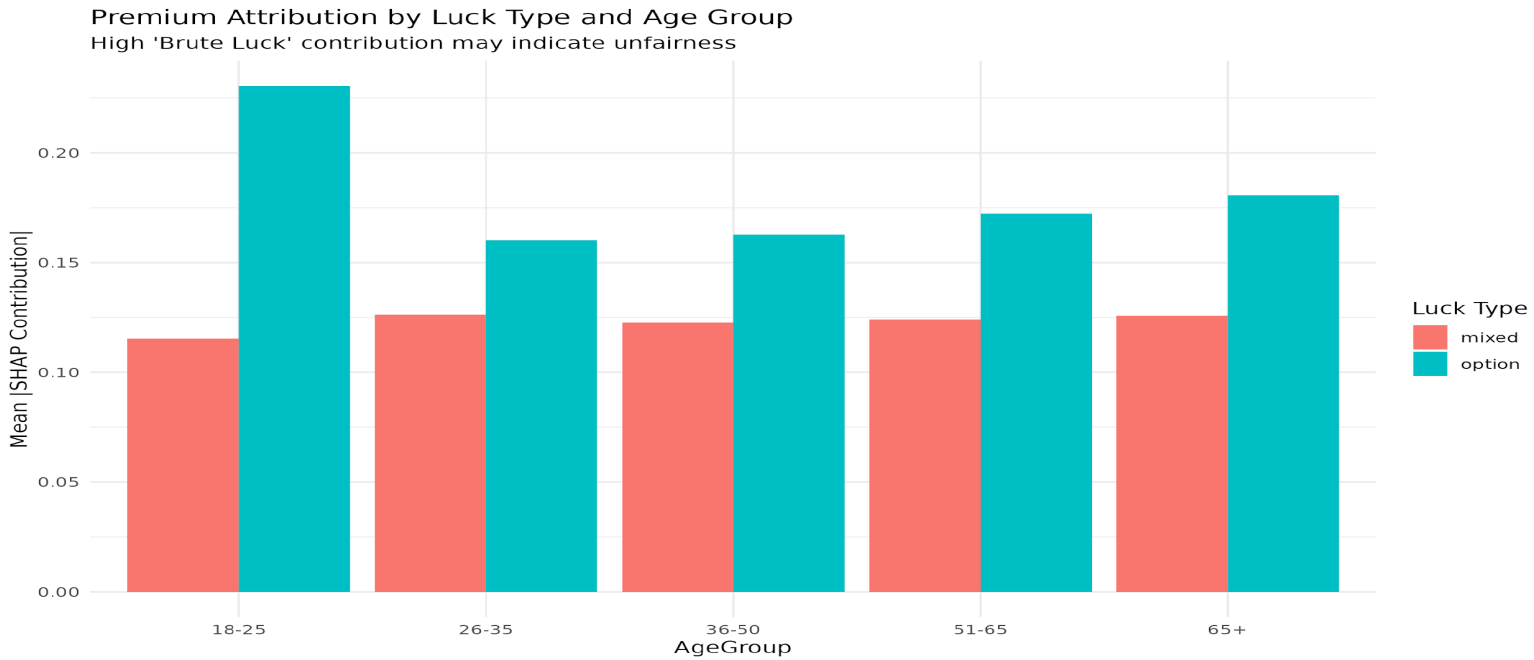
⚖️ **Feature classification:**

🟢 **Option luck** *(choices)* — VehPower, Density, Region…

🔴 **Brute luck** *(circumstances)* — Age, Gender…

🟠 **BonusMalus** — classified as option luck, but

proxy evidence says otherwise



⚠️ **If BM → reclassified as brute luck:** Brute luck exceeds **55% of total SHAP importance**

🎯 **Luck egalitarianism verdict:**

Pricing should reflect **choices**, not **circumstances**. The majority of differentiation is currently driven by factors beyond policyholder control.

PHASE 3

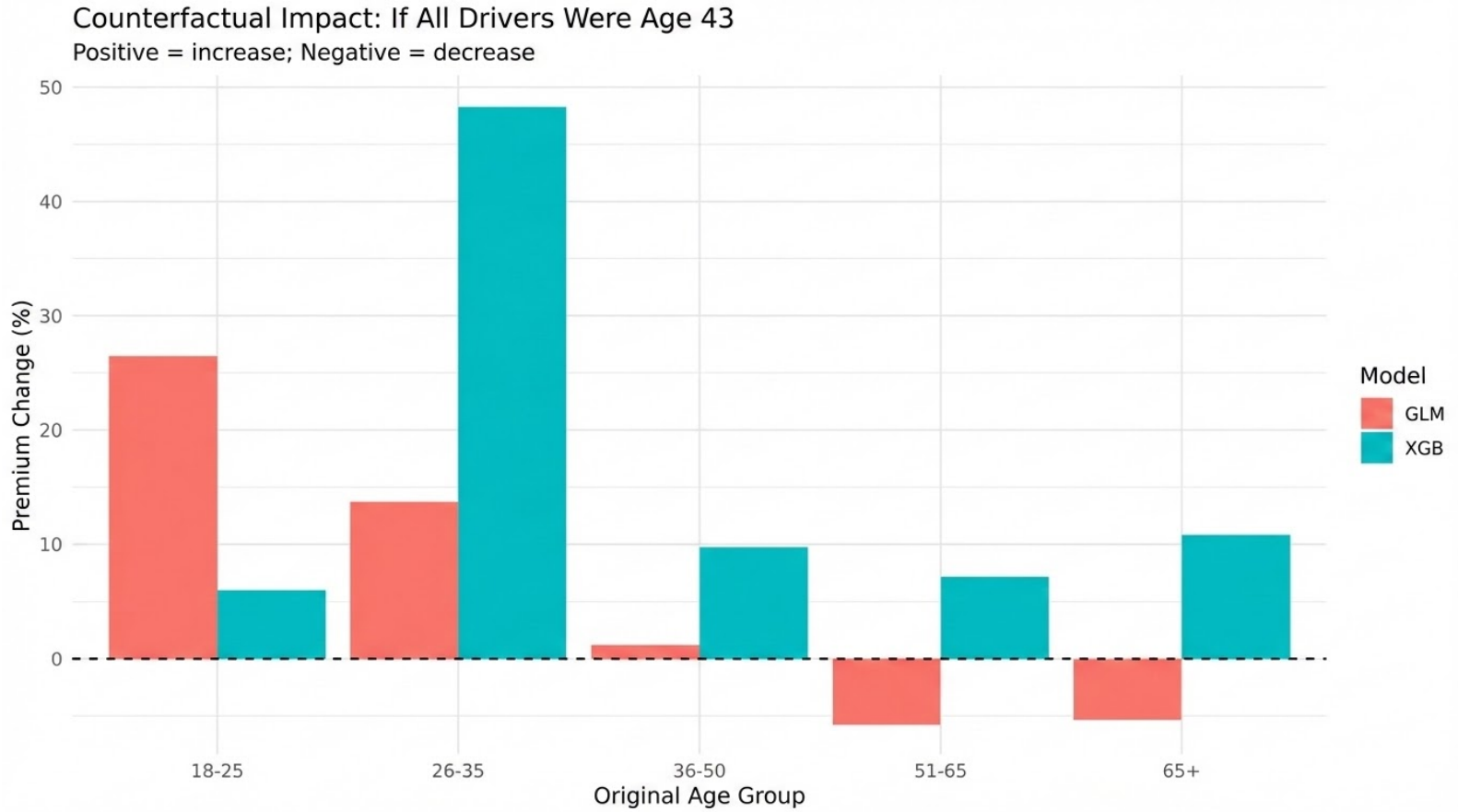
Counterfactual Fairness

Measuring the causal impact of age via do-calculus

COUNTERFACTUAL IMPACT OF AGE

We set all policyholders to the reference age band (36-45) and measure the prediction change.

The GLM shows a **+27% premium increase for the 18-25 group** due to age alone. XGBoost produces an even larger effect: **+48% for the 26-35 group**. Elderly drivers (65+) see a 15-22% decrease — effectively subsidized by younger drivers.

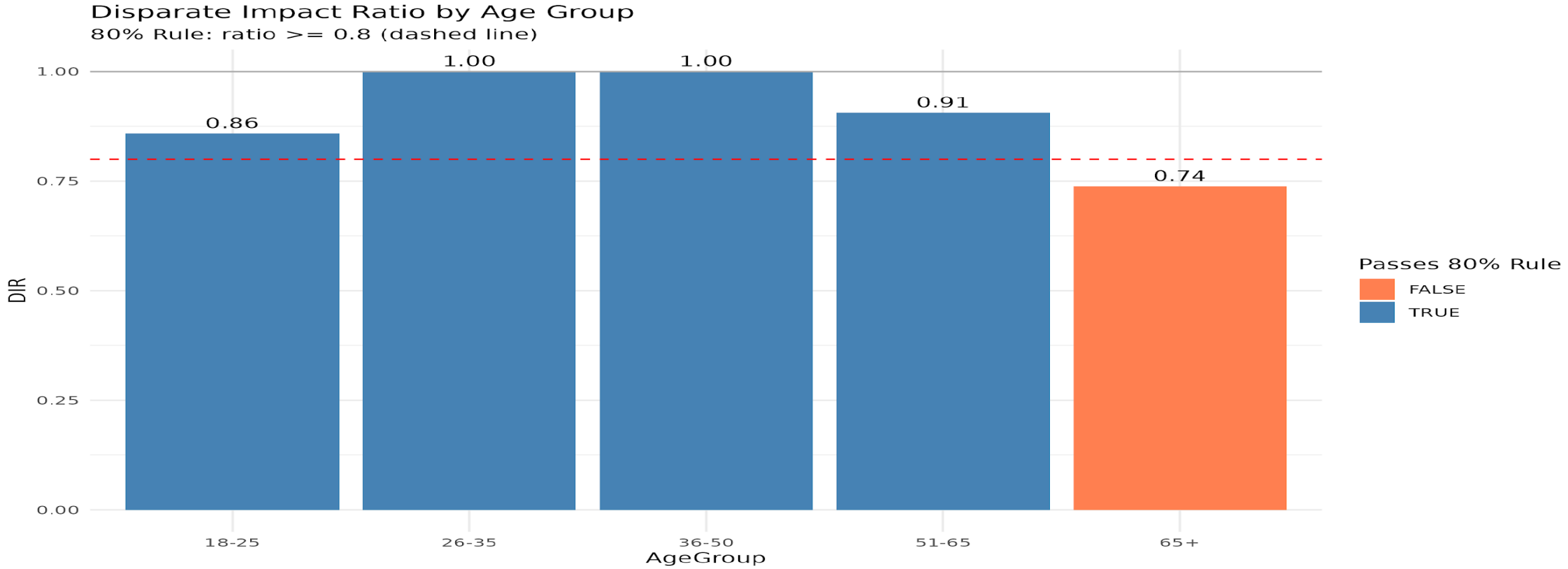


DISPARATE IMPACT – THE 80% RULE

📏 **EEOC 80% Rule** — favorable-outcome rate for protected groups must be ≥ 80% of majority group's rate.

🔴 **Age 65+** → DIR = **0.74** · ❌ FAILS threshold Despite good driving records — calibration failure prevents risk from translating into fair premiums.

🟡 **Age 18-25** → DIR = **0.82** · ⚠️ borderline PASS Requires ongoing monitoring



AUDIT VERDICT

🔴 **Phase 1 — Calibration: FAIL**

GLM fails across all age groups. XGBoost fails for 65+ (calibration slope = 0.17, far from ideal 1.0) — the model systematically misprices the oldest drivers despite their lower risk.

🔴 **Phase 2 — Proxy Detection: FAIL**

BM–age proxy correlation = 0.45. Brute luck contribution exceeds 55% of total SHAP importance — the majority of pricing power derives from circumstances, not choices.

🔴 **Phase 3 — Counterfactual Fairness: FAIL**

Causal age effect ranges 27–48% of predicted frequency. The 65+ group fails the 80% rule (DIR = 0.74) — good drivers are structurally overpriced.

⛔ **Overall verdict: FAIL**

Remediation required before deployment.

RECOMMENDATIONS

FOUR PRIORITIZED REMEDIATION ACTIONS

🔴 **HIGH — Recalibrate the 65+ group**

Slope 0.17 is catastrophic. Apply group-specific Platt scaling or age-specific intercepts to restore pricing accuracy for low-risk elderly drivers.

🟠 **MEDIUM — Document the BM proxy effect**

Required under EU AI Act Art. 13 model risk documentation. Consider BM residualization to strip out the age-correlated component while preserving legitimate risk signal.

🟠 **MEDIUM — Prefer XGBoost over GLM**

Calibration range 0.94–1.02 vs. 0.85–1.12. Tighter, but add fairness constraints — SHAP-based regularization to limit proxy discrimination.

🟡 **LOW — Monitor the 18-25 group**

DIR = 0.82 is borderline. Set automated alerts and schedule quarterly re-audits.

EU AI ACT MAPPING

EACH AUDIT PHASE MAPS TO SPECIFIC REGULATORY ARTICLES

📋 Art. 9 — Risk Management · addressed by Phase 1

Calibration audit identifies systematic prediction errors by protected group — quantified, documented, actionable.

🗃️ Art. 10 — Data Governance · addressed by Phase 2

SHAP proxy detection surfaces biases embedded in training data and transmitted through correlated features like BM.

🔍 Art. 13 — Transparency · addressed by all three phases

SHAP explanations, proxy quantification, and comprehensive audit reports provide full model explainability.

👁️ Art. 14 — Human Oversight · addressed by Phase 3

Counterfactual analysis isolates causal age effects — enabling informed, evidence-based human review.

CONTRIBUTION & POSITIONING

🧩 **Three-phase integrated audit** Calibration + SHAP proxy detection + counterfactual analysis — first framework to combine all three for insurance pricing.

⚖️ **Luck egalitarianism operationalized** First application of Dworkin's option/brute luck distinction to algorithmic insurance pricing.

🇪🇺 **EU AI Act compliance tool** Practical, auditor-ready mapping to Art. 9, 10, 13, 14 — deployment-ready documentation.

📚 **vs. the literature:** → Lindholm (2022) — we add proxy detection + ethics layer → Kusner (2017) — we extend to insurance with calibration tests → Frees & Huang (2023) — broader audit beyond single-model fixes

reach this work in GITHUB!

Access the Github repository for implementation details and collaboration!

🔎

bit.ly/fair-pricing-audit-168

- **10+ years of experience** in insurance and actuarial sector, Full Stack Actuary
- Gruppo Unipol: Life & Non-Life Risk Manager, specialized in Internal Model Premium Risk, Catastrophe Modeling, ORSA, ESG Risk, Cyber Risk
- **Expert in Legal Tech** - developing AI solutions (autonomous agents, RAG systems) for legal applications
- **IAA AI Task Force Leader** - guiding strategic initiatives at the intersection of AI and actuarial science
- Senior Actuary with expertise in AI, Machine Learning, Big Data, Software Developing

CONNECT WITH ME!

ABOUT ME

Manuel Caccone

Italian Society of Actuaries, AI TF

Insurance Data Science Conference
9 - 10 June 2026
House of Insurance, Leibniz Universität Hannover

Thank you very much for your attention

Contact

Manuel Caccone

Senior Actuary | IAA AI Task Force

manuel.caccone@gmail.com

LinkedIn: /in/manuelcaccone

Code & data: github.com/manuelcaccone/fairness-audit