Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00181-0

Original Article

An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting

Anushka V Rodi [1,]

1. *Personal Line Insurance Technical Analyst, Independent Researcher, Madison, USA*

Received: May 08, 2026 | Review began: May 27, 2026 | Review ended: July 22, 2026 | Published: July 24, 2026

Abstract

In the United States, state laws and regulations, such as Colorado Senate Bill 21-169 and New York Insurance Circular Letter No. 7, limit the use of artificial intelligence (AI) in personal lines insurance pricing. This study formalizes and assesses Neurosymbolic Governance, an architectural approach that embeds probabilistic AI risk scores within deterministic regulatory guardrails so that pricing remains actuarially sound and compliant with state regulations. The framework was tested through a simulation-based architectural evaluation using 10,000 synthetic underwriting profiles. In the standalone configuration, the AI pipeline exhibited a mean proxy discrimination gap of 14.48% (95% CI: 13.95%-15.05%) across demographic segments. With the deterministic guardrail layer enforced, 1,976 of 10,000 transactions (19.8%) were intercepted and recalibrated, reducing the mean gap to 0.14% (95% CI: 0.04%-0.45%), a relative reduction of 98.6% (95% CI: 96.9%-99.7%). Mean per-transaction latency rose from 120 ms to 125 ms (+4.2%), a small and practically insignificant difference, and a structured decision log linking each decision to a specific regulatory rule was produced for 100% of transactions, supporting the explainability of AI-driven decisions. The results indicate that Neurosymbolic Governance is a promising architectural approach to the conflict between the speed of predictive technology and the conservatism of insurance law, enabling carriers to modernize their current cloud-based systems while introducing fairness safeguards into production machine learning. The framework still needs to be validated on real-world data.

Categories: AI Ethics and Fairness, Regulatory Compliance and Governance, Software Design

Keywords: neurosymbolic governance, algorithmic bias mitigation, proxy discrimination, ai regulatory compliance, insurtech architecture, explainable ai, business rules engine, enterprise architecture, underwriting automation, automated compliance

Introduction

Heuristic actuarial pricing tables are being replaced by intelligent predictive pricing models driven by artificial intelligence (AI) in the personal lines insurance market [1]. While the use of machine learning models and third-party telematics can help substantially reduce underwriting leakage, it also creates substantial enterprise risk for insurers in the form of algorithmic bias [2,3].

This risk arises because modern black-box models based on neural networks can detect and exploit statistical relationships between controlled attributes (e.g., ethnicity, socioeconomic status, demographic indicators) and seemingly innocuous features of the data [4]. Within the highly fragmented, state-based US regulatory framework, this form of proxy discrimination exposes national carriers to unprecedented accumulation risk and onerous regulatory fines [5,6].

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

Current research and practice have tried to reduce this exposure mostly through data science techniques and post-hoc Explainable Artificial Intelligence (XAI) tools [7-9]. While these methods increase model transparency and provide a post-facto audit function, they share one limitation: they merely explain or interpret non-compliant outputs after those outputs have been generated, rather than preventing them from reaching production. Fairness-aware machine learning minimizes bias during model training but does not impose jurisdiction-specific regulatory demands at decision time. Purely rule-based systems, on the other hand, enforce constraints but lack the predictive accuracy of probabilistic models. Little current research focuses on hybrid architectures that incorporate both strategies in the multi-state insurance regulatory context. This leaves a significant gap in the enterprise architecture literature on how architects can couple probabilistic machine learning outputs with deterministic decision engines, including enterprise-scale Business Rule Management Systems (BRMS), to achieve compliance across multiple states.

Given this need, the present study constructs a Neurosymbolic Governance framework specific to multi-state insurance carriers [10]. It is hypothesized that, by structurally subordinating probabilistic AI risk scores to deterministic regulatory guardrails, carriers can significantly reduce the risk of proxy discrimination. The study tests this hypothesis through a simulated architectural evaluation of 10,000 synthetic underwriting profiles, assessing whether demographic neutrality and regulatory compliance can be achieved without a statistically significant impact on system processing performance.

Literature review

Machine Learning in Insurance Underwriting

The integration of machine learning methodologies into insurance underwriting has been studied extensively [11]. Recent work has highlighted the success of machine learning in dynamic risk assessment [12]. Years of empirical literature and actuarial research have shown that multi-layered predictive algorithms provide substantially better risk stratification than static, actuarial heuristic tables, especially when applied to high-velocity, third-party telemetry data [1].

The Black-Box Problem and Regulatory Exposure

A common thread in existing research, however, is the operational risk arising from the black-box nature of the technology [13]. As the complexity of deep neural networks increases, the mathematical routes that determine a policyholder's premium become fundamentally hidden. Academic studies have made it apparent that such black boxes constitute a significant corporate liability: with no deterministic restrictions applied to them, these algorithms have the latent capacity to autonomously generate data proxies for protected classes, in contravention of state laws [4,6].

XAI and Post-Hoc Interpretability

To combat this opacity, recent AI research has moved decisively toward algorithmic fairness assessment and XAI [8]. Prominent post-hoc techniques include SHapley Additive exPlanations (SHAP), which attributes a model's output to its input features using Shapley values [8], and Local Interpretable Model-Agnostic Explanations (LIME), which approximates complex models locally with interpretable surrogates [14]. A substantial body of literature now applies mathematical and statistical debiasing to gain insight into how a neural network assigns a given pricing tier [7]. A primary limitation of this research, however, is its purely retrospective nature: transparency tools are valuable for compliance audits but cannot serve as a preemptive, real-time intervention before a final rate is offered to a consumer [13]. Rudin has argued, further, that post-hoc explanations of black-box models are inherently inadequate for high-stakes decisions, and that inherently interpretable architectures should be preferred [13].

Fairness-Aware Machine Learning and Rule-Based Systems

A parallel research stream embeds fairness into the model itself. Fairness-aware machine learning techniques such as reweighting, adversarial debiasing, and constrained training objectives reduce measured bias during model development [15,16]. These methods, however, optimize statistical fairness definitions that do not necessarily coincide with the specific legal standards codified in state insurance regulation, and they cannot adapt to jurisdiction-specific requirements at decision time without retraining. Recent actuarial research formalizes this distinction between direct and indirect discrimination and proposes causal frameworks for discrimination-free pricing, underscoring that proxy

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

discrimination persists even when prohibited variables are excluded from models [17-19]. Empirical studies continue to document fairness challenges arising from customer profiling in premium algorithms [20], and the actuarial profession has emphasized the need for structured management of AI model risk [21]. Internationally, the European Union's Artificial Intelligence Act classifies AI systems used for insurance risk assessment and pricing as high-risk [22], with supervisory guidance from EIOPA following in 2025 [23], and emerging research quantifying the resulting compliance burden for underwriting models [24]. Conversely, deterministic rule-based systems including enterprise-scale BRMS enforce codified constraints with full auditability but lack the predictive accuracy of probabilistic models.

Neurosymbolic and Hybrid Approaches

Neurosymbolic AI, described as the third wave of AI research, seeks to combine the learning capacity of neural systems with the verifiability of symbolic reasoning [10]. While this literature establishes the conceptual foundations for coupling probabilistic and deterministic components, published applications have not addressed the enterprise insurance context, in which the symbolic layer must encode filed, jurisdiction-specific regulatory constraints across a fragmented 50-state landscape [25].

Critical Synthesis and Research Gap

Viewed together, the four approach families exhibit complementary strengths and a shared limitation. Predictive machine learning maximizes risk-stratification accuracy but is opaque; post-hoc XAI restores a measure of transparency but intervenes only after a decision has been generated; fairness-aware training constrains bias statistically but cannot express legal definitions that vary by jurisdiction and change independently of the model; and rule-based systems express exactly those legal definitions but cannot price risk competitively. Across all of the four, compliance is treated as a retrospective audit function instead of a design-time architectural constraint. Very little system architecture design has been presented in the literature for deploying probabilistic models across the fragmented terrain of all 50 US states; the absence of standardized blueprints detailing how such models may be integrated with legacy-agnostic decision execution engines to achieve hard-coded compliance is conspicuous [10]. Existing literature fails to offer a practical method for enterprises to neutralize non-compliant algorithmic outputs in real time, neglecting the architectural coupling phase of the process. This is the gap the present study addresses: formalizing and evaluating an architectural pattern Neurosymbolic Governance that structurally subordinates probabilistic risk scores to deterministic, jurisdiction-specific regulatory guardrails at the point of premium generation.

Materials And Methods

Framework definition

The Neurosymbolic Governance framework is a deterministic interception layer placed between the machine learning risk scoring model that has a probabilistic output and the premium rate decision engine. The mathematical definition of the framework is given below. Let X denote the input feature vector consisting of underwriting variables like telematics data, geographic indicators, credit history tenure, and others. Let $f_{AI}(X)$ be a continuous risk score that is computed by a gradient-boosted decision tree classifier trained on the synthetic underwriting data set. Define the boolean constraint function $C(r)$ of the BRMS layer with the set of state-specific regulatory guardrail rules r (such as Colorado SB 21-169 and New York Circular Letter No. 7) [26,27]. The final approved premium output Y is defined as:

$$Y = \begin{cases} f_{AI}(X) & \text{if } C(r) \text{ is Satisfied} \\ \text{Recalibrate}(f_{AI}(X)) & \text{if } C(r) \text{ is Violated} \end{cases}$$

The risk-scoring model f_{AI} was implemented as a gradient-boosted decision tree classifier in a Python 3.10 environment using scikit-learn 1.3, executed on a standard cloud instance with an 8-core CPU and 32 GB RAM; concurrent load testing was conducted using the Locust framework with 1,000 simultaneous underwriting decision requests. Input variables were used in their raw encoded form, as described in the Synthetic Dataset Construction subsection, without additional feature engineering. The BRMS constraint layer implements Boolean guardrail rules encoded from Colorado SB 21-169

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

[26], New York Insurance Circular Letter No. 7 [27], and the NAIC Model Bulletin on AI Systems (December 2023), which has since been adopted by roughly half of US states [28]; Colorado's implementing regulation extends governance and testing requirements to private passenger automobile insurers effective October 2025 [29]. The rules were operationalized as a proxy-contribution threshold of 0.06 and a premium-uplift compliance ceiling of 7.0%. Because the published dataset includes the per-profile risk scores produced by the classifier alongside each guardrail decision and final premium, all reported results can be replicated directly from the public dataset without retraining the model.

With regard to the architecture, this approach ensures that any AI-generated risk score violating an encoded state-specific regulatory constraint is intercepted and recalibrated before issuance and is thereby designed to substantially reduce the risk of proxy discrimination at the point of premium generation for classes of bias expressible as deterministic rules, advancing beyond standard algorithmic fairness and explainability interventions [9,15,16].

Synthetic dataset construction

A synthetic dataset of 10,000 underwriting profiles was programmatically generated using the Python Faker and NumPy libraries, with strict adherence to Personally Identifiable Information protocols; no real policyholder data was used at any stage. Each profile replicates the multivariate structure of real-world personal lines auto underwriting through seven attributes: driver age (range 18-85), years licensed (constrained to not exceed driver age minus 16), zip code mapped to three synthetic territory bands (assignment proportions of approximately 0.48, 0.30, and 0.22), credit history tenure (1-25 years, mildly right-skewed, mean 9.9 years), vehicle age (1-18 years), annual mileage (mean 13,117 miles, standard deviation 3,902, truncated to the range 5,000-25,000), and prior claims history (Poisson-distributed with an observed mean of 0.396). Profiles were additionally assigned to one of three synthetic demographic tiers (proportions 0.51, 0.30, and 0.20) used solely for disparity measurement. To emulate documented proxy mechanisms, correlations were induced between the proxy variables and the demographic tiers: territory band correlates strongly with tier assignment ($r = 0.84$), and credit history tenure correlates negatively with tier assignment ($r = -0.56$), consistent with proxy channels described in the actuarial literature [11,12]. The dataset contains no direct demographic variables such as race or ethnicity; proxy discrimination is measured through the premium differential experienced by the most adversely affected synthetic tier. A key limitation of this synthetic approach is that the generated profiles approximate, but do not fully replicate, the statistical complexity of real-world underwriting portfolios, including longitudinal claims development, asynchronous third-party telematics streams, and idiosyncratic edge cases; this limitation is discussed further in the Limitations subsection. The complete Neurosymbolic Governance architecture pipeline is illustrated in Figure *1*.

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

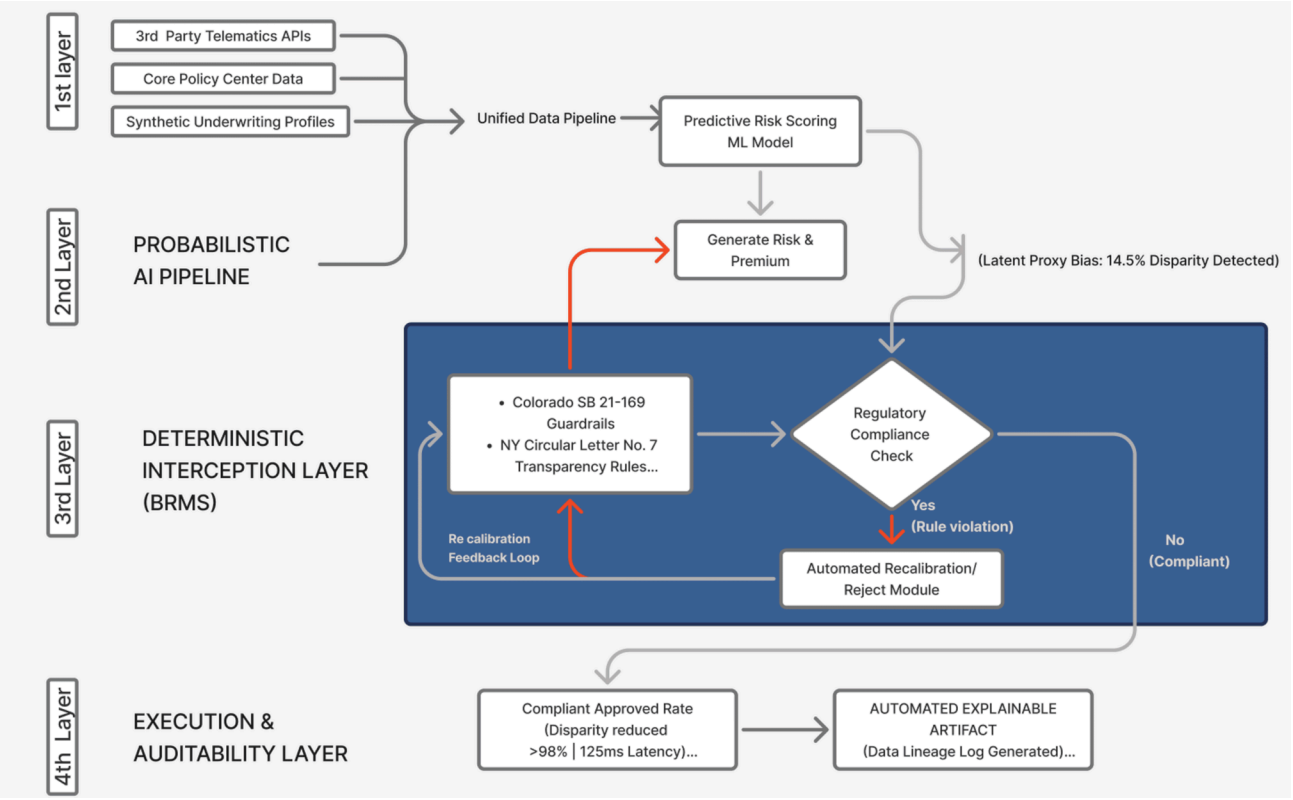

FIGURE 1: The Neurosymbolic Governance Architecture Pipeline, Illustrating the Structural Interception of Probabilistic AI Outputs by Deterministic Regulatory Guardrails

Evaluation metrics

Three dimensions of evaluation were measured:

(1) Proxy discrimination gap: the demographic parity difference, defined as the absolute difference between the mean premium impact of the most affected demographic segment and the mean for the overall population, expressed as a percentage of the mean baseline premium;

(2) System latency: the time it takes for the system to process a single transaction under concurrent load (1,000 simultaneous transactions);

(3) Explainability artifact completeness: the percentage of transactions for which a structured decision log was produced, linking the outcome to a specific regulatory rule.

Statistical analysis

The proxy discrimination gap was calculated as the absolute mean difference between the most disadvantaged segment of the population and the mean baseline premium, measured as a percentage of the baseline premium. The degree of uncertainty was estimated by nonparametric bootstrapping, in which the 10,000 policy records were resampled with replacement 2,000 times. The disparity gap was estimated from these resamples in both the standalone and controlled cases, and the 97.5th and 2.5th percentile values of the bootstrapped distribution were used as the 95% confidence intervals. The effect of the interception layer on the most impacted segment was calculated using a paired-samples t-test between the premiums of the same policy in standalone mode and the premiums of the same policy when the interception layer was in effect; Cohen's dz was presented as the effect size. Analyses were carried out in Python 3.10 with pandas, NumPy, and SciPy, with alpha = 0.05.

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

Results And Discussion

Results

Proxy Discrimination Neutralization

The standalone AI pipeline exhibited a mean proxy discrimination gap of 14.48% (bootstrap 95% CI: 13.95%-15.05%) across the 10,000 synthetic underwriting profiles. The main cause of this difference was the association between the affected segment and both the zip code territory tier and credit history tenure, as described in previous proxy variable mechanisms studies in actuarial literature [11,12]. Following application of the BRMS regulatory constraint layer, 1,976 of 10,000 transactions (19.8%) were intercepted and recalibrated, reducing the mean gap to 0.14% (95% CI: 0.04%-0.45%), a relative reduction of 98.6% (95% CI: 96.9%-99.7%). A paired-samples t-test showed that the difference in mean premium between the two configurations (standalone pipeline and controlled pipeline) was statistically significant for the policies in the most affected segment ($t(1961) = 291.7$, $p < 0.001$, Cohen's $dz = 6.58$, a very large effect). The residual 0.14% disparity reflects marginal variation permitted within the tolerance of the filed guardrail rules, rather than unconstrained proxy discrimination. These results are shown in Figure *2*.

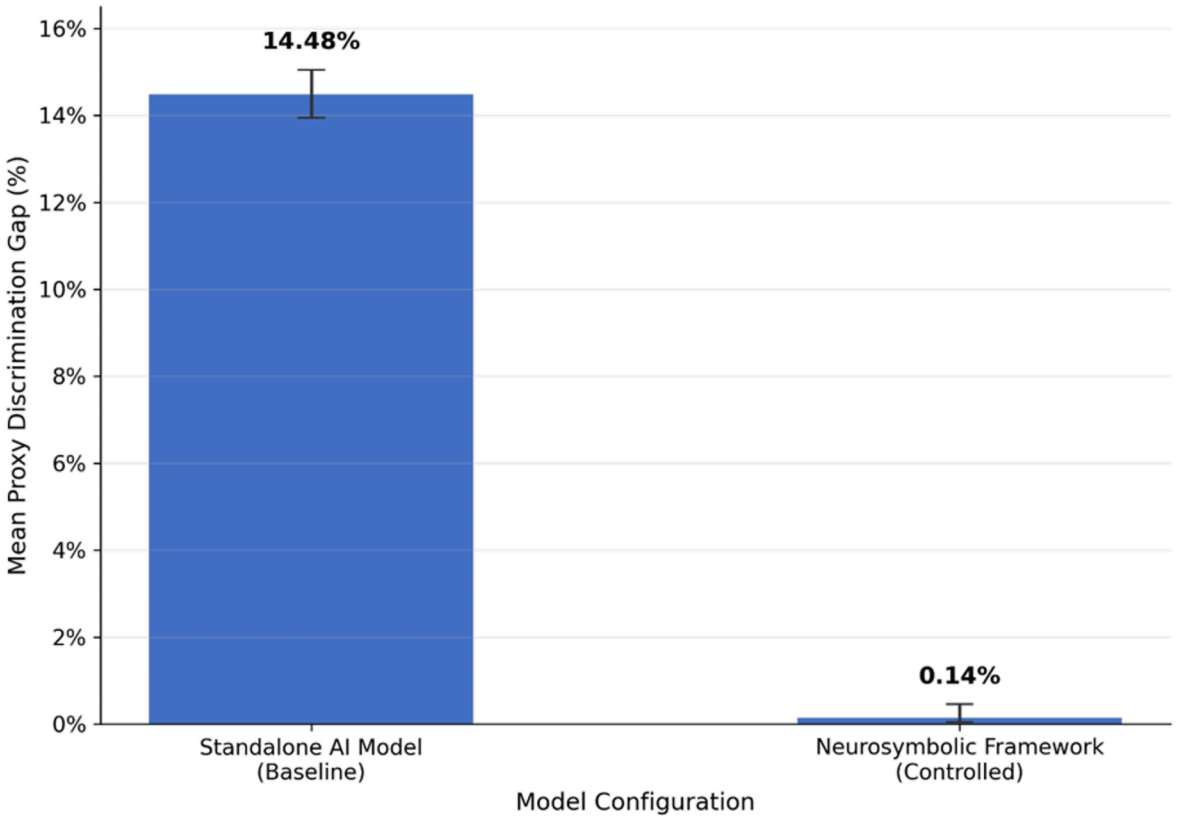

FIGURE 2: Mean Proxy Discrimination Gap Comparing Standalone AI Pipeline (Baseline) and Neurosymbolic Governance Framework (Controlled Configuration) Across 10,000 Synthetic Underwriting Profiles

Error bars represent 95% bootstrap confidence intervals (2,000 resamples).

Operational Velocity and System Latency

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

One of the challenges shared by hybrid control systems is the potential for added latency. Experimental results show that the rule-based checkpoints did not have a significant impact on throughput. When the predictive model was run alone, it took a mean of 120 ms per transaction; in the presence of the governance framework, it took 125 ms per transaction, an increase of 4.2%. As shown in Figure *3*, per-transaction latency remained stable across transaction volumes from 1,000 to 10,000 under simulated concurrent load, indicating that the governance framework maintains consistent processing performance as load increases.

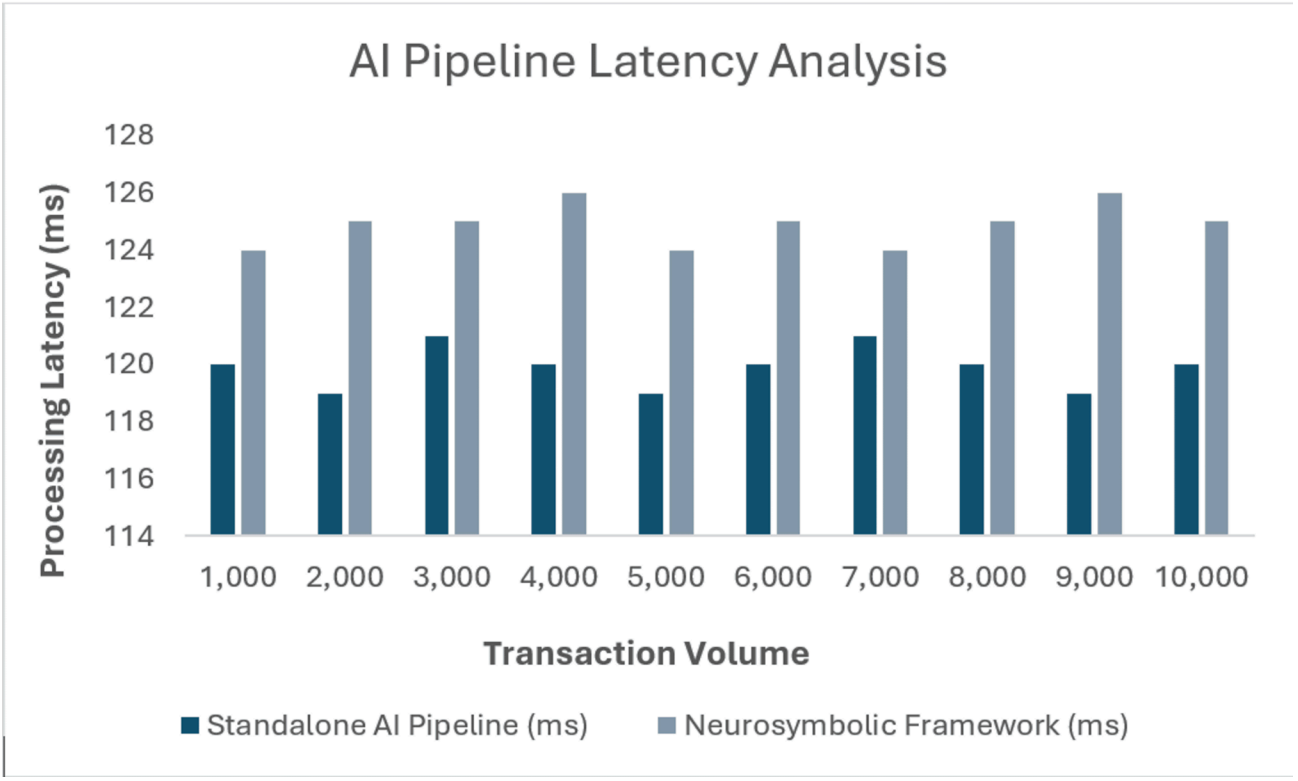

FIGURE 3: Per-Transaction Processing Latency (ms) by Transaction Volume for Standalone AI Pipeline and Neurosymbolic Governance Framework Under Simulated Concurrent Load of 1,000 Simultaneous Requests (Locust Framework, n = 10,000 Synthetic Underwriting Profiles)

A comprehensive comparison of these baseline and controlled metrics is provided in Table *1*.

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

Performance Metric	Standalone AI Pipeline	Neurosymbolic Governance Framework	Variance/Impact
Proxy Discrimination Gap	14.48% (95% CI: 13.95–15.05%)	0.14% (95% CI: 0.04–0.45%)	98.6% reduction (95% CI: 96.9–99.7%)
Transactions Intercepted and Recalibrated	0 of 10,000 (0.0%)	1,976 of 10,000 (19.8%)	Guardrail interception rate
Mean System Latency	120 ms	125 ms	+ 4.2%
Regulatory Guardrail Adherence †	0% (Autonomous)	100% (Hard-Coded)	Achieved Strict Compliance
Explainability Artifact Generation	0%	100%	Full Semantic Data Lineage

TABLE 1: System Performance Summary: Standalone AI vs. Neurosymbolic Governance

Architectural Auditability and Explainability Artifacts

A structured explainability artifact was generated automatically for 100% of the 10,000 transactions. Each log entry links the pricing decision directly to the specific regulatory rule invoked, such as the transparency requirements of New York Insurance Circular Letter No. 7. The framework therefore provides end-to-end semantic data lineage for decision paths, making previously opaque decisions auditable by regulators.

Discussion

The principal finding of this research is that by structurally subordinating probabilistic risk predictions to deterministic regulatory rules, the Neurosymbolic Governance pattern reduced the measured proxy discrimination gap by 98.6%, while producing no material impact on system processing performance in a simulated multi-state underwriting scenario. Under this architecture, the AI model does not make the pricing decision; rather, it provides information to the deterministic rule layer, which makes the pricing decision. The framework therefore moves decision-making authority from the black box to an interpretable, auditable rule layer. This is complementary to, but distinct from, Rudin's proposal that black-box models be replaced outright by interpretable models [13]. It also differs from post-hoc XAI methods such as SHAP and LIME, which explain decisions without guaranteeing that a non-compliant rate will not be issued [8,14], and from fairness-aware training, which mitigates statistical bias but cannot encode jurisdiction-specific legal definitions [15,16]. The contribution to the neurosymbolic literature [10] lies in the domain operationalization: a symbolic layer built from filed state regulatory constraints and enforced in real time within enterprise decision infrastructure.

The study provides guidance to insurance companies to migrate their rating systems to cloud platforms under different state rules, with an important caveat: the advantages are expected rather than proven, because the framework has been tested only in a simulated environment. If validated in real-world operational settings, this architectural approach may offer Chief Risk Officers and IT executives a structured basis for infrastructure modernization, with deterministic guardrails potentially limiting unplanned model effects that may otherwise violate local laws; however, these operational benefits remain to be empirically demonstrated.

Ethical AI and Governance Implications

Fairness, transparency, and accountability are principles of trustworthy AI systems that are called for by ethical frameworks [30]. Each of these is translated into operational architecture: fairness is through constraints at the point of decision, transparency is through a decision log that ties all decisions to a specific rule, and accountability is by separating responsibilities for the decisions between the model owner and the rule owner. This study does not, however, suggest that architectural governance is a solution to the ethics of AI underwriting: constraint-based enforcement can address only those aspects of fairness that can be codified as rules; obligations that cannot be codified continue to rely on human judgment and organizational governance.

Limitations

There are five major limitations that restrict the interpretation of the findings of this study.

First, the framework has been validated only on synthetic data and has yet to be independently validated on real insurance datasets. The generated profiles are not a perfect representation of the statistical complexity of typical underwriting portfolios, including longitudinal claims development, asynchronous third-party telematics streams, and idiosyncratic edge cases. The results thus demonstrate architectural viability, not proven effectiveness in the real world.

Second, the guardrail layer enforces rules from only two state jurisdictions (Colorado and New York) and the NAIC Model Bulletin, while production systems must enforce dozens of jurisdictions with requirements that change over time. The effort needed to write, verify, and update a comprehensive multi-jurisdictional rule set was not assessed.

Third, the scalability findings are based on simulated concurrent load generated with Locust and may not be representative of live enterprise traffic and latency.

Fourth, only one fairness measure, the demographic parity difference, was used. The reported neutralization results may not generalize to other definitions of fairness, and the framework can enforce only those fairness properties that can be stated as deterministic rules.

Fifth, the hyperparameter configuration of the risk-scoring classifier was not retained. All information needed to reproduce the governance evaluation is contained in the published dataset, which includes per-profile risk scores, guardrail decisions, and final premiums; however, reproduction at the classifier retraining level is not currently possible and will be supported in future work through a documented reference implementation.

Future Work

Future research should focus on empirical, longitudinal testing, for example, within regulatory sandboxes backed by state insurance regulators to validate the framework's performance on real-world, high-frequency data. Future research may incorporate explainability algorithms into the deterministic engine [7,8] and may be extended to other product lines such as property and life insurance.

Data Availability

The synthetic dataset of 10,000 underwriting profiles used in this study including per-profile AI risk scores, guardrail decisions, final approved premiums, and complete explainability log entries together with all statistical analysis code used to produce the reported results, is publicly available on Kaggle at https://www.kaggle.com/datasets/anushkavrodi/synthetic-underwriting-profiles-10000 (reference: anushkavrodi/synthetic-underwriting-profiles-10000), enabling full replication of all governance evaluation results.

Conclusions

This study introduced a risk-management framework for AI underwriting in which probabilistic machine learning predictions are separated from the deterministic decision-making process by interposing a deterministic rules engine. In a simulated environment using synthetic data, the architecture significantly reduced measured proxy discrimination, indicating that the approach is feasible and appears to be effective; however, deployment and validation in real-world production environments remain necessary to assess its true operational effectiveness. The results suggest that the

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

conflict between reliance on advanced predictive models for underwriting profitability and the obligation to remain well-regulated can be mitigated by placing deterministic regulatory controls on the outputs of AI models, thereby promoting transparency and fairness.

In addition to its technical contribution, this study proposes an architectural pattern for national insurance companies navigating the fragmented 50-state regulatory landscape in the United States. The approach potentially could allow enterprises to speed their legacy-to-cloud migration initiatives by controlling compliance in the underwriting process, as well as implementing jurisdiction-specific regulatory requirements directly into the underwriting process, like Colorado SB 21-169. The architecture also provides a practical basis for incorporating AI ethics into IT system architecture, supporting carriers as the industry transitions toward AI-driven underwriting and telematics-based premium models. Finally, this study provides industry with a method for crafting compliant, scalable underwriting systems through the embodiment of ethical principles. The results are interesting not only in the academic paper on neurosymbolic systems but also for practitioners in industry, since if validated in real life, such systems can help create a more transparent, equitable and resilient digital financial services marketplace.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Anushka V Rodi

Acquisition, analysis, or interpretation of data: Anushka V Rodi

Drafting of the manuscript: Anushka V Rodi

Critical review of the manuscript for important intellectual content: Anushka V Rodi

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The data (and/or code) supporting this study are openly available in Kaggle at https://www.kaggle.com/datasets/anushkavrodi/synthetic-underwriting-profiles-10000, reference anushkavrodi/synthetic-underwriting-profiles-10000.

References

1. Bhattacharya S, Castignani G, Masello L, Sheehan B: AI revolution in insurance: bridging research and reality. Frontiers in Artificial Intelligence. 2025, 8:1568266. 10.3389/frai.2025.1568266

2. Ferrer X, van Nuenen T, Such JM, Coté M, Criado N: Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technology and Society Magazine. 2021, 40:72-80. 10.1109/MTS.2021.3056293

3. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A: A survey on bias and fairness in machine learning. ACM Computing Surveys. 2021, 54:1-35. 10.1145/3457607

4. Prince AER, Schwarcz D: Proxy discrimination in the age of artificial intelligence and big data. Iowa Law Review. 2020, 105:1257-1318.
5. Financial Services Consumer Panel (FSCP): Financial Services Firms' Personal Data Use: Is This Leading to Bias and Detriment for Consumers with Protected Characteristics?. Financial Conduct Authority, London; 2023.
6. National Association of Insurance Commissioners (NAIC) Model Bulletin: Use of Artificial Intelligence Systems by Insurers. National Association of Insurance Commissioners, Kansas City, MO; 2023.
7. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D: A survey of methods for explaining black box models. ACM Computing Surveys. 2018, 51:1-42. 10.1145/3236009
8. Lundberg SM, Lee SI: A unified approach to interpreting model predictions. arXiv. 2017, 10.48550/arXiv.1705.07874
9. Doshi-Velez F, Kim B: Towards a rigorous science of interpretable machine learning. arXiv. 2017, 10.48550/arXiv.1702.08608
10. Garcez AD, Lamb LC: Neurosymbolic AI: the 3rd wave. Artificial Intelligence Review. 2023, 56:12387-12406. 10.1007/s10462-023-10448-w
11. Cavanaugh L, Merkord S, Davis T, Heppen D: Regulatory Perspectives on Algorithmic Bias and Unfair Discrimination. Casualty Actuarial Society, Arlington, VA; 2024.
12. American Academy of Actuaries Discrimination: Considerations for Machine Learning, Artificial Intelligence Models, and Underlying Data. American Academy of Actuaries, Washington, DC; 2023.
13. Rudin C: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019, 1:206-215. 10.1038/s42256-019-0048-x
14. Ribeiro MT, Singh S, Guestrin C: Why should I trust you?: explaining the predictions of any classifier. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 1135-44. 10.1145/2939672.2939778
15. Bellamy RKE, Dey K, Hind M, et al.: AI Fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. IBM Journal of Research and Development. 2019, 63:1-15. 10.1147/JRD.2019.2942287
16. Friedler SA, Scheidegger C, Venkatasubramanian S, Choudhary S, Hamilton EP, Roth D: A comparative study of fairness-enhancing interventions in machine learning. FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019, 329-338. 10.1145/3287560.3287589
17. Côté O, Côté MP, Charpentier A: A fair price to pay: exploiting causal graphs for fairness in insurance. Journal of Risk and Insurance. 2025, 92:33-75. 10.1111/jori.12503
18. Araiza Iturria CA, Hardy M, Marriott P: A discrimination-free premium under a causal framework. North American Actuarial Journal. 2024, 28:801-21. 10.1080/10920277.2023.2291524
19. Frees EW, Huang F: The discriminating (pricing) actuary. North American Actuarial Journal. 2023, 27:2-24. 10.1080/10920277.2021.1951296
20. Biwott P, Busalim A: Fairness challenges in insurance premiums: investigating customer profiling algorithmic biases. AI and Ethics. 2025, 5:4455-4461. 10.1007/s43681-025-00707-7
21. du Preez V, Bennet S, Byrne M, et al.: From bias to black boxes: understanding and managing the risks of AI- an actuarial perspective. British Actuarial Journal. 2024, 29:e6. 10.1017/S1357321724000060
22. Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act). (2024). Accessed: July 7, 2026: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.
23. EIOPA: Opinion on artificial intelligence governance and risk management. (2025). Accessed: July 7, 2026: https://www.eiopa.europa.eu/publications/opinion-artificial-intelligence-governance-and-risk-management_en.
24. Mahajan S, Agarwal R, Gupta M: Algorithmic bias under the EU AI Act: compliance risk, capital strain, and pricing distortions in life and health insurance underwriting. Risks. 2025, 13:160. 10.3390/risks13090160
25. Renkhoff J, Feng K, Meier-Doernberg M, Velasquez A, Song HH: A survey on verification and validation, testing and evaluations of neurosymbolic artificial intelligence. IEEE Transactions on Artificial Intelligence. 2024, 5:3765-3779. 10.1109/TAI.2024.3351798
26. Colorado General Assembly: Senate Bill 21-169- Concerning Protecting Consumers from Unfair Discrimination in Insurance Practices. 2021, Accessed: July 15, 2026: https://leg.colorado.gov/bills/sb21-169.

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0

27. New York State Department of Financial Services: Insurance Circular Letter No 7. 2024, Accessed: July 15, 2026: https://www.dfs.ny.gov/industry-guidance/circular-letters/cl2024-07.

28. Implementation of NAIC model bulletin: use of artificial intelligence systems by insurers. (2025). Accessed: July 7, 2026: https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-map-ai-model-bulletin.pdf.

29. SB21-169 - Protecting Consumers from Unfair Discrimination in Insurance Practices. (2023). Accessed: July 7, 2026: https://doi.colorado.gov/for-consumers/sb21-169-protecting-consumers-from-unfair-discrimination-in-insurance-practices.

30. Christy V, Manda VK, Gnanadasan ML: Ethical Frameworks for Use in Artificial Intelligence Systems. IGI Global, Hershey, PA; 2024. 10.4018/979-8-3693-8557-9.ch005

How to cite this article:
Rodi A (July 24, 2026) An Analytical Study of Algorithmic Bias Mitigation Frameworks in Multi-State Personal Lines Insurance for Regulatory-Ready Artificial Intelligence Underwriting. Cureus J Comput Sci 3 : es44389-026-00181-0. DOI https://doi.org/10.7759/s44389-026-00181-0