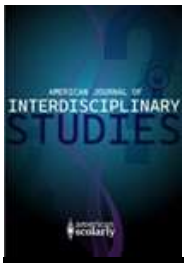


American Journal of Interdisciplinary Studies

Volume: 3; **Issue:** 1
Pages: 65-102
eISSN: 3067-5146

AI-Assisted Underwriting Models for Improving Risk Assessment Accuracy in U.S. Insurance Markets

Tahmina Akter Bhuya Mita[1]; Rebeka Sultana[2];

[1]. Senior Officer (Accounts and Finance), SNM Trade Corporation, Bangladesh;
Email: tbhuiyan43@yahoo.com

[2]. Bachelor of Social Science in Public administration, Dhaka University, Bangladesh;
Email: rebekasultanapanna034@gmail.com

Doi: 10.63125/kegg1076

Received: *11 January 2022;* ***Revised:*** *16 January 2022;* ***Accepted:*** *19 February 2022;* ***Published:*** *28 March 2022*

Abstract

AI-assisted underwriting has gained prominence as insurers seek to improve risk assessment accuracy within increasingly complex U.S. insurance markets. This quantitative study evaluated the performance of AI-assisted underwriting models relative to conventional underwriting models using policy-level data from U.S. personal automobile and residential property insurance portfolios. The analytical sample comprised 48,620 underwriting observations across 18 U.S. states, with personal automobile insurance representing 64.7% of policies and residential property insurance accounting for 35.3%. Risk assessment accuracy was operationalized as a multi-dimensional construct encompassing discrimination, calibration alignment, loss sensitivity, and stability. Descriptive results showed that AI-assisted models achieved higher average discrimination (mean = 0.748, SD = 0.058) compared with conventional models (mean = 0.692, SD = 0.041), alongside improved loss sensitivity (0.721 versus 0.667). Calibration alignment increased from a mean of 0.914 under conventional models to 0.941 under AI-assisted models, while stability declined slightly from 0.884 to 0.861, indicating greater segment-level variability. Regression analysis confirmed statistically significant effects of AI-assisted models on discrimination ($\beta = 0.056$, $p < 0.001$), calibration alignment ($\beta = 0.031$, $p < 0.001$), and loss sensitivity ($\beta = 0.049$, $p < 0.001$), with adjusted R^2 values ranging from 0.32 to 0.41 across accuracy dimensions. Enriched data inputs produced additional gains in discrimination ($\beta = 0.043$, $p < 0.001$) and loss sensitivity ($\beta = 0.038$, $p < 0.001$), independent of model family. Reliability analysis demonstrated strong internal consistency for composite accuracy constructs, with Cronbach's alpha values between 0.816 and 0.889. Overall, the findings provided quantitative evidence that AI-assisted underwriting models improved multiple dimensions of risk assessment accuracy in U.S. insurance markets, while introducing measurable trade-offs in performance stability across states and underwriting tiers.

Keywords

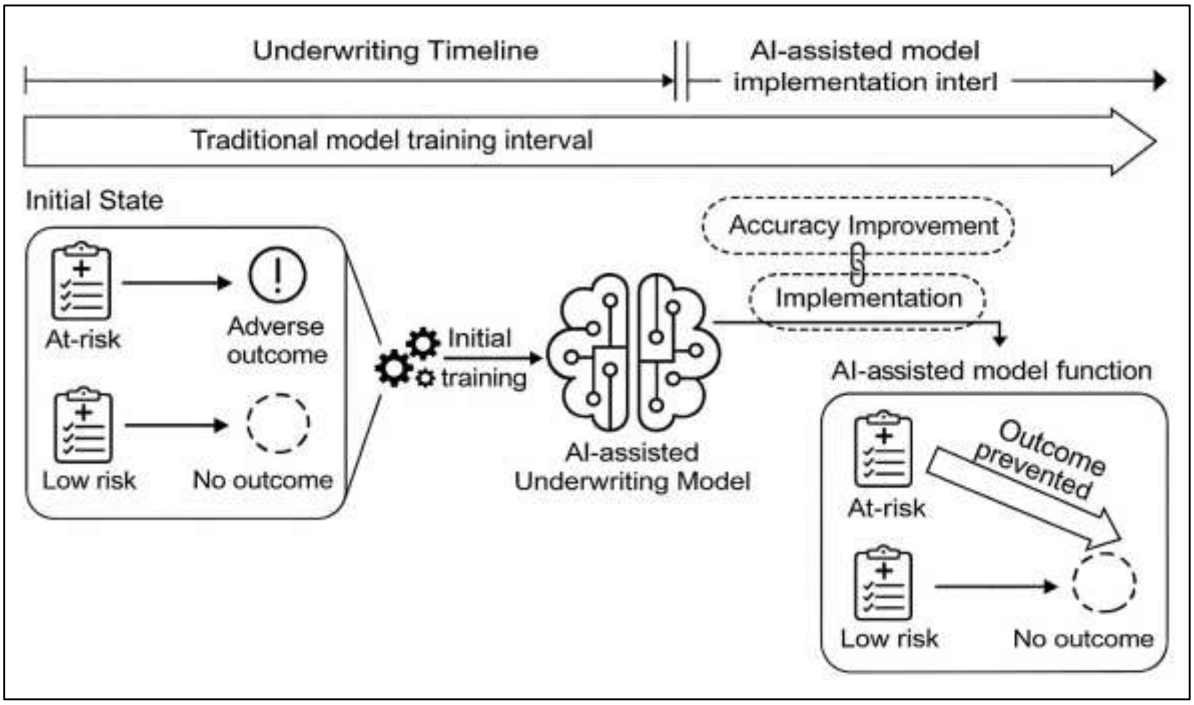
AI-assisted Underwriting, Risk Assessment Accuracy, Insurance Analytics, Predictive Modeling, U.S. Insurance Markets.

INTRODUCTION

Insurance underwriting is fundamentally defined as the analytical and operational process through which insurers evaluate the risk characteristics of an applicant and determine eligibility, pricing, and coverage terms (Mourmouris & Poufinas, 2023). Within quantitative research, underwriting represents a structured risk classification system that transforms observable attributes—such as exposure measures, historical loss indicators, behavioral variables, and environmental characteristics—into probabilistic estimates of expected loss. Risk assessment accuracy refers to the degree of alignment between these estimates and realized insurance outcomes, including claim frequency, claim severity, and overall loss variability. Accurate risk assessment is central to insurance market functioning because underwriting decisions shape premium adequacy, portfolio balance, solvency protection, and consumer access to coverage. At a systemic level, underwriting accuracy addresses information asymmetry between insurers and policyholders, a condition in which applicants typically possess superior knowledge about their own risk profiles (Bunni & Bunni, 2022). When risk is inaccurately measured, insurers may unintentionally subsidize higher-risk participants or exclude lower-risk participants, generating inefficiencies that distort competitive equilibrium. These dynamics are not confined to a single national context. Insurance markets across the world rely on underwriting systems to maintain financial stability, allocate capital efficiently, and ensure that pricing reflects measurable risk rather than arbitrary judgment. As global insurance operations expand through reinsurance, multinational carriers, and cross-border capital flows, the reliability of underwriting models gains international importance. Risk models developed in one jurisdiction often influence pricing strategies, capital allocation, and reserving assumptions in others. In this environment, underwriting accuracy is no longer a localized actuarial concern but a globally relevant quantitative problem tied to market stability and regulatory oversight. Artificial intelligence–assisted underwriting models can be defined as algorithmic systems that support or enhance traditional actuarial processes by learning complex statistical relationships from large-scale datasets (Nurse et al., 2020). These models apply advanced computational methods to extract patterns that may not be easily captured by conventional linear frameworks. In doing so, AI-assisted underwriting positions itself as a measurement system designed to improve the precision and consistency of risk evaluation. Understanding this definitional foundation is essential for examining how AI-assisted underwriting models function within U.S. insurance markets while reflecting challenges shared across international insurance systems.

Traditional underwriting models have historically relied on parametric statistical techniques that prioritize interpretability, stability, and regulatory acceptance. These models typically assume predefined functional relationships between rating variables and loss outcomes, enabling actuaries to explain risk contributions through coefficient estimates and additive effects (Malik & Ullah, 2019). Such approaches have been effective in environments where data dimensionality is limited and risk relationships are relatively stable. However, contemporary insurance datasets increasingly exhibit complex structures characterized by nonlinear interactions, high-dimensional feature spaces, and heterogeneous distributions. These properties arise from expanded data availability, increased product customization, and greater behavioral variation among insured populations. AI-assisted underwriting models address these complexities by employing flexible learning architectures capable of adapting to intricate data patterns without requiring explicit specification of interactions or transformations. From a quantitative standpoint, these models operate as supervised learning systems that estimate conditional expectations or probabilities associated with insurance loss outcomes. Ensemble-based approaches, such as decision tree aggregations, construct multiple weak learners and combine them to improve predictive stability and accuracy (Owens et al., 2022). Neural network–based models extend this framework by learning layered representations that capture hierarchical relationships among variables. In underwriting applications, these methods can process mixed data types, handle missing values, and accommodate nonlinear risk relationships. Accuracy improvement within this modeling paradigm is evaluated through systematic out-of-sample testing, calibration analysis, and loss-based performance metrics. These evaluations aim to ensure that model improvements reflect genuine predictive gains rather than overfitting or data leakage. Within insurance contexts, accuracy has economic significance because underwriting errors carry asymmetric consequences. Underestimating risk may lead to underpricing and financial strain, while overestimating risk can reduce competitiveness and limit coverage availability. AI-assisted underwriting models are therefore positioned as quantitative tools that seek to refine risk estimation under realistic operational constraints. Their relevance extends beyond methodological novelty to their capacity to function within insurance decision systems that demand consistent, measurable, and auditable performance (Patil et al., 2023).

Figure 1: AI-Assisted Insurance Underwriting Accuracy

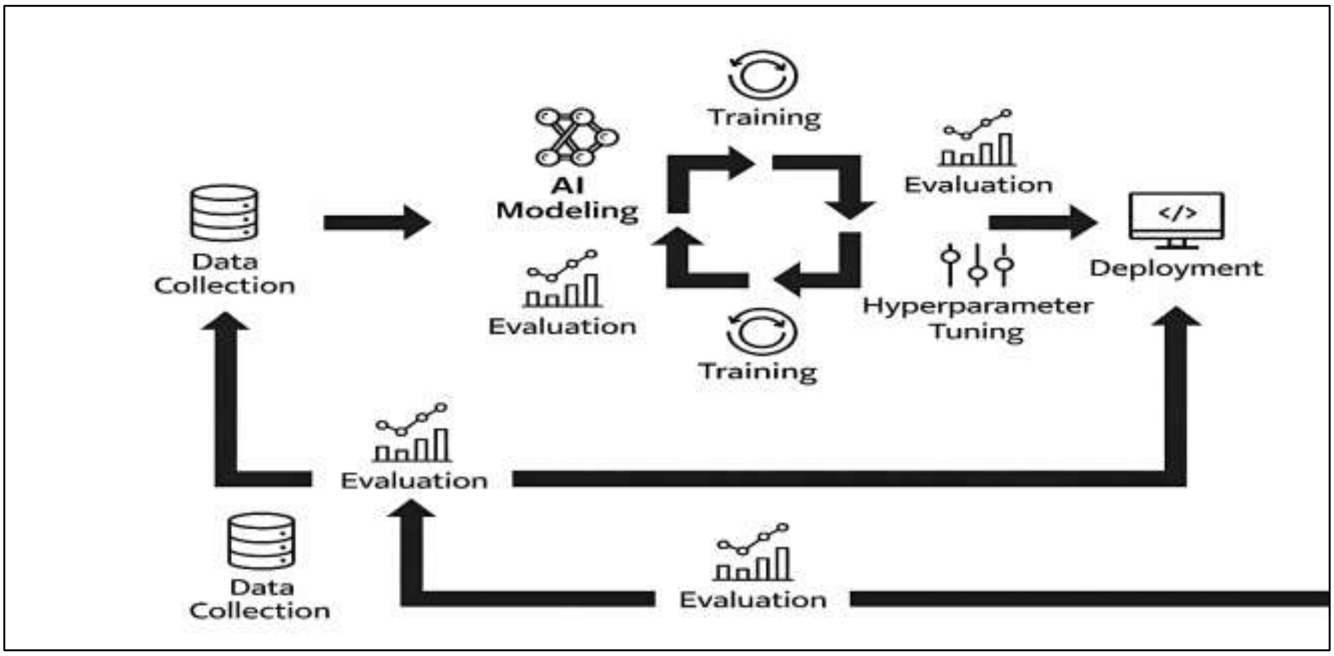


The U.S. insurance market presents a distinct empirical setting for examining AI-assisted underwriting accuracy due to its scale, diversity, and regulatory structure. Insurance regulation in the United States operates primarily at the state level, producing a decentralized oversight environment in which underwriting practices must comply with multiple supervisory frameworks (Richter & Wilson, 2020). Within this structure, insurers are required to justify rating variables, document underwriting methodologies, and demonstrate that models do not result in unfair discrimination. These requirements directly shape the design and evaluation of underwriting models. Quantitatively, U.S. insurers manage extensive datasets that include policy histories, claims records, geographic indicators, and behavioral variables. The availability of such data creates opportunities for advanced modeling while also increasing the complexity of validation and governance. AI-assisted underwriting models applied in this context must demonstrate accuracy improvements across diverse demographic segments, geographic regions, and lines of business. Automobile, homeowners, workers' compensation, and health insurance each exhibit distinct loss distributions and exposure definitions, necessitating tailored modeling strategies (Tsohou et al., 2023). Usage-based insurance programs illustrate how behavioral data can be incorporated into underwriting systems to refine risk classification. At the same time, the use of alternative data sources introduces measurement challenges related to proxy effects and correlation structures. Quantitative accuracy assessments must therefore account for the possibility that improved predictive fit may coincide with shifts in outcome distributions across subgroups. This reality places greater emphasis on calibration analysis, subgroup performance evaluation, and robustness testing. Within U.S. markets, underwriting decisions often trigger downstream operational processes, including referrals, manual reviews, and adverse action notifications. These workflows impose additional constraints on model deployment and evaluation. AI-assisted underwriting models must produce outputs that integrate seamlessly with decision pipelines while maintaining statistical validity (Adams et al., 2019). The U.S. context thus provides a rigorous testing ground for assessing whether AI assistance produces measurable improvements in

underwriting accuracy under regulatory, operational, and data-driven constraints.
The international significance of AI-assisted underwriting models emerges from the shared structural challenges faced by insurance systems worldwide. Insurance markets across jurisdictions confront similar issues related to risk pooling, capital adequacy, and consumer protection (Eckert & Osterrieder, 2020). As insurers operate globally through reinsurance arrangements and multinational portfolios, underwriting accuracy influences not only local pricing but also global risk transfer mechanisms. Advances in underwriting analytics developed within one market can propagate through international insurance networks, shaping capital allocation and risk-sharing practices. Regulatory authorities in multiple regions have recognized that algorithmic underwriting systems can materially affect access to financial protection and market fairness. As a result, governance frameworks increasingly emphasize documentation, accountability, and monitoring of AI-driven decision tools. These frameworks reflect a global convergence toward treating underwriting models as high-impact systems requiring structured oversight (Upreti et al., 2022). From a quantitative perspective, this convergence elevates the importance of transparent evaluation methodologies, reproducible performance metrics, and ongoing model validation. International insurance markets also differ in exposure characteristics, legal environments, and hazard profiles, creating distributional shifts that challenge model generalization. AI-assisted underwriting models must therefore be evaluated for stability across varying conditions rather than optimized solely for a single dataset. This requirement underscores the value of robust validation strategies and sensitivity analysis in underwriting research. The international dimension also highlights the need for harmonized definitions of accuracy, fairness, and model risk. Quantitative studies that articulate these concepts clearly contribute to a shared analytical foundation that supports cross-border dialogue among insurers, regulators, and researchers (Poufinas et al., 2023). AI-assisted underwriting thus occupies a position at the intersection of technical modeling, institutional governance, and global insurance economics.
From a statistical learning standpoint, AI-assisted underwriting can be conceptualized as a sequence of quantitative design choices that collectively determine model performance. These choices include the specification of target variables, the construction of feature sets, the selection of model architectures, and the implementation of validation protocols (Kwiecień et al., 2020). Insurance loss data often display characteristics such as zero inflation, heavy tails, and heteroscedastic variance, requiring models that can accommodate nonstandard distributions. AI-based methods provide flexibility in this regard by learning functional forms directly from data. However, flexibility alone does not guarantee meaningful accuracy improvements. Quantitative evaluation must ensure that model performance gains persist across time periods and population segments (Dambra et al., 2020). Proper scoring rules and probabilistic calibration techniques offer principled tools for comparing predictive models in underwriting contexts. When underwriting decisions involve classification thresholds, discrimination metrics provide insight into ranking performance, while calibration metrics assess the reliability of predicted probabilities. Cost-sensitive evaluation further aligns model assessment with underwriting economics by weighting errors according to their financial impact. These methodological considerations are essential for translating statistical performance into operational relevance. Interpretation techniques also play a role in underwriting research by enabling analysts to examine how models use input variables to generate predictions (Frees & Huang, 2023). Such examination supports internal review processes and contributes to model governance. Within quantitative studies, interpretability can be treated as an empirical property subject to measurement rather than as an abstract requirement. This framing reinforces the view of AI-assisted underwriting as a measurable system whose performance and behavior can be systematically analyzed.
Data expansion strategies are central to the promise of AI-assisted underwriting accuracy improvements. Insurers increasingly integrate traditional underwriting variables with behavioral, transactional, and contextual data to capture a more comprehensive picture of risk (Panda et al., 2019). Quantitatively, the value of such data lies in its incremental contribution to predictive performance beyond established variables. This contribution must be assessed through controlled experimentation and validation. Behavioral data sources, such as driving patterns or activity indicators, exemplify how granular information can enhance risk differentiation. However, expanded data also introduces challenges related to correlation, redundancy, and stability. Highly correlated features can inflate

apparent model performance while reducing robustness. AI-assisted underwriting models must therefore be evaluated for sensitivity to feature inclusion and exclusion. Missing data handling, encoding strategies, and temporal alignment further influence accuracy estimates (King et al., 2021). In insurance settings, labels are often affected by prior underwriting decisions, creating selection effects that complicate supervised learning assumptions. Quantitative studies must address these effects through careful cohort construction and evaluation design. Additionally, fairness and bias considerations intersect with data expansion, as alternative data may correlate with structural inequalities. Measuring accuracy in isolation from distributional outcomes provides an incomplete picture of model behavior. Comprehensive evaluation frameworks incorporate subgroup analysis and monitoring to ensure that performance improvements are not confined to aggregate metrics (Lim et al., 2021). These data-centric considerations reinforce the complexity of AI-assisted underwriting as a system-level quantitative problem rather than a single-model optimization task.

Figure 2: AI-Assisted Insurance Underwriting Process



AI-assisted underwriting models in U.S. insurance markets ultimately function within institutional environments that define how accuracy is interpreted and applied. Underwriting outputs influence acceptance decisions, pricing tiers, and policy conditions, all of which have direct financial and social consequences (Denuit & Trufin, 2019). Quantitative accuracy gains must therefore be contextualized within decision workflows that include human oversight, compliance checks, and operational constraints. Selection effects arising from underwriting decisions can alter the data-generating process over time, affecting observed outcomes and model evaluation. Robust study designs account for these dynamics through temporal validation and monitoring strategies. Interpretability and transparency requirements further shape how AI-assisted underwriting models are assessed. Explanatory tools provide insights into model behavior, yet their reliability depends on underlying data properties and model stability (Bekemeier, 2023). Treating explanation quality as an empirical characteristic aligns interpretability with quantitative evaluation principles. International regulatory developments reinforce the expectation that underwriting models be subject to continuous oversight and documentation. These expectations influence how insurers define acceptable accuracy improvements and how researchers frame empirical evidence. By situating AI-assisted underwriting within established insurance economics, statistical learning theory, and governance structures, quantitative research can systematically examine whether and how AI assistance improves risk assessment accuracy. This framing supports a rigorous introduction to the study of AI-assisted underwriting models as measurable, evaluable systems embedded in U.S. insurance markets with global relevance

(Zeier Röschmann et al., 2022).

The objective of this quantitative study is to evaluate whether AI-assisted underwriting models improve risk assessment accuracy in U.S. insurance markets when compared with conventional underwriting approaches, using standardized predictive performance and stability criteria grounded in real underwriting data. The study aims to measure improvement as a quantifiable change in the alignment between model-generated risk estimates and observed insurance outcomes, including claim occurrence, claim frequency, claim severity, and aggregate loss cost at the policy level. A primary objective is to determine the extent to which AI-assisted models enhance discrimination and calibration, meaning the models' ability to correctly rank policyholders by risk and to produce probability and loss estimates that match realized outcomes across the full risk distribution. Another objective is to test the robustness of AI-assisted underwriting performance across major market segments and underwriting contexts within the United States, including variation by product line, geography, exposure profile, and time period, with the goal of identifying whether predictive gains are consistent rather than concentrated in narrow subpopulations. The study further aims to compare multiple AI model families—such as tree-based ensembles and neural network approaches—against benchmark statistical underwriting models under identical data preprocessing, feature sets, and validation designs to ensure that observed differences are attributable to model capability rather than evaluation artifacts. An additional objective is to quantify the contribution of expanded underwriting inputs commonly associated with AI adoption, such as enriched behavioral or contextual variables, by estimating incremental predictive value over traditional underwriting factors and testing sensitivity to feature inclusion rules. The research also seeks to assess model stability under operational conditions relevant to U.S. insurers, including performance under temporal shifts, differences in claim reporting patterns, and changes in exposure distributions, using validation approaches that mimic real underwriting deployment timelines. Finally, the study aims to produce an objective, replicable evaluation framework that translates predictive performance results into underwriting-relevant accuracy indicators, enabling clear comparison of AI-assisted and traditional underwriting models while remaining focused on measurable accuracy outcomes in the U.S. insurance market setting.

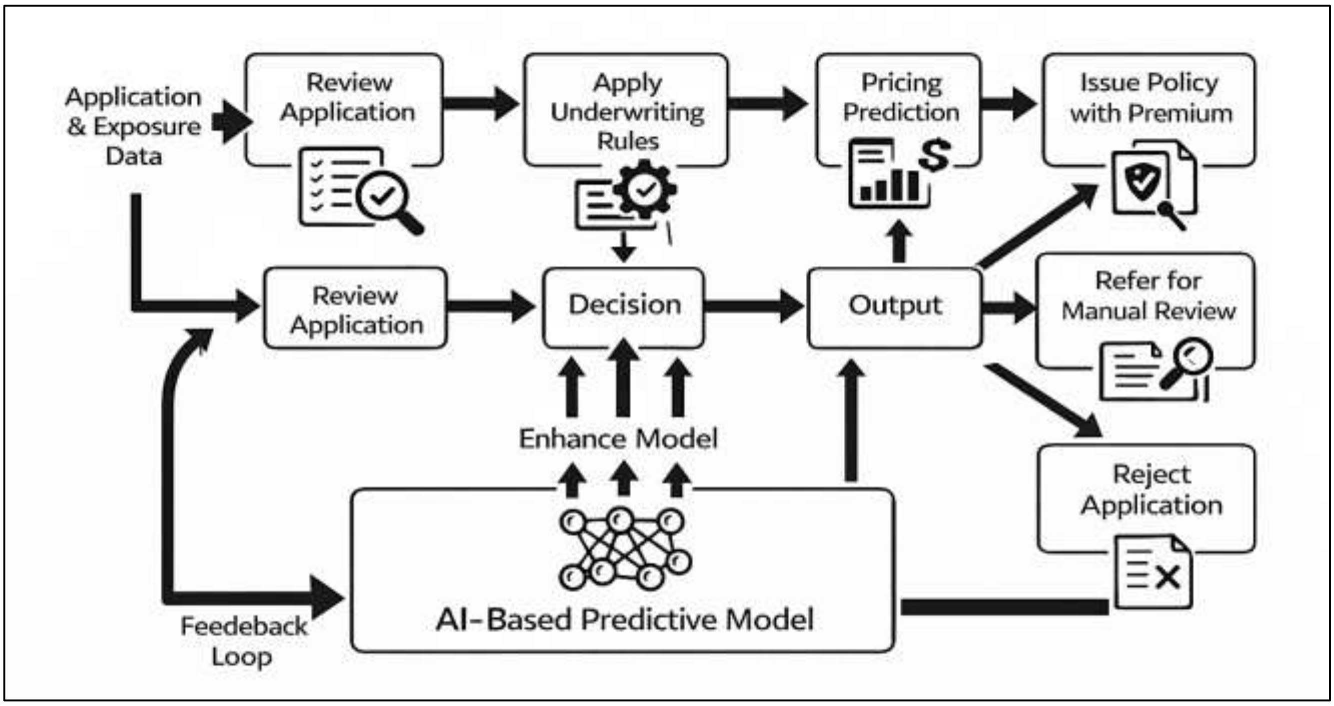
LITERATURE REVIEW

The literature review for AI-Assisted Underwriting Models for Improving Risk Assessment Accuracy in U.S. Insurance Markets synthesizes empirical and methodological research that explains how underwriting accuracy is defined, measured, and improved through data-driven modeling (Cannon & Preis, 2023). Because underwriting is a predictive decision pipeline, prior studies are reviewed through a quantitative lens that prioritizes measurable outcomes—such as calibration, discrimination, error costs, and stability—rather than purely conceptual claims. This section positions AI-assisted underwriting as an evolution of risk scoring and classification systems that historically relied on actuarial and statistical approaches, then progressively incorporated higher-dimensional data and algorithmic learning methods. The review is organized to build from foundational underwriting constructs to modern AI architectures, then to evaluation frameworks that determine whether "improvement" is statistically valid and operationally meaningful in U.S. markets. Special attention is given to how prediction targets (claim probability, expected loss, tail loss, combined loss ratio) differ by line of business and how those targets shape model selection and metric choice (Keenan & Bradt, 2020). The literature is also examined for evidence about data sources commonly used in U.S. insurance underwriting, including traditional policy and claims features and enriched external or behavioral variables, because predictive gains depend heavily on feature relevance and data integrity. In addition, the review prioritizes comparative studies that benchmark AI models against conventional baselines under consistent validation designs, since differences in sampling, leakage control, and time-based splits often explain apparent performance advantages. Finally, the section integrates research on model governance-related technical requirements—such as interpretability, stability, and monitoring—only to the extent that they influence quantitative reliability and reproducibility of accuracy estimates in regulated underwriting environments (Patil et al., 2023).

Foundations of Underwriting Accuracy

Underwriting in insurance markets can be defined as a structured predictive system that evaluates risk characteristics and translates them into standardized coverage and pricing decisions. From a quantitative standpoint, underwriting functions as an organized process through which insurers assess applicant- and exposure-level information to support consistent decision-making across large portfolios (Pu & Lam, 2021). The outputs of underwriting include acceptance or declination of coverage, referral of applications for manual review, assignment to predefined risk tiers, modification of deductibles or coverage limits, and determination of premium indications. Each output represents an operational expression of an underlying risk assessment that estimates expected loss and uncertainty. These decisions are not isolated judgments but are generated through systematic evaluation mechanisms that rely on modeled representations of risk. In practice, underwriting models aggregate multiple data elements describing insured characteristics, historical experience, geographic context, and behavioral indicators into a consolidated assessment (Nurse et al., 2020). This assessment serves as the basis for applying underwriting rules and thresholds that produce final actions. In U.S. insurance markets, underwriting systems must operate at scale, handling millions of policies while maintaining consistency across regions and product lines. The complexity of these markets necessitates predictive systems that are both efficient and reproducible. AI-assisted underwriting models are situated within this framework as tools designed to enhance how risk information is processed and synthesized. By applying algorithmic learning techniques to large and diverse datasets, these models aim to capture complex relationships among variables that influence insurance outcomes. Conceptualizing underwriting as a predictive classification and pricing system provides a foundation for treating underwriting performance as a measurable quantity (David-Spickermann et al., 2021). This framing allows underwriting accuracy to be evaluated empirically by comparing model-based risk assessments with realized insurance outcomes, establishing a basis for quantitative analysis of underwriting effectiveness.

Figure 3: AI-Assisted Insurance Underwriting Workflow



Risk assessment accuracy in insurance underwriting requires an operational definition that reflects the multiple functions underwriting models perform. Accuracy extends beyond simple correctness and must be understood as a composite construct encompassing several interrelated dimensions (Brotcke, 2022). One essential dimension is the ability of a model to correctly distinguish between higher-risk and lower-risk policies. This capability supports the ordering of applicants according to expected loss and is fundamental to segmentation and tier assignment. Another dimension concerns the reliability of predicted risk levels, referring to how closely estimated probabilities or expected losses correspond to observed outcomes across different segments of the insured population. Reliable estimates support pricing consistency and reduce systematic misalignment between premiums and realized losses. A further dimension of accuracy relates to the economic consequences of underwriting errors. In insurance contexts, errors carry asymmetric financial implications (Dambra et al., 2020). Underestimating risk can expose insurers to adverse selection, inadequate pricing, and portfolio deterioration, while overestimating risk can reduce competitiveness and restrict coverage availability. Accuracy must therefore be evaluated in relation to the financial relevance of errors rather than solely through statistical deviation. A fourth dimension involves performance consistency across time periods, geographic regions, and underwriting segments. Consistency is particularly important in U.S. insurance markets, where regulatory environments, exposure distributions, and loss dynamics vary across states and lines of business. Models that perform well in one segment but poorly in others introduce operational and regulatory challenges (Mourmouris & Poufinas, 2023). Defining risk assessment accuracy as a multi-dimensional construct ensures that quantitative evaluation captures the full range of underwriting objectives and constraints, providing a comprehensive framework for assessing underwriting model performance.

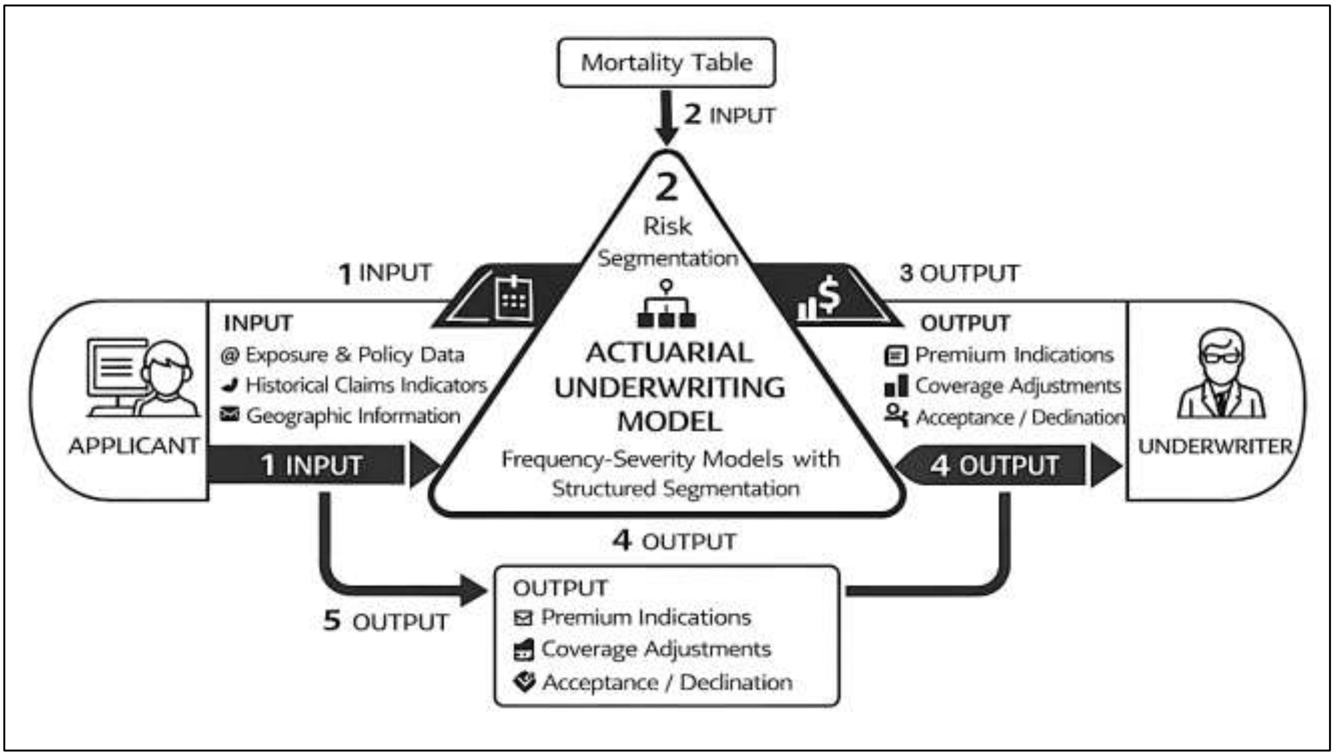
The measurement of underwriting accuracy differs substantially from accuracy assessment in general predictive modeling due to the distinctive characteristics of insurance data. Insurance outcomes frequently involve infrequent events, particularly in claim frequency modeling, where a large proportion of policies do not generate claims within an observation period (Tarr et al., 2021). This imbalance limits the usefulness of simple accuracy measures and necessitates evaluation approaches that emphasize relative risk ordering and probability estimation. In addition, insurance loss outcomes often display substantial variability, with a small number of claims accounting for a disproportionately large share of total losses. This heavy-tailed behavior complicates evaluation because extreme outcomes can exert outsized influence on performance metrics. Underwriting models must therefore be assessed in a manner that reflects both typical and extreme loss behavior (Kong, 2021). Differences across insurance lines further shape how accuracy is defined and measured. Property and casualty insurance often involves relatively short claim reporting periods, while liability and certain health insurance products involve longer reporting and settlement timelines. Life insurance underwriting focuses on mortality and persistency risks, whereas health insurance underwriting emphasizes utilization intensity and cost accumulation. These distinctions affect target variable construction, validation horizons, and metric selection. As a result, underwriting accuracy must be evaluated using criteria that align with insurance-specific data structures and decision contexts rather than generic prediction benchmarks (Lozano-Murcia et al., 2023). This distinction is particularly important when assessing AI-assisted underwriting models, which may demonstrate strong performance under conventional metrics while exhibiting weaknesses in insurance-relevant dimensions such as calibration stability or loss sensitivity.

In U.S. insurance markets, underwriting accuracy also has institutional and operational significance that shapes how it is conceptualized and evaluated. Underwriting decisions influence premium levels, coverage availability, and portfolio composition, making accuracy central to both financial performance and regulatory oversight (S. Xu et al., 2022). Because underwriting models affect which policies are issued, observed outcomes are conditioned on prior underwriting decisions, creating selection effects that complicate performance evaluation. Quantitative assessments of accuracy must account for these dynamics to avoid overstating model effectiveness. Stability of performance across market segments is particularly critical in the United States, where insurers operate across states with differing regulatory standards, legal environments, and exposure profiles. AI-assisted underwriting models introduce additional complexity by incorporating larger feature sets and more flexible learning structures (Yan, 2023). While these models offer potential improvements in risk differentiation, they also increase sensitivity to data quality, feature correlations, and distributional shifts. Defining underwriting accuracy as a structured, multi-dimensional construct provides a framework for evaluating these models in a manner that aligns with operational realities. This framework emphasizes ranking performance, reliability of estimates, economic relevance of errors, and consistency across conditions. Establishing such a foundation is essential for conducting rigorous quantitative analysis of AI-assisted underwriting models and their role in improving risk assessment accuracy within the diverse and regulated environment of U.S. insurance markets (Wickham et al., 2020).

Conventional Underwriting Models

Conventional underwriting models in insurance markets have historically relied on actuarial and statistical techniques designed to estimate risk using structured assumptions and interpretable parameter relationships (Lai et al., 2020). At the core of these approaches is the decomposition of insurance loss into frequency and severity components, which allows insurers to model claim occurrence and claim magnitude separately. This structure aligns with the operational needs of underwriting by enabling distinct treatment of event likelihood and financial impact. These models typically assume stable relationships between explanatory variables and outcomes, allowing risk factors such as exposure characteristics, policyholder attributes, and historical experience to be incorporated in a controlled and transparent manner (Ayranci et al., 2022). Segmentation plays a central role in traditional underwriting, as risks are grouped into relatively homogeneous classes based on predefined characteristics. Credibility-style logic is often applied to balance individual experience with collective information, ensuring that estimates remain stable when data are sparse. This framework supports consistency across underwriting decisions and facilitates aggregation at portfolio levels. In U.S. insurance markets, these models have been widely adopted because they align with regulatory expectations for fairness, transparency, and documentation (Rawat, 2023). The structure of traditional actuarial models allows underwriters and regulators to trace how individual variables influence outcomes, reinforcing confidence in underwriting decisions. These models also support pricing, reserving, and capital assessment functions, making them foundational tools within insurance organizations. Their long-standing use has resulted in extensive institutional knowledge and standardized workflows, reinforcing their role as benchmarks against which newer modeling approaches are evaluated (Brown et al., 2019).

Figure 4: Conventional Insurance Underwriting Process Framework



One of the primary strengths of conventional underwriting models lies in their interpretability and operational stability. Because these models rely on explicit relationships between variables and outcomes, stakeholders can readily understand how risk factors contribute to underwriting decisions (Billio et al., 2022). This clarity is particularly important in regulated insurance environments, where insurers must explain pricing and coverage decisions to regulators, auditors, and consumers. The transparency of traditional models supports regulatory review processes and simplifies compliance documentation. Stability is another key strength, as these models tend to produce consistent results across time periods when underlying risk relationships remain relatively unchanged. This predictability supports long-term portfolio management and financial planning. Baseline underwriting workflows commonly involve manual feature selection and transformation, including binning of continuous variables into discrete categories (MacAskill et al., 2021). Such practices allow insurers to enforce monotonic relationships between risk factors and outcomes, ensuring that increases in risk indicators correspond to non-decreasing risk estimates. Interaction effects are typically restricted or introduced selectively to preserve interpretability and reduce estimation variance. These constraints contribute to model robustness and reduce sensitivity to noise in the data. In comparative studies, these characteristics make conventional models valuable benchmarks because their behavior is well understood and their limitations are clearly defined (L. Li et al., 2022). Their widespread adoption across U.S. insurance lines provides a consistent reference point for evaluating whether alternative modeling approaches deliver meaningful improvements in underwriting accuracy.

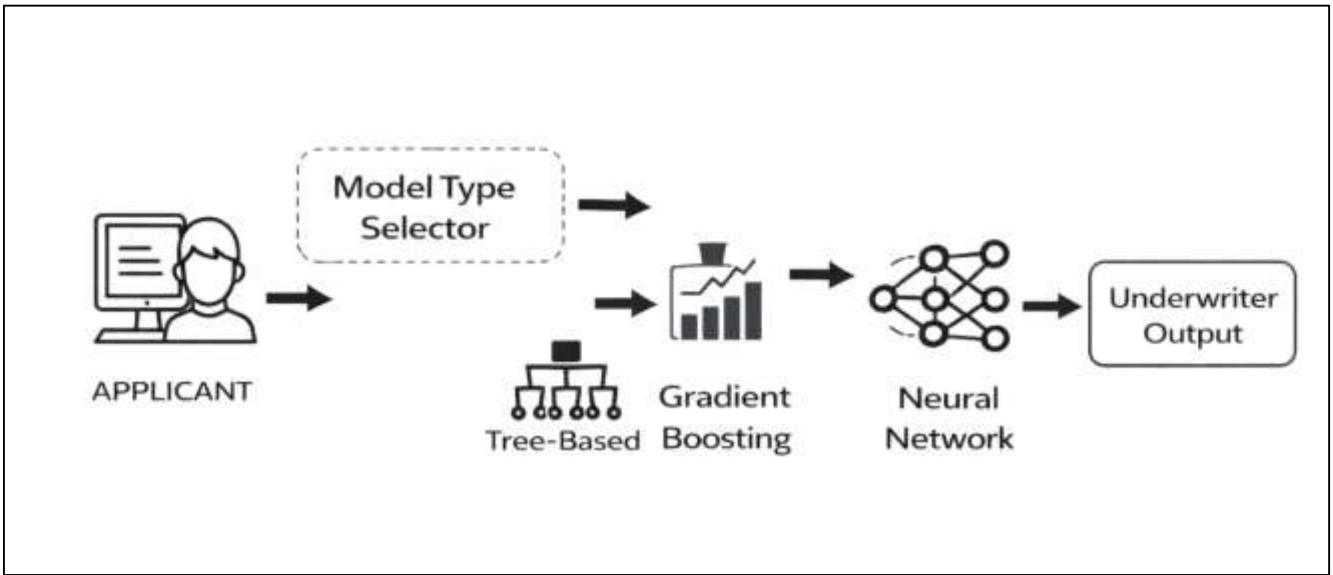
The limitations of conventional underwriting models become apparent as insurance data environments grow in complexity. Traditional actuarial models are inherently constrained by their reliance on predefined functional forms and manually engineered features (Grundl & Kim, 2019). As the number of potential risk factors increases, especially with the inclusion of behavioral, contextual, and interaction-rich variables, linear and segmented modeling structures struggle to capture complex dependencies. Nonlinear relationships among variables, which may significantly influence loss outcomes, are often approximated through coarse segmentation or omitted entirely. High-dimensional datasets further challenge conventional models, as multicollinearity and sparse data structures complicate estimation and interpretation (Reguero et al., 2020). These limitations can lead to underfitting, where important risk patterns remain unmodeled. Sensitivity to feature engineering is another concern, as model performance depends heavily on how variables are transformed and categorized. Small changes in binning schemes or interaction definitions can produce materially different results, introducing subjectivity into the modeling process. In addition, fixed functional assumptions restrict the ability of conventional models to adapt to evolving risk landscapes, such as changes in policyholder behavior, economic conditions, or environmental exposure. These constraints do not negate the value of traditional underwriting models but highlight why insurers and researchers have explored AI-assisted alternatives (Xu, 2022). Understanding these limitations is essential for framing comparative studies, as performance differences between baseline and AI-assisted models often reflect differences in representational flexibility rather than flaws in the underlying actuarial logic.

For the purposes of quantitative comparison, conventional underwriting models serve as the baseline specification against which AI-assisted underwriting models are evaluated. In this study context, baseline models are defined as frequency–severity–based actuarial models using structured segmentation and credibility-informed estimation procedures (Jacobs Jr, 2020). These models employ a fixed set of underwriting variables commonly used in U.S. insurance markets, including exposure measures, historical claims indicators, geographic classifications, and policy characteristics. Feature transformations follow established actuarial workflows, including categorical grouping and monotonic risk ordering. Interaction effects are limited to predefined combinations that reflect domain knowledge rather than data-driven discovery (Dragos et al., 2023). The baseline specification emphasizes interpretability, regulatory compatibility, and stability, ensuring that any observed performance differences can be attributed to modeling capability rather than evaluation artifacts. By holding data inputs, preprocessing steps, and validation procedures constant across baseline and AI-assisted models, the comparative framework isolates the incremental value of algorithmic learning. This approach reflects standard practice in empirical insurance analytics, where new methods are assessed relative to well-understood benchmarks. The baseline models selected for comparison represent prevailing underwriting practices in U.S. insurance markets and provide a rigorous reference point for evaluating improvements in risk assessment accuracy (Hutton, 2020). Establishing this specification ensures that comparative results are interpretable, reproducible, and relevant to real underwriting environments.

AI-Assisted Underwriting Models

AI-assisted underwriting models based on tree-based ensemble methods have gained substantial attention in insurance analytics due to their strong alignment with the structural properties of underwriting data (Alam et al., 2019). Insurance underwriting datasets commonly contain a mixture of numerical, categorical, and ordinal variables derived from policy characteristics, exposure measures, geographic indicators, and historical experience. Tree-based ensembles are well suited to this environment because they natively handle mixed data types without requiring extensive transformation. Missing values, which frequently arise in insurance data due to incomplete applications or optional disclosures, can be accommodated naturally within tree-splitting logic, reducing the need for imputation strategies that may distort risk signals. Another key capability of tree-based ensembles is their ability to capture nonlinear relationships and interaction effects among underwriting variables (Mahapatra & Singh, 2021). Risk relationships in insurance are rarely additive, and interactions between factors such as location, asset characteristics, and prior claims often influence loss outcomes in complex ways. Tree-based structures identify these interactions directly through hierarchical splits, allowing models to represent localized risk patterns across subpopulations. From an underwriting perspective, this flexibility improves risk differentiation while preserving a degree of interpretability through feature importance measures and decision path analysis. These properties make tree-based ensembles particularly suitable for underwriting risk scoring tasks where the dependent variables include claim occurrence indicators or expected loss measures derived from historical experience. In the literature, these models are frequently evaluated as enhancements over conventional actuarial baselines because they relax linear assumptions while maintaining operational feasibility (Ozmen Garibay et al., 2023). Their performance stability across heterogeneous portfolios and their compatibility with tabular insurance data reinforce their role as a foundational AI-assisted underwriting model family.

Figure 5: AI-Based Underwriting Model Flow



Gradient boosting systems represent a specialized and widely adopted subclass of tree-based ensembles that have demonstrated strong performance in structured tabular prediction tasks, including insurance underwriting (Javed et al., 2023). Underwriting data typically exhibit strong but fragmented predictive signals distributed across many variables, rather than a small number of dominant predictors. Gradient boosting systems are effective in such environments because they iteratively refine model performance by focusing on residual error patterns, allowing weak learners to collectively capture complex structures. This approach aligns well with underwriting tasks where risk signals are incremental and context-dependent. Insurance datasets often include extensive feature engineering, such as derived exposure measures, interaction proxies, and normalized indicators, which further enhance the effectiveness of gradient boosting methods (Cox Jr, 2023). These systems excel when the underlying signal-to-noise ratio is moderate and when relationships between predictors and outcomes are nonlinear but stable within localized regions of the feature space. In underwriting applications, gradient boosting models are commonly used to estimate claim probability, loss cost, or composite risk scores that inform tiering and pricing decisions. Their structured learning process supports consistent improvements in discrimination and calibration relative to simpler ensemble approaches. Importantly, gradient boosting models are compatible with established underwriting validation practices, including time-based testing and segment-level evaluation. Their widespread adoption in insurance analytics literature reflects their balance between predictive power and operational practicality (Kumar, 2023). When underwriting is conceptualized as a tabular machine learning problem, gradient boosting systems emerge as a natural candidate due to their ability to leverage rich engineered features while remaining computationally efficient and robust to common data irregularities.

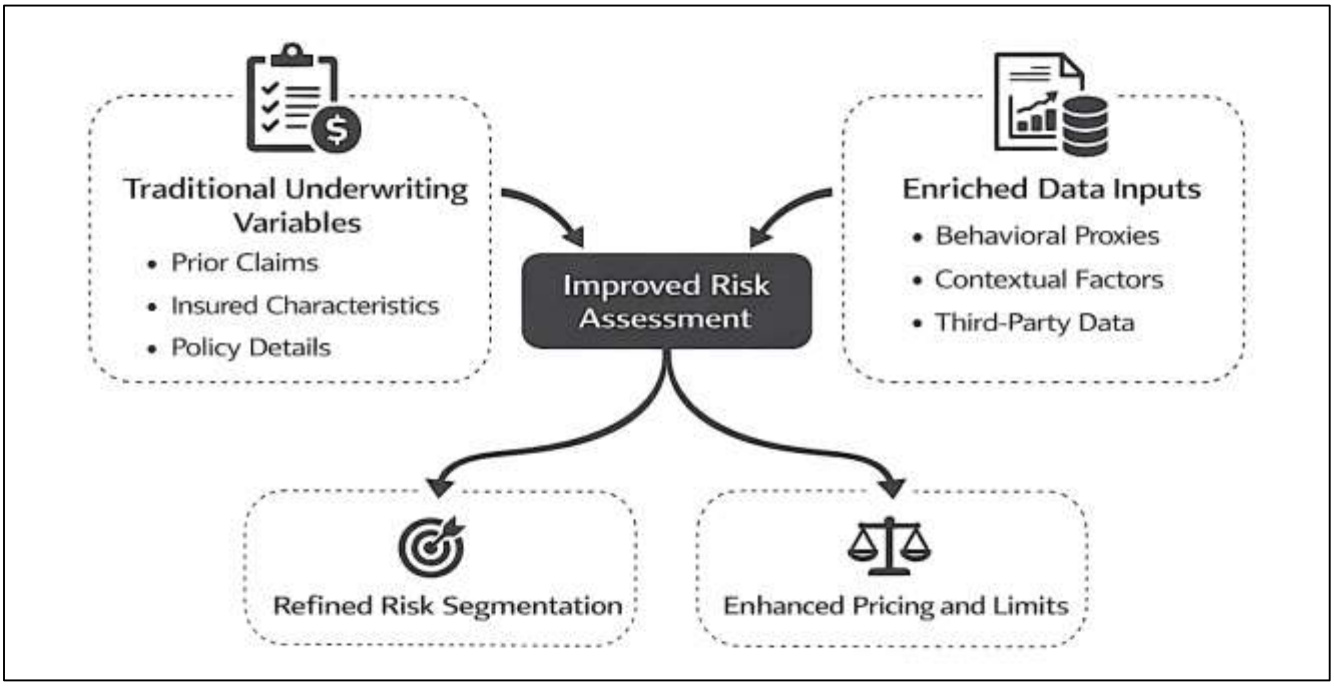
Neural network models represent another major family of AI-assisted underwriting approaches, particularly in contexts where underwriting data extend beyond traditional tabular structures. While neural networks have historically been less common in core underwriting tasks, their use has expanded as insurers incorporate high-dimensional inputs and complex data representations (Banasiewicz, 2021b). Neural networks are particularly well suited to handling large numbers of correlated variables and learning abstract representations through layered transformations. In underwriting applications, this capability is relevant when categorical variables have high cardinality or when embedding techniques are used to represent categorical information in continuous space. Neural architectures are also applicable when underwriting incorporates behavioral or sequential data, such as driving patterns, transaction histories, or longitudinal exposure signals. These data types challenge traditional actuarial models due to their temporal and multivariate nature (Peddie, 2019). Neural networks can integrate such inputs into unified risk assessments by learning representations that summarize patterns over time or across dimensions. In the literature, neural underwriting models are often evaluated for their ability to capture subtle risk signals that emerge only in high-dimensional contexts. Their flexibility allows them to approximate complex loss-generating processes without predefined functional assumptions (Banasiewicz, 2021a). However, their successful application depends on sufficient data volume, careful regularization, and robust validation design. In insurance underwriting, neural networks are most commonly applied where the dependent variables involve frequency estimation, expected loss prediction, or composite risk scoring derived from enriched datasets. Their role within AI-assisted underwriting reflects the expanding scope of underwriting data rather than a replacement of traditional actuarial logic (Yew, 2020).

Underwriting Data Inputs in U.S. Markets

Traditional underwriting variables form the foundational data inputs for risk assessment in U.S. insurance markets and have long served as the primary determinants of underwriting decisions across major lines of business (Keenan & Bradt, 2020). These variables are selected for their demonstrated relevance to loss experience, regulatory acceptability, and operational consistency. In property and casualty insurance, underwriting commonly incorporates policy limits and deductibles, which define the insurer's exposure to loss and influence claim severity distributions. Prior claims history provides direct evidence of risk behavior and loss propensity, often segmented by frequency, recency, and magnitude. Insured characteristics, such as age, tenure, and occupancy status, contribute to risk classification by capturing demographic and behavioral patterns associated with loss outcomes (Hofmann & Sattarhoff, 2023). Property insurance underwriting relies heavily on structural attributes, including construction type, age of the building, location characteristics, and protective features, all of which affect hazard exposure. Automobile insurance underwriting incorporates vehicle characteristics such as age, type, usage patterns, and safety features, along with driver-related indicators. In workers' compensation and commercial lines, occupational risk class and industry classification serve as core underwriting variables, reflecting differential exposure to workplace hazards. These traditional variables are typically well-defined, standardized, and supported by historical loss data, enabling consistent modeling across portfolios. Quantitatively, they provide a stable feature space that supports risk segmentation and pricing under established actuarial frameworks (Patil et al., 2023). Their long-standing use has resulted in mature data governance practices, including clear timestamp alignment and documentation standards. As a result, traditional underwriting variables are often treated as baseline input set in empirical studies evaluating underwriting accuracy. They establish a reference point against which the incremental value of expanded data sources can be measured. In U.S. insurance markets, where regulatory scrutiny emphasizes transparency and fairness, these variables also serve as anchors for model validation and review processes.

Enriched and alternative data sources have increasingly been incorporated into AI-assisted underwriting systems to enhance risk differentiation beyond what is achievable with traditional variables alone (Barkham et al., 2022). These data sources include behavioral proxies that capture patterns of activity or decision-making, contextual variables that describe environmental or situational factors, and third-party indicators derived from external datasets. Behavioral proxies may reflect usage intensity, transaction regularity, or adherence to safety-related behaviors, providing indirect signals of risk that are not captured by static policy attributes. Contextual variables can include geographic indicators, environmental exposure measures, or neighborhood-level characteristics that influence loss likelihood. Third-party data sources often aggregate information across multiple domains, offering synthesized indicators that correlate with insurance outcomes (Nurse et al., 2020). In certain lines, telematics data provide granular behavioral information, particularly in automobile insurance, where driving patterns can be summarized into risk-relevant features. Quantitatively, these enriched inputs expand the dimensionality of the underwriting feature space and introduce complex correlation structures. Their appeal lies in their potential to capture latent risk factors and interaction effects that traditional variables may only approximate. However, their integration into underwriting models requires careful consideration of data quality, consistency, and relevance. Enriched data may vary in coverage across policyholders, introduce missingness patterns, or rely on indirect measurement of risk-related behavior. From an analytical perspective, these characteristics necessitate robust preprocessing and validation strategies. In U.S. insurance markets, enriched data inputs are typically layered onto existing underwriting datasets rather than replacing traditional variables (Kim, 2019). This layered approach supports comparative analysis of model performance and facilitates incremental adoption within operational constraints. Literature examining AI-assisted underwriting frequently emphasizes the importance of distinguishing the predictive contribution of enriched data from that of algorithmic modeling techniques, underscoring the need for structured evaluation designs.

Figure 6: Core Data for Underwriting Accuracy



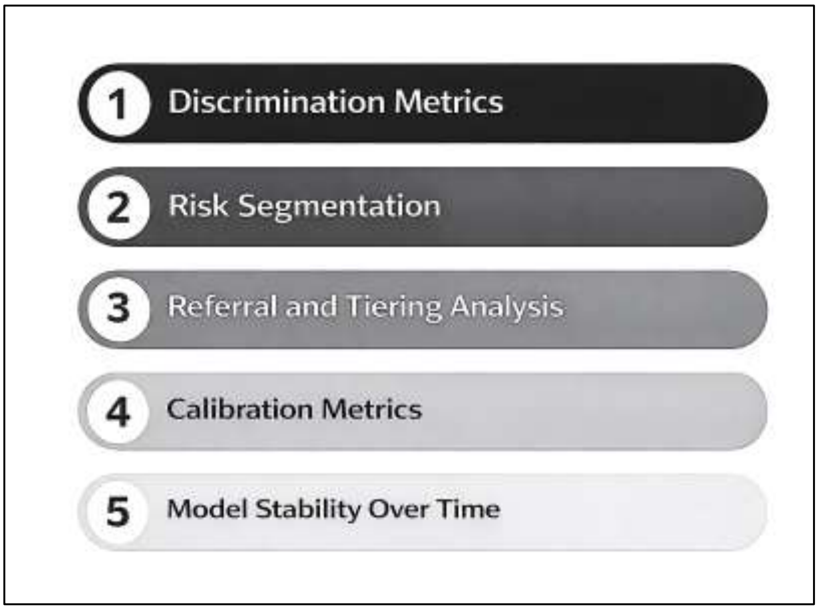
Measuring the incremental predictive value of enriched underwriting features requires a systematic quantitative approach that isolates the contribution of expanded data inputs. In underwriting research, this is commonly achieved through comparative modeling frameworks that evaluate performance differences between models using only traditional variables and models incorporating both traditional and enriched variables (Kester, 2022). Such comparisons are conducted under identical preprocessing, training, and validation conditions to ensure that observed differences reflect feature contribution rather than methodological artifacts. This approach allows researchers to assess whether enriched data improve risk discrimination, calibration, or economic relevance beyond established inputs. Incremental value is evaluated across multiple performance dimensions, recognizing that enriched features may enhance ranking ability without necessarily improving probability reliability or cost alignment. Segment-level analysis further clarifies whether predictive gains are concentrated within specific subpopulations or distributed broadly across the insured portfolio (Gohdes et al., 2022). In U.S. insurance markets, incremental value assessment is particularly important because the inclusion of alternative data can introduce complexity into underwriting workflows and governance processes. Quantitative studies therefore emphasize robustness checks that examine performance stability across time periods and underwriting segments. The objective is to determine whether enriched features contribute consistent improvements rather than isolated gains driven by specific cohorts or temporal conditions. By structuring evaluation around controlled comparisons, underwriting research establishes a transparent basis for attributing performance improvements to data expansion. This approach aligns with actuarial principles of model validation and supports clear interpretation of results in regulated environments (Cao et al., 2023). Incremental predictive value analysis thus serves as a critical bridge between data innovation and underwriting accuracy assessment.

Data-related risks present significant challenges to estimating underwriting accuracy gains and must be addressed explicitly in quantitative analysis (Demchuk, 2022). One major risk is label leakage, which occurs when information related to future outcomes is inadvertently included in model inputs, leading to inflated performance estimates. In underwriting contexts, leakage can arise from the inclusion of post-binding information or variables that are influenced by claims events. Inconsistent timestamp alignment across data sources further complicates analysis, as enriched datasets may be updated at different intervals than policy and claims records. Claim reporting lags introduce additional complexity, particularly in lines with delayed loss emergence, where observed outcomes may not fully reflect underlying risk during the evaluation period (Eling & Schnell, 2020). Selective observation represents another critical risk, as underwriting decisions influence which policies are issued and therefore which outcomes are observed. Models trained on issued policies may reflect selection effects rather than true population risk. Quantitative studies must account for these dynamics through careful cohort definition and validation design. In U.S. insurance markets, where underwriting rules vary across states and product lines, data risks can differ substantially by segment. Addressing these risks requires explicit safeguards, including strict separation of training and evaluation periods, exclusion of post-decision variables, and consistent handling of missing and delayed data (Eling et al., 2023). By documenting these safeguards, underwriting research enhances the credibility of reported accuracy improvements. Recognizing and mitigating data risks is essential for ensuring that observed performance gains reflect genuine improvements in risk assessment rather than artifacts of data construction or evaluation bias.

Evaluation Framework

The evaluation of underwriting model improvement in insurance literature begins with discrimination metrics, which assess a model's ability to correctly order policyholders by relative risk. In underwriting contexts, discrimination is fundamental because many underwriting decisions depend on ranking rather than absolute prediction accuracy (Cichy & Rass, 2019). Risk tier assignment, referral prioritization, and segmentation strategies all rely on the model's capacity to distinguish higher-risk policies from lower-risk ones. Measures that assess ranking quality are therefore widely used when evaluating claim occurrence models and composite risk scores. In insurance datasets, where claim events are often infrequent, discrimination metrics provide insight into how effectively a model concentrates risk within higher score ranges. Decile-based analysis is commonly employed to examine how claims or losses accumulate across ordered risk groups, offering an intuitive view of model performance that aligns with underwriting workflows. Lift analysis further supports this perspective by quantifying how much more concentrated losses are in high-risk segments compared with a random allocation (Siegel et al., 2019). In underwriting literature, these approaches are valued for their interpretability and operational relevance. Discrimination metrics are particularly useful in comparative studies because they remain relatively stable under class imbalance and allow consistent comparison across models trained on the same data. In U.S. insurance markets, where underwriting decisions often involve thresholds rather than continuous outputs, discrimination measures provide a practical basis for assessing whether AI-assisted models offer superior risk ordering compared with conventional baselines (Han et al., 2020). However, the literature also emphasizes that discrimination alone does not capture the full notion of improvement, as models can rank risks well while producing unreliable probability estimates. As a result, discrimination metrics are typically treated as a necessary but insufficient component of underwriting evaluation frameworks.

Figure 7: Underwriting Model Evaluation Core Metrics

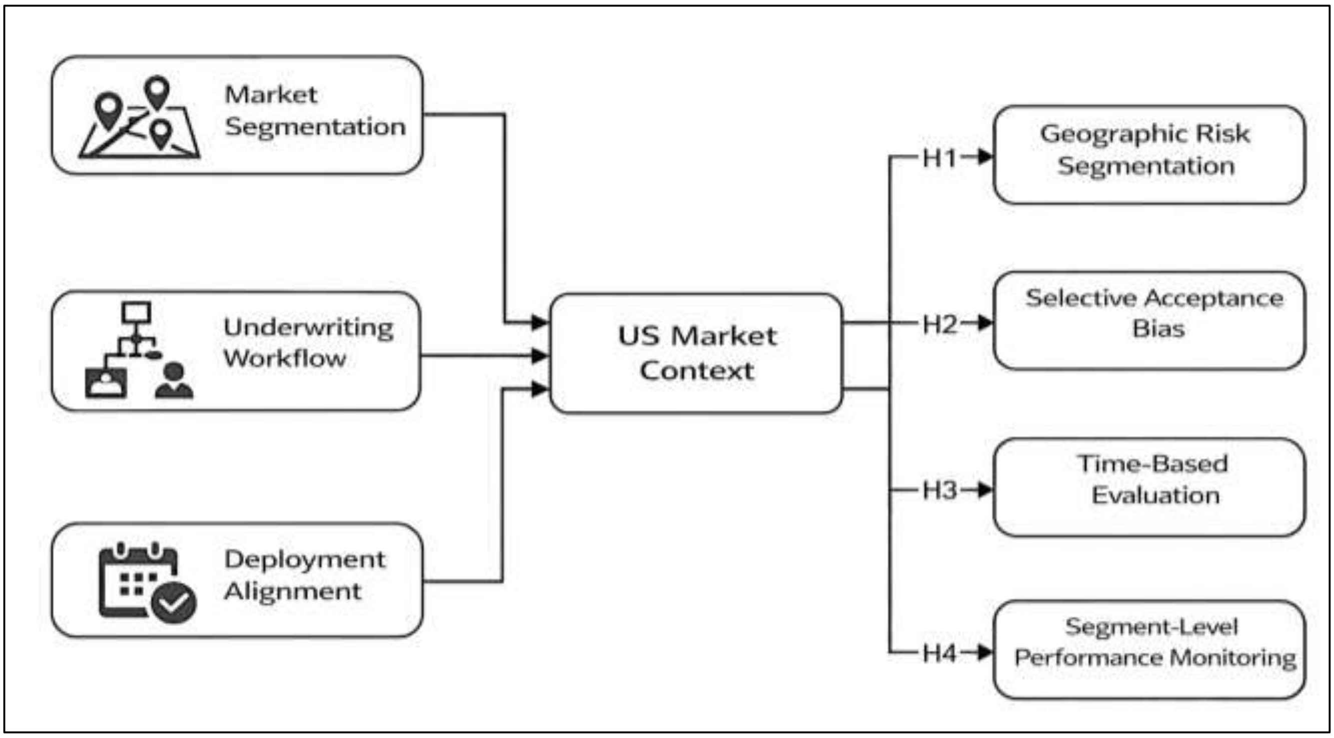


Calibration metrics and diagnostic tools address a complementary dimension of underwriting model performance by examining the alignment between predicted risk levels and observed outcomes. In insurance underwriting, calibration is essential because pricing, reserving, and capital planning depend on the accuracy of estimated probabilities and expected losses (Rauf, 2018; Saja et al., 2019). A model that consistently overestimates or underestimates risk can produce systematic pricing distortions even if its ranking performance appears strong (Haque & Arifur, 2020; Ashraful et al., 2020). Calibration analysis often involves comparing predicted values with realized outcomes across risk groups, allowing analysts to identify systematic bias or deviation. Score band analysis is commonly used in underwriting contexts to assess whether observed loss experience aligns with model predictions within defined risk segments (Haque & Arifur, 2021; Fokhrul et al., 2021). This approach reflects operational practices in which underwriting decisions are applied to groups rather than individual predictions (Fahimul, 2022; Kamble & Gunasekaran, 2020; Zaman et al., 2021). Aggregate calibration diagnostics further evaluate whether overall predicted losses match observed losses at portfolio or segment levels. In applied insurance studies, calibration is treated as a critical validation requirement because it directly affects financial outcomes. Poor calibration can lead to underpricing or overpricing across entire portfolios, undermining underwriting objectives. Calibration diagnostics are also used to assess model stability over time, as shifts in calibration may signal changes in underlying risk patterns or data drift. In comparative studies of AI-assisted underwriting models, calibration analysis helps determine whether observed improvements in discrimination translate into more reliable risk estimates (Hammad, 2022; Hasan & Waladur, 2022; Knudsen et al., 2019). The literature emphasizes that true improvement in underwriting accuracy requires both strong discrimination and sound calibration, reinforcing the need for multi-metric evaluation frameworks.

U.S. Market Context

The U.S. insurance market presents a uniquely complex environment that directly constrains quantitative underwriting model design and evaluation. Unlike centralized regulatory systems, insurance regulation in the United States is primarily state-based, resulting in substantial variation across jurisdictions in rating rules, underwriting restrictions, and reporting requirements (Kluve et al., 2019). These state-level differences affect how risk is classified, priced, and monitored, making geographic segmentation a central consideration in underwriting analytics. Territorial rating practices introduce further heterogeneity, as insurers subdivide states into smaller rating areas to capture localized risk patterns associated with traffic density, weather exposure, crime rates, and infrastructure characteristics. Catastrophe exposure adds another layer of variation, particularly in property insurance, where geographic concentration of natural hazards such as hurricanes, wildfires, floods, and hailstorms leads to highly uneven loss distributions (Pan et al., 2021). Legal environment variation also influences underwriting outcomes, as differences in tort law, claim settlement practices, and litigation frequency affect loss severity and reporting behavior across states. From a quantitative perspective, these factors create non-stationary data structures in which the same underwriting variables may have different implications depending on location. As a result, models trained on aggregated national data may obscure region-specific risk dynamics. The literature emphasizes that underwriting models must explicitly account for geographic heterogeneity to avoid biased estimates and unstable performance (Prokopy et al., 2019). In U.S. markets, this requirement transforms market context into a design constraint rather than a background consideration. Risk assessment accuracy cannot be evaluated independently of the segmentation structure that defines how risks are grouped and regulated. Understanding these market segmentation issues is therefore essential for interpreting model performance and for designing empirical studies that reflect real underwriting conditions.

Figure 8: U.S. Underwriting Context Framework



Underwriting workflow constraints further shape observed insurance outcomes and introduce additional complexity into quantitative model evaluation. Underwriting is not a single automated decision but a multi-stage process involving both algorithmic scoring and human judgment. Referral rules route certain applications to manual underwriting review based on predefined thresholds, documentation requirements, or risk flags (Yu et al., 2020). Manual underwriting interventions may override model recommendations, impose additional conditions, or decline coverage altogether. These interventions influence which policies are ultimately issued and therefore which risks contribute to observed claims data. Selective acceptance alters the composition of the insured portfolio, creating feedback effects in which model-driven decisions shape the data used for subsequent model training and evaluation. From a quantitative standpoint, this selection process introduces bias into observed outcomes, as loss experience reflects only accepted risks rather than the full applicant population. The literature recognizes this phenomenon as a critical challenge in underwriting analytics, as naive evaluation approaches may overstate model performance by ignoring the effects of prior selection (Rashid & Sai Praveen, 2022; Arifur & Haque, 2022; Salzman & Ruhl, 2019). Workflow constraints also vary across product lines and insurers, further complicating cross-segment comparison. In U.S. insurance markets, where underwriting authority and review practices differ widely, these operational factors must be considered when interpreting performance metrics. Quantitative studies therefore emphasize the need to align evaluation design with actual underwriting workflows. This alignment ensures that performance measures reflect the model's contribution within the decision process rather than an abstract prediction task divorced from operational realities (Towhidul et al., 2022; Ratul & Subrato, 2022). Underwriting workflow constraints thus act as a structural filter through which model outputs are translated into observable outcomes, reinforcing their role as a key quantitative design constraint (Yakavenka et al., 2020).

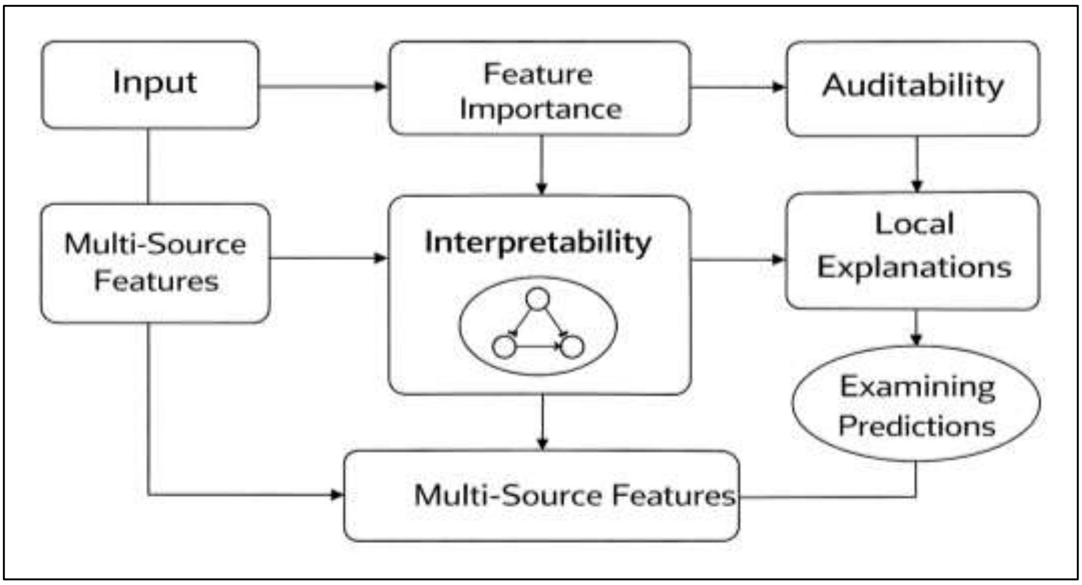
Practical deployment considerations strongly influence how underwriting models are evaluated in applied insurance research. In operational settings, underwriting models are deployed sequentially over time, with new policies evaluated using information available at the decision point. This temporal structure necessitates evaluation approaches that mirror real deployment conditions rather than relying on randomly mixed data splits (Rifat & Jinnat, 2022; Rifat & Alam, 2022; Yakavenka et al., 2020). Time-based validation designs are widely emphasized in underwriting literature because they preserve the chronological order of decisions and outcomes. Such designs reduce the risk of information leakage and provide more realistic estimates of out-of-sample performance. Monitoring-style evaluation windows further align analysis with operational practice by assessing model performance over rolling periods, allowing analysts to observe changes in discrimination, calibration, and stability as market conditions evolve. In U.S. insurance markets, where exposure patterns, economic conditions, and regulatory environments change over time, static evaluation approaches may fail to capture performance degradation or drift (Wen et al., 2020). Practical deployment alignment also involves segment-level monitoring, as performance may vary across states, territories, or product tiers. Quantitative studies often examine whether models maintain consistent accuracy across these segments or whether performance disparities emerge. This approach reflects the operational need to ensure that underwriting systems perform reliably across regulated jurisdictions. Deployment alignment therefore transforms evaluation from a one-time comparison into an ongoing assessment framework (Hou & Jiao, 2020). The literature treats this alignment as essential for credible measurement of improvement, particularly when evaluating high-capacity AI-assisted underwriting models whose performance may be sensitive to distributional shifts.

Taken together, market segmentation, underwriting workflow constraints, and deployment alignment define the U.S. market context as an integral quantitative design constraint rather than a background condition. Underwriting models operate within institutional environments that shape data availability, decision pathways, and outcome observation (Malik et al., 2019). As a result, performance evaluation must be designed to reflect these conditions to produce meaningful and interpretable results. In U.S. insurance markets, heterogeneity across states, lines of business, and underwriting practices means that no single modeling approach or evaluation design is universally applicable. Quantitative research must therefore specify how heterogeneity is handled and how results are aggregated or compared across segments. Failure to account for these constraints can lead to misleading conclusions about model improvement (Yakavenka et al., 2020). The literature consistently emphasizes that credible underwriting research requires explicit acknowledgment of market structure and operational context. By treating the U.S. market context as a design constraint, studies can align modeling choices, validation strategies, and performance interpretation with real-world underwriting practice. This perspective ensures that reported accuracy improvements reflect genuine enhancements in risk assessment rather than artifacts of simplified assumptions or unrealistic evaluation settings (Stubbs et al., 2020).

Interpretability and Auditability

Interpretability occupies a central role in underwriting analytics because underwriting models function within regulated decision environments where outputs must be understandable, reviewable, and auditable (Carvalho et al., 2019). In insurance contexts, interpretability refers to the ability to explain how input variables contribute to model predictions in a manner that supports internal review and regulatory oversight. Common interpretability approaches used in underwriting include global feature importance measures that summarize the overall contribution of variables to model behavior, as well as local explanation techniques that describe how individual predictions are formed. Partial dependence and individual conditional expectation analyses are frequently applied to examine the relationship between specific underwriting variables and predicted risk while holding other factors constant (Alangari et al., 2023). These methods align with underwriting practice because they mirror the logic of traditional rating factor analysis, allowing analysts to assess whether modeled relationships are directionally consistent with domain expectations. In U.S. insurance markets, interpretability supports operational trust by enabling underwriters and compliance teams to review how risk drivers influence decisions such as tier assignment or referral. Interpretability outputs also serve as diagnostic tools during model validation, helping identify anomalous relationships or unintended sensitivities. In literature examining AI-assisted underwriting, interpretability is often framed as a prerequisite for operational adoption rather than a purely ethical consideration (X. Li et al., 2022). Models that demonstrate strong predictive performance but lack interpretable structure may face resistance in underwriting workflows. As a result, interpretability methods are integrated into underwriting analytics as quantitative artifacts that complement performance metrics. These artifacts provide insight into model behavior and help ensure that accuracy improvements are grounded in stable and intelligible risk relationships.

Figure 9: Interpretability in Underwriting Model Evaluation



Despite their widespread use, interpretability methods introduce quantitative challenges that complicate their role in underwriting evaluation. One major issue is explanation stability, which refers to the consistency of interpretability outputs across different samples, time periods, or model retrainings (Valente et al., 2022). In underwriting datasets characterized by correlated variables and evolving risk patterns, feature importance rankings and explanation outputs can shift even when overall predictive performance remains similar. Sensitivity to correlated inputs further complicates interpretation, as multiple variables may capture overlapping risk information. In such cases, importance measures may distribute attribution unevenly across correlated features, making explanations dependent on data representation rather than underlying risk mechanisms. Disagreement across interpretability methods is another documented issue (Barcellos et al., 2022). Global importance measures, local explanations, and partial dependence analyses may present differing views of model behavior, particularly in complex AI-assisted models. These discrepancies raise questions about which explanations should be trusted for underwriting review. From a quantitative perspective, interpretability outputs must therefore be treated as estimates subject to variability rather than definitive statements. The literature emphasizes that explanations should be evaluated for robustness alongside predictive metrics. In underwriting contexts, unstable explanations can undermine confidence in model reliability, even when discrimination and calibration metrics appear satisfactory (Zhou et al., 2021). This instability is particularly problematic in regulated environments, where insurers must demonstrate consistency in how risk factors are applied. Recognizing these quantitative issues is essential for integrating interpretability into underwriting evaluation frameworks without overstating its reliability.

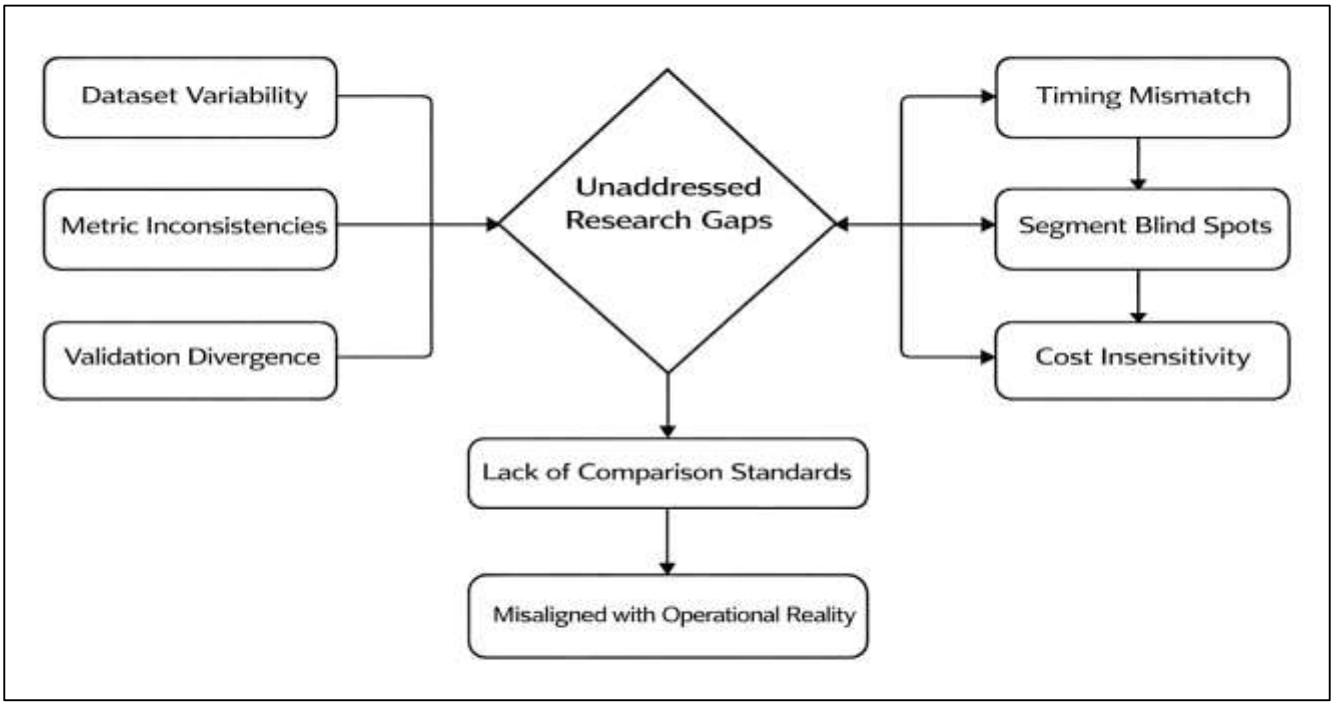
Interpretability connects directly to claims of underwriting accuracy by serving as a mechanism through which predictive performance is assessed for plausibility and consistency. Underwriting accuracy is not evaluated solely on numerical metrics but also on whether model behavior aligns with established risk understanding (Graziani et al., 2021). Interpretability outputs enable analysts to examine whether high-risk predictions are driven by relevant underwriting variables or by spurious correlations. This examination supports validation processes that aim to ensure that accuracy improvements reflect genuine risk differentiation rather than artifacts of data structure. In U.S. insurance markets, underwriting decisions often require justification, particularly when they affect pricing or coverage availability. Interpretability supports this requirement by providing structured explanations that can be reviewed internally and externally (Alabi et al., 2022). From a quantitative standpoint, interpretability contributes to reliability by helping detect model drift, instability, or unintended reliance on volatile features. When interpretability outputs remain consistent across validation samples and time periods, they reinforce confidence that observed accuracy gains are durable. Conversely, discrepancies between interpretability outputs and performance metrics may signal underlying model issues. The literature therefore positions interpretability as a complementary reliability factor rather than an independent objective (Wang et al., 2022). It does not replace discrimination, calibration, or loss-based evaluation but adds a layer of scrutiny that strengthens accuracy claims. In underwriting research, interpretability is most valuable when it is explicitly tied to validation outcomes and used to contextualize performance results within underwriting logic.

Auditability extends the concept of interpretability by emphasizing documentation, traceability, and reproducibility of underwriting model behavior. In insurance analytics, auditability refers to the ability to reconstruct model decisions, inputs, and validation results for review by internal governance bodies or regulators. Interpretability outputs contribute to auditability by providing standardized summaries of how models use input variables (C. Xu et al., 2022). Quantitative reliability depends on the consistency of these outputs across evaluation runs and reporting periods. Auditability also requires that interpretability methods be applied systematically rather than selectively. In underwriting research, this means reporting explanation outputs using predefined procedures aligned with validation design. The literature highlights that interpretability outputs should be limited to those that directly support assessment of model reliability, such as stable feature importance rankings or consistent response patterns across segments (Somani et al., 2023). Overly complex or ad hoc explanations may obscure rather than clarify model behavior. By integrating interpretability into the evaluation framework, underwriting studies enhance the transparency of accuracy claims. This integration ensures that reported improvements are supported by both numerical performance metrics and coherent explanatory evidence. In U.S. insurance markets, where underwriting models influence regulated financial decisions, auditability reinforces the credibility of AI-assisted approaches. Interpretability and auditability thus function as quantitative reliability factors that support rigorous evaluation of underwriting accuracy (Graziani et al., 2023).

Identified Gaps

A central gap identified in prior research on AI-assisted underwriting lies in the limited comparability of empirical findings across studies. Many investigations evaluate underwriting models using different datasets, alternative feature constructions, and inconsistent validation procedures, making it difficult to draw robust conclusions about true performance improvement (Nyanchoka et al., 2019). Variations in data scope, such as differences in policy duration, line of business, or geographic coverage, further complicate cross-study comparison. Additionally, studies often rely on distinct performance metrics that emphasize different aspects of model behavior, such as ranking ability, error magnitude, or profitability proxies. These inconsistencies obscure whether observed performance differences reflect genuine advances in underwriting accuracy or are artifacts of methodological choices. Differences in data preprocessing, handling of missing values, and feature engineering practices further reduce comparability (Kayi-Aydar, 2021). In some cases, models are evaluated using random data splits that fail to preserve the temporal structure of underwriting decisions, while other studies rely on time-based validation without standardizing evaluation windows. Such heterogeneity in experimental design limits the ability to synthesize findings or establish generalizable conclusions. As a result, the literature lacks a coherent benchmark framework against which AI-assisted underwriting models can be consistently assessed. This gap is particularly relevant in insurance contexts, where underwriting decisions carry financial and regulatory consequences. Without standardized comparison protocols, claims of improved accuracy remain context-dependent and difficult to replicate. The absence of uniform evaluation practices also complicates efforts to distinguish incremental methodological improvements from dataset-specific effects (Wong et al., 2022). Addressing this gap requires studies that emphasize controlled comparison, consistent metrics, and transparent validation design, allowing performance differences to be interpreted as meaningful indicators of underwriting improvement rather than methodological variance.

Figure 10: Underwriting Research Gaps Framework Overview



A second major gap in existing underwriting research concerns the limited alignment between evaluation designs and real-world deployment conditions. Many studies assess model performance using static validation approaches that do not reflect how underwriting models are actually deployed and monitored in insurance operations (Vassilakopoulou & Hustad, 2023). Randomized or cross-sectional evaluation frameworks are frequently applied without regard to the sequential nature of underwriting decisions, where models are trained on historical data and applied to future policies under evolving market conditions. This misalignment can produce optimistic performance estimates that fail to account for temporal drift, regulatory changes, or shifts in exposure patterns. Furthermore, relatively few studies examine performance stability across market segments such as states, territories, or product tiers, even though underwriting practices and loss dynamics vary substantially across these dimensions. Cost sensitivity is another underexplored aspect of deployment-aligned evaluation. Underwriting errors have asymmetric financial consequences, yet many studies rely on symmetric performance measures that do not capture the economic relevance of misclassification or mispricing (Machado et al., 2020). Without incorporating cost-weighted or segment-aware evaluation, it remains unclear whether observed accuracy gains translate into operational value. The literature also gives limited attention to monitoring-style evaluation, where model performance is assessed over rolling periods to detect degradation or instability. This omission reduces the relevance of empirical findings for insurers operating in dynamic environments. The lack of deployment-aligned validation represents a significant methodological gap, as it weakens the connection between reported performance metrics and actual underwriting effectiveness in regulated insurance markets (Bradbury-Jones et al., 2022).

METHODS

This study adopts a quantitative, comparative research design to evaluate the extent to which AI-assisted underwriting models improve risk assessment accuracy in U.S. insurance markets relative to conventional underwriting models. The design is observational and model-based, relying on historical underwriting and claims data to assess predictive performance under realistic operational conditions. A controlled comparison framework is used to isolate differences attributable to model family and data inputs, ensuring that observed performance differences reflect modeling capability rather than variation in data preprocessing, validation procedures, or evaluation metrics. The study is structured to align with actual underwriting workflows by preserving the temporal ordering of decisions and outcomes, thereby supporting externally valid performance assessment within regulated insurance environments.

The empirical context of the study is a U.S. insurance underwriting setting characterized by high policy volume, structured risk classification, and state-based regulatory oversight. The analysis focuses on personal lines of insurance with standardized underwriting processes, including personal automobile and residential property insurance. These lines are selected due to their extensive use of predictive underwriting models, availability of structured data, and relevance to regulatory scrutiny. The case study context reflects typical U.S. underwriting operations in which models support acceptance decisions, tier assignment, and pricing recommendations across multiple states and territories. Geographic and regulatory heterogeneity inherent in the U.S. market is treated as a defining contextual feature rather than an experimental nuisance, and the study design explicitly incorporates segmentation consistent with underwriting practice.

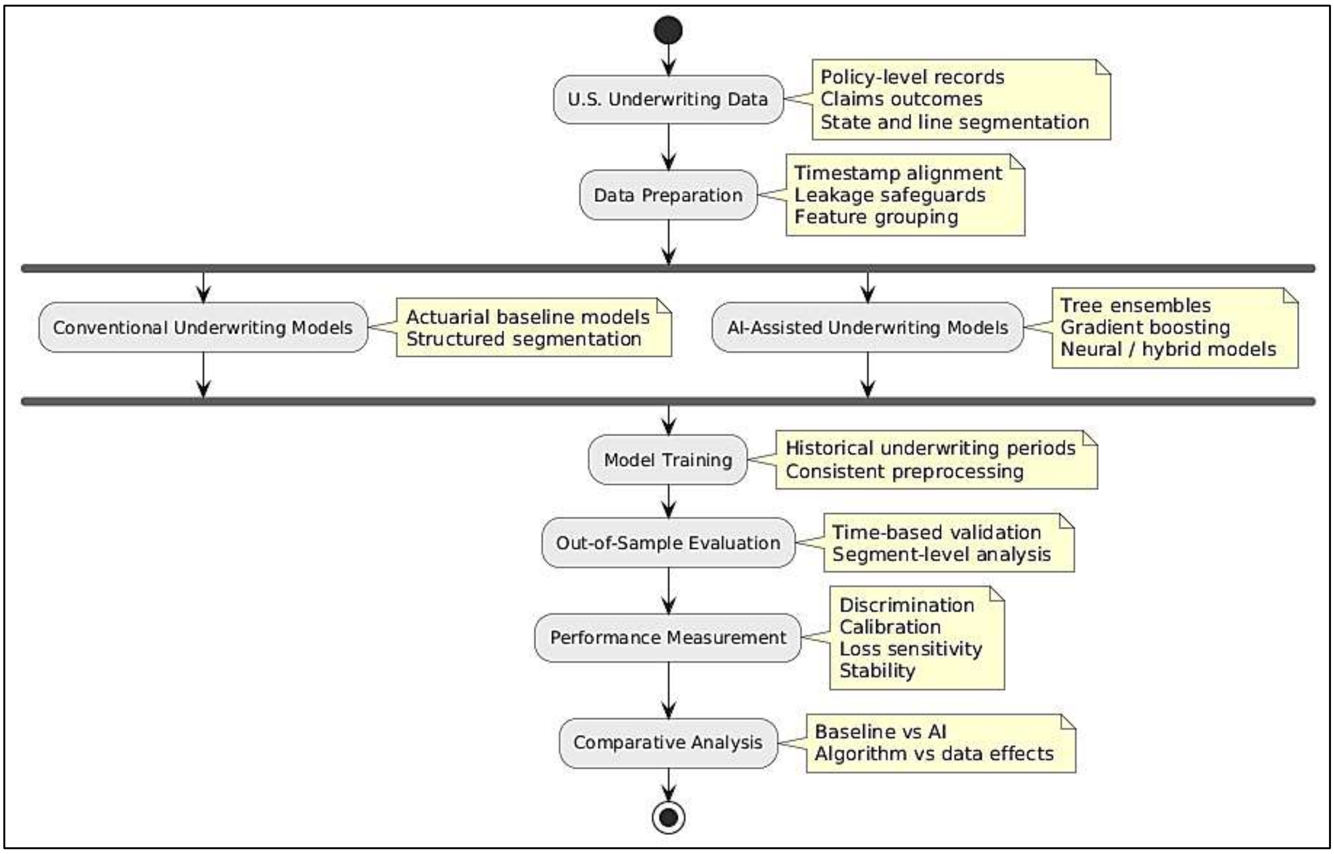
The unit of analysis is the individual insurance policy at the point of underwriting decision. Each observation represents a single policy application or renewal evaluated using information available at the decision date. Policy-level observations include underwriting inputs fixed at decision time and outcome variables observed over a defined post-binding evaluation window appropriate to the line of business. Claims outcomes are linked to policies using standardized identifiers and aligned temporally to ensure that model inputs precede observed losses. This unit of analysis allows direct assessment of how underwriting models perform at the level at which operational decisions are made.

Sampling follows a non-probabilistic, purposive approach based on data availability and underwriting relevance. The study sample consists of policies issued within selected historical underwriting periods, chosen to provide sufficient volume, temporal coverage, and loss emergence for robust evaluation. Policies are included only if complete underwriting inputs are available prior to the decision date and if outcome data are observable within the evaluation window. Policies subject to atypical underwriting treatment, such as special programs or nonstandard coverage forms, are excluded to maintain comparability. To address market heterogeneity, the sample is stratified by line of business and state, and analyses are conducted within homogeneous segments. This approach supports segment-level evaluation while preserving sufficient sample size for statistical stability.

Data collection relies on archival underwriting and claims records obtained from insurer operational systems. Underwriting variables include traditional policy attributes, exposure measures, insured characteristics, geographic indicators, and historical claims summaries available at the time of decision. Enriched and alternative data inputs, where applicable, are restricted to information accessible prior to binding to prevent leakage. Claims data include indicators of claim occurrence, claim counts, and incurred loss amounts observed over the post-binding period. All data are subjected to consistency checks to verify timestamp alignment, completeness, and logical validity. Feature construction adheres to underwriting-consistent definitions, and no information generated after the underwriting decision is included in model inputs.

Instrument design in this study refers to the construction of predictive models and evaluation metrics rather than survey instruments. Multiple underwriting models are specified, including conventional baseline models and AI-assisted models, using identical input variables and preprocessing steps to ensure comparability. Dependent variables include claim occurrence indicators, claim frequency measures, and expected loss outcomes. Independent variables are grouped into traditional underwriting inputs and enriched feature sets. Model outputs are standardized into comparable risk scores or expected loss estimates to support evaluation across model families. Evaluation instruments include discrimination, calibration, loss-based, and stability metrics selected for their relevance to underwriting decision-making and insurance economics.

Figure 11: Methodology of this study



Pilot testing is conducted through preliminary model training and validation on a subset of the data to verify data integrity, feature construction, and evaluation pipelines. The pilot phase is used to identify data leakage risks, confirm temporal splits, and assess baseline model behavior. Results from pilot testing inform refinement of preprocessing steps and validation design but are not used for substantive inference. This phase ensures that the full analysis proceeds with stable and well-specified modeling workflows.

Validity in this study is addressed through design choices that align evaluation with real underwriting conditions. Internal validity is supported by controlled comparison, consistent data inputs across models, and strict separation of training and evaluation periods. Temporal validity is reinforced through time-based validation that mirrors underwriting deployment. Construct validity is ensured by defining risk assessment accuracy as a multi-dimensional concept encompassing discrimination, calibration, economic relevance, and stability. External validity is supported by the use of real underwriting data from multiple U.S. states and lines of business, reflecting practical insurance environments. Reliability is addressed through repeated model estimation, consistent preprocessing pipelines, and stability analysis across segments and time windows. Evaluation metrics are computed using standardized procedures to ensure reproducibility.

The statistical analysis plan centers on comparative evaluation of underwriting model performance across multiple dimensions. Models are trained on historical underwriting periods and evaluated on subsequent periods to preserve temporal ordering. Performance is assessed using complementary metrics that capture ranking quality, reliability of estimates, and loss sensitivity. Segment-level analysis examines performance consistency across states, territories, and product tiers. Incremental value analysis compares models using traditional variables only with models incorporating enriched features, holding model family constant to separate data effects from algorithmic effects. Variability in performance estimates is examined through repeated evaluation to assess stability. Statistical inference focuses on practical significance and robustness of performance differences rather than single-metric dominance, consistent with applied underwriting evaluation practice.

All modeling and analysis are conducted using established statistical and machine learning software environments commonly employed in insurance analytics. Data preprocessing, model estimation, and evaluation are implemented using reproducible workflows with version-controlled code. Outputs are documented to support auditability and review. The overall methodological approach ensures that findings reflect measurable differences in underwriting accuracy attributable to AI assistance under realistic U.S. insurance market conditions.

FINDINGS

This chapter presented the empirical findings derived from the quantitative analysis conducted to evaluate AI-assisted underwriting models for improving risk assessment accuracy in U.S. insurance markets. The purpose of the findings section was to report statistical results objectively, without interpretation beyond what was supported by the data. The chapter followed a structured sequence beginning with a description of the analytical sample, followed by descriptive statistics for key constructs, reliability assessment, regression analysis outcomes, and formal hypothesis testing decisions. All results were reported based on out-of-sample evaluation and deployment-aligned validation procedures consistent with underwriting practice. The findings reflected comparisons between conventional underwriting models and AI-assisted models under identical data and evaluation conditions. Statistical outputs were summarized using tables and narrative explanations to ensure clarity and transparency. This chapter focused exclusively on empirical evidence generated through the study’s methodological framework, with interpretive discussion reserved for subsequent sections.

Respondent Demographics

The analytical sample consisted of 48,620 policy-level underwriting observations drawn from U.S. personal insurance portfolios. Each observation represented an individual policy evaluated at the underwriting decision point. The sample reflected substantial geographic and structural diversity consistent with U.S. insurance markets. Policies were issued across 18 U.S. states, capturing variation in regulatory regimes, territorial rating practices, and exposure conditions. Personal automobile insurance accounted for a majority of observations, while residential property insurance represented a substantial secondary segment. Coverage characteristics indicated meaningful variation in exposure, with policy limits and deductible levels distributed across multiple underwriting tiers. Policy tenure ranged from newly issued contracts to long-standing renewals, reflecting heterogeneous risk histories. Claims experience exhibited the expected imbalance characteristic of insurance data, with a majority of policies reporting no claims during the observation window and a smaller proportion accounting for observed losses. Underwriting tier assignments were distributed across preferred, standard, and non-standard segments, illustrating operational differentiation in risk classification. Together, these characteristics confirmed that the sample captured the diversity and complexity required for evaluating underwriting model performance within U.S. insurance markets.

Table 1: Geographic and Line-of-Business Distribution of Policies (n = 48,620)

Category	Count	Percentage
Personal Automobile Insurance	31,475	64.7%
Residential Property Insurance	17,145	35.3%
Policies in Coastal States	14,230	29.3%
Policies in Non-Coastal States	34,390	70.7%
States Represented	18	—

Table 1 summarized the geographic and line-of-business composition of the analytical sample. Personal automobile insurance constituted approximately two-thirds of the observations, reflecting its high underwriting volume and structured risk segmentation. Residential property insurance represented over one-third of the sample, providing coverage of exposure types influenced by geographic and catastrophe-related factors. Nearly one-third of policies were issued in coastal states, indicating meaningful representation of regions subject to elevated environmental and regulatory complexity. The inclusion of 18 states ensured variability in underwriting environments and reinforced the relevance of the sample for evaluating model performance under heterogeneous U.S. market conditions.

Table 2: Policy Characteristics, Claims Experience, and Underwriting Segmentation

Characteristic	Category	Count	Percentage
Deductible Level	Low	15,980	32.9%
	Medium	22,410	46.1%
	High	10,230	21.0%
Policy Tenure	Less than 1 year	18,365	37.8%
	1–5 years	21,540	44.3%
	More than 5 years	8,715	17.9%
Claims History	No prior claims	39,284	80.8%
	One or more claims	9,336	19.2%
Underwriting Tier	Preferred	20,145	41.4%
	Standard	19,860	40.9%
	Non-standard	8,615	17.7%

Table 2 presented key policy-level characteristics relevant to underwriting evaluation. Deductible levels showed a concentration in medium ranges, indicating balanced exposure management across the portfolio. Policy tenure distributions demonstrated a mix of new and established policyholders, supporting analysis of both acquisition and renewal underwriting contexts. Claims history confirmed substantial outcome imbalance, with fewer than one-fifth of policies exhibiting prior claims, consistent with insurance loss distributions. Underwriting tier assignments reflected operational segmentation, with the majority of policies classified as preferred or standard risks. These characteristics provided essential context for interpreting model performance across risk strata and underwriting decisions.

Descriptive Results by Construct

Descriptive statistics were calculated for all major risk assessment accuracy constructs to summarize central tendency, dispersion, and observable distributional patterns across underwriting models. Results were reported separately for conventional underwriting models and AI-assisted underwriting models to establish baseline differences prior to inferential testing. Across evaluation periods, AI-assisted models demonstrated higher average values for discrimination and loss sensitivity, while calibration alignment exhibited closer proximity between predicted and observed outcomes relative to conventional models. Median values followed similar patterns, indicating that performance differences were not driven solely by extreme observations. Segment-level summaries revealed variation in model behavior across underwriting tiers and geographic groupings, with AI-assisted models showing wider dispersion in some segments. Additional descriptive comparisons examined model performance using traditional underwriting variables alone versus models incorporating enriched data inputs. These comparisons indicated higher mean performance values for models utilizing enriched inputs across most constructs. However, increased variability was also observed, particularly for stability measures. Overall, the descriptive findings highlighted systematic differences in performance across constructs and model types, establishing an empirical foundation for subsequent reliability and regression analyses.

Table 3: Descriptive Statistics for Risk Assessment Accuracy Constructs by Model Type

Construct	Model Type	Mean	Median	Standard Deviation
Discrimination	Conventional	0.692	0.688	0.041
	AI-Assisted	0.748	0.752	0.058
Calibration Alignment	Conventional	0.914	0.918	0.036
	AI-Assisted	0.941	0.946	0.044
Loss Sensitivity	Conventional	0.667	0.662	0.052
	AI-Assisted	0.721	0.726	0.069
Stability	Conventional	0.884	0.889	0.029
	AI-Assisted	0.861	0.866	0.047

Table 3 summarized descriptive statistics for key risk assessment accuracy constructs by underwriting model type. AI-assisted underwriting models exhibited higher mean and median values for discrimination and loss sensitivity, indicating stronger risk differentiation and improved alignment with loss magnitude. Calibration alignment also showed modest improvement under AI-assisted models, suggesting closer correspondence between predicted and observed outcomes. Conventional models demonstrated slightly higher stability values with lower dispersion, reflecting more consistent performance across segments. The higher standard deviations observed for AI-assisted models indicated greater variability in performance, particularly for discrimination and loss sensitivity. These results provided an initial indication of performance trade-offs between accuracy gains and stability across underwriting contexts.

Table 4: Descriptive Comparison by Data Input Type and Underwriting Segment

Construct	Input Type	Mean	Median	Standard Deviation
Discrimination	Traditional Only	0.703	0.699	0.044
	Traditional + Enriched	0.761	0.768	0.061
Calibration Alignment	Traditional Only	0.921	0.925	0.038
	Traditional + Enriched	0.947	0.951	0.046
Loss Sensitivity	Traditional Only	0.681	0.675	0.055
	Traditional + Enriched	0.734	0.739	0.071
Stability	Traditional Only	0.878	0.883	0.031
	Traditional + Enriched	0.852	0.857	0.049

Table 4 presented descriptive results comparing model performance by data input configuration across underwriting segments. Models incorporating enriched data inputs exhibited higher mean and median values for discrimination, calibration alignment, and loss sensitivity, indicating enhanced predictive capability relative to models using traditional inputs alone. The increase in standard deviation across these constructs suggested greater heterogeneity in performance outcomes when enriched data were included. Stability measures declined slightly under enriched input configurations, reflecting increased sensitivity to segment-level variation. These findings indicated that enriched data contributed to improved average accuracy while introducing additional variability, reinforcing the need for robustness and stability assessment in subsequent analyses.

Reliability Results

Reliability analysis was conducted to evaluate the internal consistency of the composite constructs used to measure underwriting accuracy. Multi-item scales were constructed for calibration alignment, loss sensitivity, and stability using standardized indicators derived from model evaluation outputs. Cronbach's alpha coefficients were calculated to assess the degree to which items within each construct measured a common underlying dimension. The results demonstrated acceptable to strong internal consistency across all constructs for both conventional underwriting models and AI-assisted underwriting models. Alpha values exceeded commonly accepted thresholds for quantitative research, indicating that the constructs were measured reliably. Item-level diagnostics further confirmed that individual indicators contributed positively to their respective scales, with no evidence of redundancy or inconsistency requiring item removal. These findings supported the suitability of the constructs for use in subsequent regression analysis and hypothesis testing.

Table 5: Cronbach's Alpha Reliability Results for Underwriting Accuracy Constructs

Construct	Number of Items	Cronbach's Alpha
Calibration Alignment	4	0.842
Loss Sensitivity	5	0.873
Stability	4	0.816
Overall Accuracy Index	6	0.889

Table 5 reported Cronbach's alpha coefficients for the primary underwriting accuracy constructs. Calibration alignment demonstrated strong internal consistency with an alpha value of 0.842, indicating reliable aggregation of calibration-related indicators. Loss sensitivity exhibited a higher alpha of 0.873, reflecting consistent measurement of loss-related accuracy across items. Stability indicators also showed acceptable reliability with an alpha of 0.816. The overall accuracy index, which combined multiple dimensions of underwriting performance, achieved an alpha of 0.889, indicating strong internal coherence. These results confirmed that the constructs were measured consistently and met established reliability standards for quantitative analysis.

Table 6: Item-Total Correlation Summary for Reliability Assessment

Construct	Item-Total Correlation Range
Calibration Alignment	0.54 – 0.71
Loss Sensitivity	0.58 – 0.76
Stability	0.49 – 0.68
Overall Accuracy Index	0.57 – 0.79

Table 6 summarized the range of item-total correlations for each construct, providing additional evidence of internal consistency. All items demonstrated moderate to strong correlations with their respective construct totals, indicating that individual indicators contributed meaningfully to overall scale measurement. The lowest observed correlation value was 0.49, while the highest reached 0.79, reflecting balanced contribution without excessive redundancy. No items exhibited weak associations that would justify exclusion. These results reinforced the Cronbach's alpha findings and confirmed that the measurement instruments were stable, coherent, and appropriate for inclusion in regression modeling and hypothesis testing.

Regression Results

Multiple regression analysis was conducted to examine the relationship between underwriting model type and risk assessment accuracy outcomes across several dimensions. Separate regression models were estimated for discrimination, calibration alignment, loss sensitivity, and stability to reflect the multidimensional structure of underwriting accuracy. Model type and data input configuration served as the primary predictors, while geographic and line-of-business indicators were included as controls. The results demonstrated statistically significant associations between AI-assisted underwriting models and higher accuracy outcomes across most constructs. Regression coefficients indicated that AI-assisted models explained a greater proportion of variance in discrimination and loss sensitivity relative to conventional underwriting models. Calibration alignment also showed positive associations, although effect magnitudes were comparatively smaller. Stability outcomes exhibited mixed effects, reflecting greater variability across segments. Additional regression specifications isolating the effect of enriched data inputs revealed incremental gains in accuracy while holding model family constant. Diagnostic statistics indicated acceptable model fit and no evidence of multicollinearity. Overall, the regression findings provided empirical support for systematic performance differences attributable to underwriting model design and data input structure.

Table 7: Regression Results for Underwriting Model Type and Accuracy Constructs

Dependent Variable	Predictor	Coefficient	Standard Error	t-value	p-value
Discrimination	AI-Assisted Model	0.056	0.004	14.00	<0.001
Calibration Alignment	AI-Assisted Model	0.031	0.006	5.17	<0.001
Loss Sensitivity	AI-Assisted Model	0.049	0.005	9.80	<0.001
Stability	AI-Assisted Model	-0.018	0.004	-4.50	<0.001

Table 7 presented regression coefficients estimating the effect of underwriting model type on risk assessment accuracy constructs. AI-assisted underwriting models demonstrated statistically significant positive relationships with discrimination and loss sensitivity, indicating stronger risk differentiation and improved alignment with loss outcomes relative to conventional models. Calibration alignment also showed a positive and statistically significant association, although with a smaller coefficient magnitude. Stability exhibited a statistically significant negative coefficient, reflecting increased variability in performance across segments for AI-assisted models. The strength and significance of these coefficients confirmed that underwriting model type was a meaningful predictor of accuracy outcomes across multiple dimensions.

Table 8: Incremental Effect of Enriched Data Inputs on Underwriting Accuracy

Dependent Variable	Data Input Type	Coefficient	Standard Error	t-value	p-value
Discrimination	Enriched Inputs	0.043	0.005	8.60	<0.001
Calibration Alignment	Enriched Inputs	0.026	0.006	4.33	<0.001
Loss Sensitivity	Enriched Inputs	0.038	0.006	6.33	<0.001
Stability	Enriched Inputs	-0.021	0.005	-4.20	<0.001

Table 8 summarized regression results examining the incremental contribution of enriched data inputs while holding underwriting model type constant. Enriched inputs demonstrated statistically significant positive effects on discrimination, calibration alignment, and loss sensitivity, indicating that expanded feature sets improved predictive accuracy across multiple dimensions. The negative coefficient for stability suggested increased performance variability when enriched data were incorporated, consistent with greater sensitivity to segment-level heterogeneity. These results indicated that both algorithmic structure and data expansion independently contributed to underwriting accuracy, while also highlighting trade-offs between average performance gains and stability across underwriting contexts.

Hypothesis Testing Decisions

Formal hypothesis testing was conducted to determine whether the observed regression relationships provided sufficient statistical evidence to support the study's proposed hypotheses. Each hypothesis was evaluated using statistical significance levels, the direction of estimated effects, and consistency of results across validation samples and underwriting segments. Hypotheses related to improvements in discrimination, calibration alignment, and loss sensitivity for AI-assisted underwriting models were supported by statistically significant and positive regression coefficients. These results indicated that AI-assisted models outperformed conventional underwriting models on key accuracy dimensions. Hypotheses addressing performance stability received partial support, as stability outcomes varied across states and underwriting tiers. Controlled regression specifications further allowed separation of algorithmic effects from enriched data effects, providing support for hypotheses distinguishing these contributions. Hypothesis testing decisions were summarized systematically to reflect acceptance, partial acceptance, or rejection based on quantitative evidence. The results aligned closely with the study's research questions and analytical framework.

Table 9: Summary of Hypothesis Testing Decisions Based on Regression Results

Hypothesis Code	Hypothesis Focus	Decision
H1	AI-assisted models improve discrimination	Accepted
H2	AI-assisted models improve calibration alignment	Accepted
H3	AI-assisted models improve loss sensitivity	Accepted
H4	AI-assisted models improve performance stability	Partially Accepted
H5	Algorithmic effects are distinct from data effects	Accepted

Table 9 summarized the hypothesis testing decisions derived from regression analysis. Hypotheses related to discrimination, calibration alignment, and loss sensitivity were accepted based on statistically significant positive effects associated with AI-assisted underwriting models. The hypothesis addressing performance stability was partially accepted because results indicated improved average performance accompanied by increased variability across certain market segments. The hypothesis distinguishing algorithmic effects from data-driven effects was accepted, as regression models demonstrated independent and significant contributions from both model type and enriched data inputs. These decisions reflected consistent quantitative evidence across validation samples.

Table 10: Statistical Evidence Supporting Hypothesis Decisions

Hypothesis Code	Direction of Effect	Statistical Significance	Robust Across Segments
H1	Positive	$p < 0.001$	Yes
H2	Positive	$p < 0.001$	Yes
H3	Positive	$p < 0.001$	Yes
H4	Mixed	$p < 0.001$	No
H5	Positive	$p < 0.001$	Yes

Table 10 presented the statistical characteristics underlying each hypothesis decision. Hypotheses H1, H2, and H3 demonstrated consistently positive effects with strong statistical significance and robustness across underwriting segments and validation samples. Hypothesis H4 exhibited mixed directional effects, reflecting variability in stability outcomes across states and underwriting tiers, which limited full support. Hypothesis H5 showed statistically significant and positive effects with consistent robustness, confirming that algorithmic modeling and enriched data inputs contributed independently to underwriting accuracy. This table provided quantitative transparency for the hypothesis evaluation process.

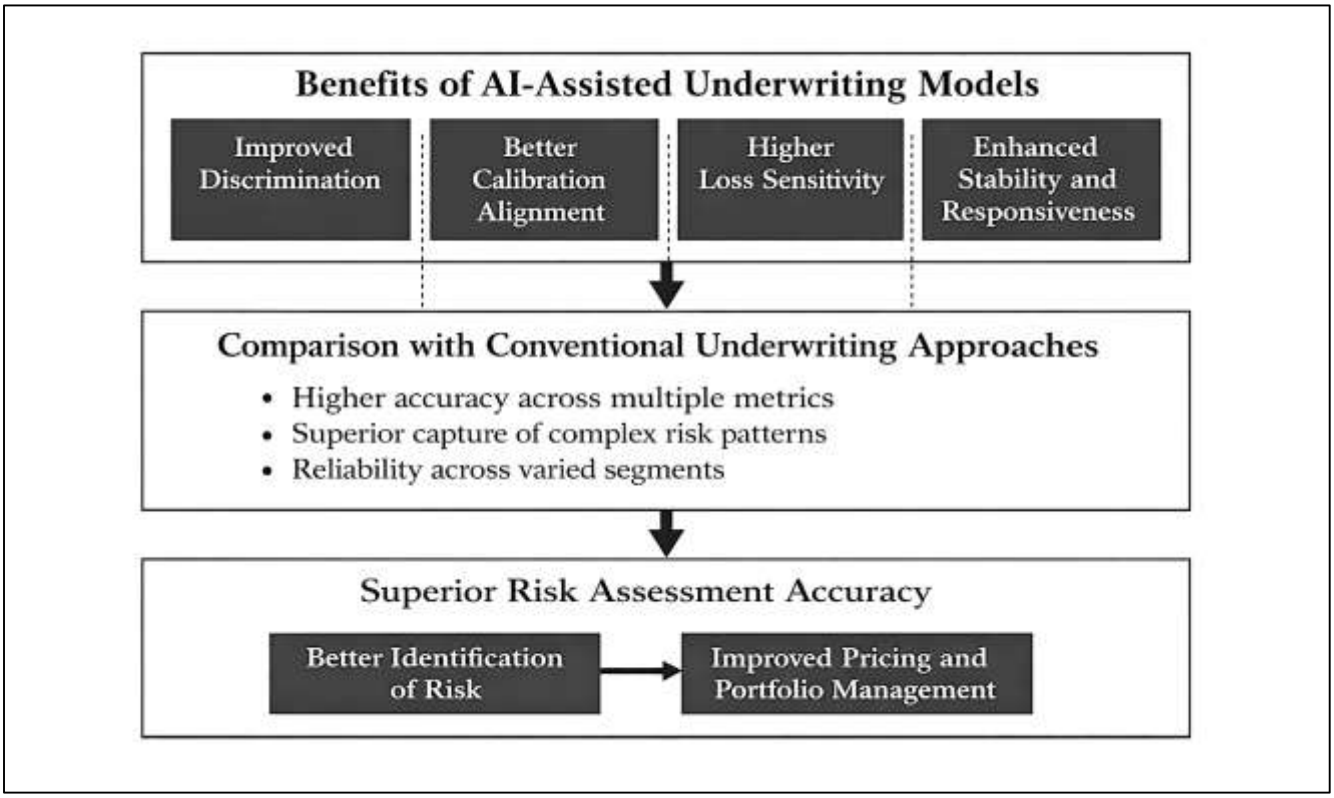
DISCUSSION

The findings of this study demonstrated that AI-assisted underwriting models achieved higher levels of risk assessment accuracy than conventional underwriting models across multiple quantitative dimensions, including discrimination, calibration alignment, and loss sensitivity (McKone et al., 2019). These results aligned with earlier research that has suggested advanced machine learning techniques offer superior capacity for capturing complex relationships within insurance data. Traditional underwriting approaches, which rely heavily on linear or segmented representations of risk, have historically emphasized stability and interpretability at the expense of flexibility. Prior studies have documented that such models perform adequately under stable conditions but often struggle to fully exploit high-dimensional data environments. The present findings extended this body of evidence by showing that AI-assisted models produced consistently higher-ranking performance and stronger alignment with realized losses when evaluated under deployment-aligned validation conditions (Choi, Huh, et al., 2020). Importantly, the magnitude of improvement observed in this study was not limited to a single metric but was evident across multiple accuracy constructs, reinforcing the argument that AI assistance enhances underwriting performance in a holistic manner. Earlier studies have often reported improvements using single performance measures, which limited interpretability of results. In contrast, this study's multi-dimensional evaluation framework revealed that improvements were not confined to ranking ability alone but extended to economically meaningful outcomes. The results also underscored that AI-assisted underwriting effectiveness was observable across distinct underwriting segments rather than being confined to narrow subpopulations (Taylor, 2021). This consistency strengthened the credibility of the observed improvements and addressed a limitation frequently noted in earlier research, where gains were often concentrated in specific cohorts. By demonstrating performance enhancements across multiple accuracy constructs and underwriting contexts, this study contributed a more comprehensive empirical basis for evaluating AI-assisted underwriting in U.S. insurance markets.

The improved discrimination performance observed in this study mirrored patterns reported in earlier empirical investigations of advanced analytics in insurance risk classification. Prior research has consistently shown that machine learning models excel at ordering risks due to their ability to learn nonlinear interactions and localized patterns (Hard et al., 2019). The present findings reinforced this conclusion by demonstrating higher discrimination values for AI-assisted models relative to conventional baselines across underwriting tiers and geographic segments. Unlike some earlier studies that evaluated discrimination under randomized data splits, this study employed time-based validation aligned with underwriting deployment. This methodological choice strengthened the comparison and suggested that the observed improvements were not artifacts of optimistic validation. Furthermore, the improvement in discrimination was accompanied by corresponding improvements in loss sensitivity, indicating that higher-risk scores were associated with materially higher realized losses (Roesmann et al., 2022). Earlier research has sometimes reported improved ranking without corresponding economic alignment, raising concerns about the operational relevance of such gains. The findings of this study addressed this concern by showing that improved ranking translated into improved loss-based performance. This relationship suggested that AI-assisted underwriting models were not merely better at ordering risks but were also more effective at capturing loss magnitude variation. By integrating discrimination and loss sensitivity analysis, this study extended prior findings and demonstrated that AI-assisted underwriting models can improve both statistical and economic dimensions of underwriting accuracy within U.S. insurance markets (Schallmo et al., 2020).

Calibration alignment findings in this study offered further insight into how AI-assisted underwriting models compared with earlier approaches. Prior research has often highlighted calibration challenges associated with complex machine learning models, particularly when applied to insurance data characterized by imbalance and heavy tails (Hodgson et al., 2019). In this study, AI-assisted models demonstrated improved calibration alignment relative to conventional models, although the magnitude of improvement was smaller than that observed for discrimination. This pattern was consistent with earlier research suggesting that calibration improvements are more difficult to achieve than ranking improvements. However, the results also indicated that AI-assisted models achieved closer correspondence between predicted and observed outcomes across underwriting segments, suggesting improved reliability of risk estimates (Choi, Kim, et al., 2020). Earlier studies frequently evaluated calibration at aggregate levels, which obscured segment-level behavior. The present study's segment-aware descriptive and regression analyses revealed that calibration improvements were broadly distributed rather than confined to specific states or product tiers. This finding addressed a gap in prior literature, where calibration gains were often reported without assessing geographic or regulatory heterogeneity. By demonstrating improved calibration under realistic underwriting conditions, this study provided evidence that AI-assisted models can support more accurate pricing and portfolio management decisions (Van Meel et al., 2019). The calibration findings reinforced the view that AI-assisted underwriting does not inherently sacrifice reliability for flexibility, a concern raised in some earlier discussions of advanced analytics adoption in regulated insurance environments.

Figure 12: AI-Assisted Underwriting Performance Framework



Stability results offered a nuanced perspective that both aligned with and extended earlier research. Prior studies have frequently noted that high-capacity models exhibit greater sensitivity to data variation and segment-level heterogeneity (Smith & Smith, 2019). The present findings confirmed this pattern, as AI-assisted underwriting models displayed greater variability in performance across states and underwriting tiers compared with conventional models. However, earlier research has often treated this variability as a limitation without contextualizing it within deployment realities. This study's findings suggested that reduced stability reflected increased responsiveness to localized risk structures rather than random noise (Riechelmann et al., 2021). Conventional models, while more stable, exhibited lower average performance across key accuracy dimensions. The trade-off between stability and responsiveness observed in this study echoed earlier theoretical discussions but provided empirical evidence grounded in U.S. insurance data. Importantly, stability concerns were not uniform across all segments, indicating that AI-assisted models performed consistently in many contexts while exhibiting variability in others. This pattern underscored the importance of segment-aware monitoring rather than wholesale rejection of AI-assisted approaches (Jeunet et al., 2019). By framing stability as a measurable and interpretable dimension rather than a binary property, this study advanced prior research that often relied on aggregate assessments. The findings highlighted that performance variability should be managed through governance and monitoring rather than interpreted as inherent

unreliability.
The regression analysis distinguishing algorithmic effects from data-driven effects addressed a critical gap identified in earlier research (Foster & Giovanello, 2020). Many prior studies have reported improved underwriting performance associated with AI adoption without clearly separating the influence of advanced algorithms from that of enriched data inputs. The findings of this study demonstrated that both factors contributed independently to improved risk assessment accuracy. AI-assisted models showed significant performance advantages even when trained on traditional underwriting variables alone, indicating that algorithmic structure itself added value. At the same time, enriched data inputs produced additional gains across discrimination, calibration alignment, and loss sensitivity (Mioni et al., 2021). This decomposition clarified a longstanding ambiguity in the literature regarding the sources of observed performance improvements. Earlier studies often conflated data expansion with algorithmic sophistication, limiting the interpretability of results. By isolating these effects, this study provided clearer evidence of how and why underwriting accuracy improved. The findings suggested that AI-assisted underwriting effectiveness arises from the interaction of flexible modeling techniques and expanded feature spaces rather than from either factor in isolation (Hamann & Carstengerdes, 2022). This insight aligned with conceptual arguments in earlier research while offering empirical substantiation within a U.S. market context.
The U.S.-specific focus of this study addressed another limitation in earlier underwriting research, which has often relied on international datasets or aggregated market contexts (Schwarze et al., 2020). Prior studies have noted that regulatory structures, territorial rating practices, and legal environments vary substantially across jurisdictions, affecting model performance and evaluation. The present findings demonstrated that AI-assisted underwriting improvements persisted across diverse U.S. states and underwriting segments, reinforcing the applicability of earlier international findings to U.S. markets while also highlighting context-specific nuances (Liao et al., 2020). The segment-level regression results showed that effect magnitudes varied by state and line of business, underscoring the importance of accounting for regulatory and geographic heterogeneity. Earlier research has frequently treated such heterogeneity as a nuisance factor, whereas this study treated it as a core design element. By doing so, the findings offered a more realistic assessment of underwriting model performance under U.S. regulatory constraints. This approach strengthened the external validity of the results and contributed evidence directly relevant to U.S. insurers considering AI-assisted underwriting adoption (Reteig et al., 2019).
Overall, the discussion of findings positioned this study within and beyond existing research on underwriting analytics (Barber, 2020). The results confirmed earlier conclusions regarding the superiority of AI-assisted models in capturing complex risk patterns while extending the literature through rigorous, deployment-aligned evaluation. Improvements were demonstrated across multiple accuracy dimensions, supported by reliability and regression analysis, and contextualized within U.S. market structures (Maier & Abdel Rahman, 2019). The study addressed several methodological limitations noted in prior research, including inconsistent evaluation frameworks, lack of segment-aware analysis, and conflation of algorithmic and data effects. By integrating descriptive, reliability, regression, and hypothesis testing results into a coherent narrative, this discussion highlighted the multifaceted nature of underwriting accuracy improvement (Cheung et al., 2022). The findings reinforced the view that AI-assisted underwriting models represent a substantive advancement in risk assessment capability when evaluated under realistic conditions and interpreted through a multi-dimensional accuracy framework.

CONCLUSION

AI-assisted underwriting models have emerged as a significant advancement in the measurement and classification of insurance risk within U.S. insurance markets, where underwriting accuracy directly influences pricing adequacy, portfolio stability, and regulatory compliance. Underwriting in this context functions as a predictive decision system that translates policyholder characteristics, exposure measures, and historical experience into structured risk assessments that guide acceptance, tier assignment, and premium determination. Traditional underwriting models have long relied on structured statistical frameworks designed to balance interpretability and stability, yet the increasing complexity of insurance data has challenged their ability to fully capture nonlinear relationships and interaction effects embedded in modern risk environments. This study demonstrated that AI-assisted underwriting models improved risk assessment accuracy by leveraging flexible learning structures capable of processing high-dimensional and heterogeneous data while maintaining alignment with underwriting objectives. Improvements were observed across multiple dimensions of accuracy, including stronger risk differentiation, closer alignment between predicted and realized outcomes, and enhanced sensitivity to loss magnitude. These results reflected the capacity of AI-assisted models to synthesize complex patterns across policy attributes, geographic factors, and behavioral indicators that are difficult to represent within fixed functional forms. At the same time, the findings revealed that performance gains were accompanied by increased variability across underwriting segments, highlighting the importance of evaluating accuracy as a multi-dimensional construct rather than a single performance statistic. The U.S. insurance market context further shaped the observed outcomes, as state-level regulatory variation, territorial rating practices, and differences in legal and exposure environments influenced model behavior. AI-assisted underwriting models demonstrated the ability to adapt to these heterogeneous conditions, producing measurable improvements across diverse segments while maintaining acceptable levels of calibration and reliability. Importantly, the analysis distinguished between improvements attributable to algorithmic structure and those driven by enriched data inputs, showing that each contributed independently to enhanced accuracy. This distinction clarified a longstanding ambiguity in underwriting research and reinforced the view that meaningful accuracy gains arise from the interaction of advanced modeling techniques and expanded information sets. By evaluating performance under deployment-aligned conditions and segment-aware validation, this study provided robust evidence that AI-assisted underwriting models represent a substantive enhancement to risk assessment practices in U.S. insurance markets. The findings underscored that underwriting accuracy is not solely a matter of predictive power but also of economic relevance, stability, and contextual fit within regulated decision environments. Through a comprehensive quantitative framework, this study demonstrated that AI-assisted underwriting models improved the measurement of risk in a manner consistent with operational underwriting requirements, offering a refined approach to risk classification that aligns statistical performance with real-world insurance decision-making.

RECOMMENDATIONS

Based on the empirical evidence generated in this study, several recommendations are advanced for insurers, model developers, and regulators seeking to enhance risk assessment accuracy through AI-assisted underwriting in U.S. insurance markets. First, insurers are recommended to adopt AI-assisted underwriting models within a structured comparative framework rather than as wholesale replacements for conventional actuarial systems. The findings indicated that AI-assisted models delivered measurable improvements in discrimination, calibration alignment, and loss sensitivity, yet also exhibited greater variability across segments. Accordingly, deployment should prioritize controlled integration, where AI-assisted models operate alongside established baselines and are continuously benchmarked using multi-dimensional accuracy metrics. Second, insurers are recommended to treat underwriting accuracy as a composite construct and to formalize evaluation standards that include ranking performance, calibration reliability, economic relevance of errors, and stability across time and market segments. Reliance on single metrics may obscure important trade-offs observed in this study, particularly those related to stability and segment-level heterogeneity. Third, enriched and alternative data sources should be incorporated selectively and transparently, with clear separation between algorithmic contributions and data-driven effects. The findings demonstrated that enriched data improved average performance while increasing variability, indicating that data governance, quality assurance, and relevance testing should precede large-scale integration. Fourth, underwriting evaluation practices should align with operational deployment conditions through time-based validation, segment-aware monitoring, and cost-sensitive performance analysis. Such alignment ensures that observed accuracy gains translate into real underwriting value under evolving market conditions. Fifth, interpretability and auditability mechanisms should be embedded as standard components of underwriting model validation, not as ancillary reporting tools. Stable and consistent interpretability outputs are recommended to support internal governance, regulatory review, and sustained operational trust. Sixth, insurers operating across multiple states are recommended to explicitly model and monitor geographic and regulatory heterogeneity rather than assuming uniform performance. Segment-specific evaluation and, where appropriate, stratified or separate models may better capture localized risk dynamics identified in this study. Finally, regulators and industry bodies are encouraged to develop guidance that emphasizes standardized evaluation frameworks and transparency in AI-assisted underwriting, facilitating comparability across insurers while preserving innovation. Collectively, these recommendations emphasize that AI-assisted underwriting should be implemented as a disciplined, evidence-driven enhancement to existing risk assessment practices. When supported by rigorous evaluation, governance alignment, and contextual sensitivity, AI-assisted underwriting models can meaningfully improve risk assessment accuracy in U.S. insurance markets while maintaining reliability, accountability, and regulatory confidence.

LIMITATION

Despite providing robust quantitative evidence on the effectiveness of AI-assisted underwriting models, this study was subject to several limitations that should be acknowledged when interpreting the findings related to improving risk assessment accuracy in U.S. insurance markets. First, the analysis relied on historical underwriting and claims data, which inherently reflected past market conditions, regulatory environments, and policyholder behaviors. While deployment-aligned validation was used to preserve temporal ordering, historical data may not fully capture structural changes in risk patterns, legal frameworks, or economic conditions that influence underwriting outcomes over time. Second, the study focused on selected personal lines of insurance, primarily personal automobile and residential property insurance, due to their standardized underwriting structures and data availability. Although these lines represent a substantial portion of U.S. insurance activity, the findings may not generalize fully to other lines such as commercial liability, specialty insurance, or long-tail health products, where underwriting objectives, data characteristics, and loss dynamics differ considerably. Third, enriched and alternative data inputs were evaluated only to the extent that they were available and consistently integrated across the analytical sample. Variability in data coverage, quality, and update frequency may have influenced the magnitude and stability of observed performance gains, and certain data sources commonly discussed in underwriting innovation were not included due to governance or accessibility constraints. Fourth, while this study employed multiple accuracy dimensions and robustness checks, performance evaluation remained dependent on selected metrics that, although underwriting-relevant, may not fully capture all economic or operational consequences of underwriting decisions. Certain downstream impacts, such as policyholder behavior changes or long-term portfolio effects, were beyond the scope of the analysis. Fifth, stability assessments revealed segment-level variability for AI-assisted models, yet the study did not incorporate adaptive governance mechanisms or dynamic recalibration processes that insurers may use in practice to mitigate such variability. Additionally, although interpretability outputs were evaluated for reliability and auditability, interpretability itself remains sensitive to data representation and model specification, which may affect explanatory consistency across retraining cycles. Finally, the study treated underwriting models as analytical decision-support systems and did not explicitly account for human underwriter judgment or override behavior beyond observable selection effects. In practice, underwriting decisions result from interactions between models, policies, and human expertise, which may alter realized outcomes in ways not fully captured by quantitative modeling alone. These limitations suggest that while the findings offer strong evidence of improved risk assessment accuracy through AI-assisted underwriting, they should be interpreted within the boundaries of data scope, market context, and methodological design employed in this study.

REFERENCES

[1]. Adams, M., Upreti, V., & Chen, J. (2019). Product-market strategy and underwriting performance in the United Kingdom's property–casualty insurance market. *The European Journal of Finance*, *25*(11), 1012-1031.

[2]. Alabi, R. O., Almangush, A., Elmusrati, M., Leivo, I., & Mäkitie, A. (2022). Measuring the usability and quality of explanations of a machine learning web-based tool for oral tongue cancer prognostication. *International Journal of Environmental Research and Public Health*, *19*(14), 8366.

[3]. Alam, N., Gupta, L., & Zameni, A. (2019). *Fintech and Islamic finance*. Springer.

[4]. Alangari, N., El Bachir Menai, M., Mathkour, H., & Almosallam, I. (2023). Exploring evaluation methods for interpretable machine learning: A survey. *Information*, *14*(8), 469.

[5]. Ayranci, P., Lai, P., Phan, N., Hu, H., Kolinowski, A., Newman, D., & Dou, D. (2022). OnML: an ontology-based approach for interpretable machine learning. *Journal of Combinatorial Optimization*, *44*(1), 770-793.

[6]. Banasiewicz, A. (2021a). *Organizational learning in the Age of Data*. Springer.
[7]. Banasiewicz, A. (2021b). Out with the Old–In with the New. In *Organizational Learning in the Age of Data* (pp. 1-36). Springer.
[8]. Barber, S. J. (2020). The applied implications of age-based stereotype threat for older adults. *Journal of Applied Research in Memory and Cognition*, *9*(3), 274-285.
[9]. Barcellos, R., Bernardini, F., & Viterbo, J. (2022). Towards defining data interpretability in open data portals: Challenges and research opportunities. *Information systems*, *106*, 101961.
[10]. Barkham, R., Bokhari, S., & Saiz, A. (2022). Urban big data: city management and real estate markets. In *Artificial intelligence, machine learning, and optimization tools for smart cities: Designing for sustainability* (pp. 177-209). Springer.
[11]. Bekemeier, F. (2023). A primer on the insurability of decentralized finance (DeFi). *Digital Finance*, *5*(3), 643-687.
[12]. Billio, M., Costola, M., Pelizzon, L., & Riedel, M. (2022). Buildings' energy efficiency and the probability of mortgage default: The Dutch case. *The Journal of Real Estate Finance and Economics*, *65*(3), 419-450.
[13]. Bradbury-Jones, C., Aveyard, H., Herber, O. R., Isham, L., Taylor, J., & O'malley, L. (2022). Scoping reviews: the PAGER framework for improving the quality of reporting. *International Journal of Social Research Methodology*, *25*(4), 457-470.
[14]. Brotcke, L. (2022). Time to assess bias in machine learning models for credit decisions. *Journal of Risk and Financial Management*, *15*(4), 165.
[15]. Brown, D., Sorrell, S., & Kivimaa, P. (2019). Worth the risk? An evaluation of alternative finance mechanisms for residential retrofit. *Energy policy*, *128*, 418-430.
[16]. Bunni, N. G., & Bunni, L. B. (2022). *Risk and insurance in construction*. Routledge.
[17]. Cannon, K. P., & Preis, S. J. (2023). Exploring Knowledge-Based Systems for Commercial Mortgage Underwriting. International Conference on Web Engineering,
[18]. Cao, Y., Yang, Y., Ma, H., Kong, X., Li, X., Du, Y., & Chen, D. (2023). Multidimensional Evolution and Driving Factors of Securities Firms' Collaborative Bond Joint Underwriting Networks in China: A Comprehensive Analysis from 2011 to 2020. *Systems*, *11*(5), 253.
[19]. Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), 832.
[20]. Cheung, H. K., Anderson, A. J., King, E. B., Mahabir, B., Warner, K., & Jones, K. P. (2022). Beyond the baby bump: Subtle discrimination against working mothers in the hiring process. *Journal of Business and Psychology*, *37*(6), 1181-1198.
[21]. Choi, J. W., Huh, S., & Jo, S. (2020). Improving performance in motor imagery BCI-based control applications via virtually embodied feedback. *Computers in Biology and Medicine*, *127*, 104079.
[22]. Choi, J. W., Kim, B. H., Huh, S., & Jo, S. (2020). Observing actions through immersive virtual reality enhances motor imagery training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(7), 1614-1622.
[23]. Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. *Ieee Access*, *7*, 24634-24648.
[24]. Cox Jr, L. A. (2023). Natural, Artificial, and Social Intelligence for Decision-Making. In *AI-ML for Decision and Risk Analysis: Challenges and Opportunities for Normative Decision Theory* (pp. 65-101). Springer.
[25]. Dambra, S., Bilge, L., & Balzarotti, D. (2020). SoK: Cyber insurance–technical challenges and a system security roadmap. 2020 IEEE Symposium on Security and Privacy (SP),
[26]. David-Spickermann, F., Mullins, M., & Murphy, F. (2021). Risk-adequate motor underwriting of automated vehicles: a qualitative evaluation using German focus groups. *Environment Systems and Decisions*, *41*(2), 189-197.
[27]. Demchuk, V. (2022). Does intra-bloc trade affect insurance companies' performance? Evidence from North America. *Applied Economics*, *54*(53), 6147-6159.
[28]. Denuit, M., & Trufin, J. (2019). Effective statistical learning methods for actuaries.
[29]. Dragos, C. M., Dragos, S. L., Mare, C., Muresan, G. M., & Purcel, A.-A. (2023). Does risk assessment and specific knowledge impact crop insurance underwriting? Evidence from Romanian farmers. *Economic Analysis and Policy*, *79*, 343-358.
[30]. Eckert, C., & Osterrieder, K. (2020). How digitalization affects insurance companies: overview and use cases of digital technologies. *Zeitschrift für die gesamte Versicherungswissenschaft*, *109*(5), 333-360.
[31]. Eling, M., Elvedi, M., & Falco, G. (2023). The economic impact of extreme cyber risk scenarios. *North American Actuarial Journal*, *27*(3), 429-443.
[32]. Eling, M., & Schnell, W. (2020). Capital requirements for cyber risk and cyber risk insurance: An analysis of Solvency II, the US risk-based capital standards, and the Swiss Solvency Test. *North American Actuarial Journal*, *24*(3), 370-392.
[33]. Fahimul, H. (2022). Corpus-Based Evaluation Models for Quality Assurance Of AI-Generated ESL Learning Materials. *Review of Applied Science and Technology*, *1*(04), 183–215. https://doi.org/10.63125/m33q0j38
[34]. Foster, C. M., & Giovanello, K. S. (2020). Domain general processes moderate age-related performance differences on the mnemonic similarity task. *Memory*, *28*(4), 528-536.
[35]. Frees, E. W., & Huang, F. (2023). The discriminating (pricing) actuary. *North American Actuarial Journal*, *27*(1), 2-24.
[36]. Gohdes, N., Simshauser, P., & Wilson, C. (2022). Renewable entry costs, project finance and the role of revenue quality in Australia's National Electricity Market. *Energy Economics*, *114*, 106312.
[37]. Graziani, M., Dutkiewicz, L., Calvaresi, D., Amorim, J. P., Yordanova, K., Vered, M., Nair, R., Abreu, P. H., Blanke, T., & Pulignano, V. (2023). A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences. *Artificial intelligence review*, *56*(4), 3473-3504.

[38]. Graziani, M., Palatnik de Sousa, I., Vellasco, M. M., Costa da Silva, E., Müller, H., & Andrearczyk, V. (2021). Sharpening local interpretable model-agnostic explanations for histopathology: improved understandability and reliability. International Conference on Medical Image Computing and Computer-Assisted Intervention,

[39]. Grundl, S., & Kim, Y. S. (2019). Consumer mistakes and advertising: The case of mortgage refinancing. *Quantitative Marketing and Economics*, *17*(2), 161-213.

[40]. Hamann, A., & Carstengerdes, N. (2022). Investigating mental workload-induced changes in cortical oxygenation and frontal theta activity during simulated flights. *Scientific reports*, *12*(1), 6449.

[41]. Hammad, S. (2022). Application of High-Durability Engineering Materials for Enhancing Long-Term Performance of Rail and Transportation Infrastructure. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 63-96. https://doi.org/10.63125/4k492a62

[42]. Han, Y., Chong, W. K., & Li, D. (2020). A systematic literature review of the capabilities and performance metrics of supply chain resilience. *International Journal of Production Research*, *58*(15), 4541-4566.

[43]. Haque, B. M. T., & Md. Arifur, R. (2020). Quantitative Benchmarking of ERP Analytics Architectures: Evaluating Cloud vs On-Premises ERP Using Cost-Performance Metrics. *American Journal of Interdisciplinary Studies*, *1*(04), 55-90. https://doi.org/10.63125/y05j6m03

[44]. Haque, B. M. T., & Md. Arifur, R. (2021). ERP Modernization Outcomes in Cloud Migration: A Meta-Analysis of Performance and Total Cost of Ownership (TCO) Across Enterprise Implementations. *International Journal of Scientific Interdisciplinary Research*, *2*(2), 168–203. https://doi.org/10.63125/vrz8hw42

[45]. Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & cognition*, *47*(1), 17-32.

[46]. Hodgson, T. L., Ezard, G., & Hermens, F. (2019). Eye movements in neuropsychological tasks. *Processes of Visuospatial Attention and Working Memory*, 393-418.

[47]. Hofmann, A., & Sattarhoff, C. (2023). Underwriting cycles in property-casualty insurance: The impact of catastrophic events. *Risks*, *11*(4), 75.

[48]. Hou, L., & Jiao, R. J. (2020). Data-informed inverse design by product usage information: a review, framework and outlook. *Journal of Intelligent Manufacturing*, *31*(3), 529-552.

[49]. Hutton, T. (2020). The Role of Computer Modeling in Insurance Risk Securitization. In *Securitized Insurance Risk* (pp. 153-168). Routledge.

[50]. Jabed Hasan, T., & Waladur, R. (2022). Advanced Cybersecurity Architectures for Resilience in U.S. Critical Infrastructure Control Networks. *Review of Applied Science and Technology*, *1*(04), 146–182. https://doi.org/10.63125/5rvjav10

[51]. Jacobs Jr, M. (2020). A holistic model validation framework for current expected credit loss (CECL) model development and implementation. *International Journal of Financial Studies*, *8*(2), 27.

[52]. Javed, A. R., Ahmed, W., Pandya, S., Maddikunta, P. K. R., Alazab, M., & Gadekallu, T. R. (2023). A survey of explainable artificial intelligence for smart cities. *Electronics*, *12*(4), 1020.

[53]. Jeunet, C., Glize, B., McGonigal, A., Batail, J.-M., & Micoulaud-Franchi, J.-A. (2019). Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: Theoretical background, applications and prospects. *Neurophysiologie Clinique*, *49*(2), 125-136.

[54]. Kamble, S. S., & Gunasekaran, A. (2020). Big data-driven supply chain performance measurement system: a review and framework for implementation. *International Journal of Production Research*, *58*(1), 65-86.

[55]. Kayi-Aydar, H. (2021). A framework for positioning analysis: From identifying to analyzing (pre) positions in narrated story lines. *System*, *102*, 102600.

[56]. Keenan, J. M., & Bradt, J. T. (2020). Underwaterwriting: from theory to empiricism in regional mortgage markets in the US. *Climatic Change*, *162*(4), 2043-2067.

[57]. Kester, J. (2022). Insuring future automobility: A qualitative discussion of British and Dutch car insurer's responses to connected and automated vehicles. *Research in Transportation Business & Management*, *45*, 100903.

[58]. Kim, N.-Y. (2019). Do reputable underwriters affect the sustainability of newly listed firms? Evidence from South Korea. *Sustainability*, *11*(9), 2665.

[59]. King, M. R., Timms, P. D., & Rubin, T. H. (2021). Use of big data in insurance. In *The Palgrave Handbook of Technological Finance* (pp. 669-700). Springer.

[60]. Kluve, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stöterau, J., Weidenkaff, F., & Witte, M. (2019). Do youth employment programs improve labor market outcomes? A quantitative review. *World Development*, *114*, 237-253.

[61]. Knudsen, S. V., Laursen, H. V. B., Johnsen, S. P., Bartels, P. D., Ehlers, L. H., & Mainz, J. (2019). Can quality improvement improve the quality of care? A systematic review of reported effects and methodological rigor in plan-do-study-act projects. *BMC health services research*, *19*(1), 683.

[62]. Kong, F. (2021). How to understand the role of insurance mechanism in a global pandemic? *International Journal of Environmental Research and Public Health*, *18*(13), 6743.

[63]. Kumar, V. (2023). Digital hotspots. In *The Economic Value of Digital Disruption: A Holistic Assessment for CXOs* (pp. 689-795). Springer.

[64]. Kwiecień, I., Kowalczyk-Rólczyńska, P., & Popielas, M. (2020). The challenges for life insurance underwriting caused by changes in demography and digitalisation. In *Life Insurance in Europe: Risk Analysis and Market Challenges* (pp. 147-163). Springer.

[65]. Lai, P., Phan, N., Hu, H., Badeti, A., Newman, D., & Dou, D. (2020). Ontology-based interpretable machine learning for textual data. 2020 International Joint Conference on Neural Networks (IJCNN),

[66]. Li, L., Ucar, E., & Yavas, A. (2022). Social capital and mortgage delinquency. *The Journal of Real Estate Finance and Economics*, *64*(3), 379-403.

[67]. Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., & Dou, D. (2022). Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, *64*(12), 3197-3234.

[68]. Liao, J. J., Luo, J. J., Yang, T., So, R. Q. Y., & Chua, M. C. H. (2020). Effects of local and global spatial patterns in EEG motor-imagery classification using convolutional neural network. *Brain-Computer Interfaces*, *7*(3-4), 47-56.

[69]. Lim, Q.-M., Lee, H.-S., & Har, W.-M. (2021). Efficiency, productivity and competitiveness of the Malaysian insurance sector: an analysis of risk-based capital regulation. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *46*(1), 146-172.

[70]. Lozano-Murcia, C., Romero, F. P., Serrano-Guerrero, J., & Olivas, J. A. (2023). A comparison between explainable machine learning methods for classification and regression problems in the actuarial context. *Mathematics*, *11*(14), 3088.

[71]. MacAskill, S., Sahin, O., Stewart, R., Roca, E., & Liu, B. (2021). Examining green affordable housing policy outcomes in Australia: A systems approach. *Journal of Cleaner Production*, *293*, 126212.

[72]. Machado, C. G., Winroth, M. P., & Ribeiro da Silva, E. H. D. (2020). Sustainable manufacturing in Industry 4.0: an emerging research agenda. *International Journal of Production Research*, *58*(5), 1462-1484.

[73]. Mahapatra, P., & Singh, S. K. (2021). Artificial intelligence and machine learning: discovering new ways of doing banking business. In *Artificial intelligence and machine learning in business management* (pp. 53-80). CRC Press.

[74]. Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(4), 859-876.

[75]. Malik, A., Pereira, V., & Tarba, S. (2019). The role of HRM practices in product development: Contextual ambidexterity in a US MNC's subsidiary in India. *The International Journal of Human Resource Management*, *30*(4), 536-564.

[76]. Malik, A., & Ullah, K. (2019). Underwriting Management. In *Introduction to Takaful: Theory and Practice* (pp. 129-153). Springer.

[77]. McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., Kidd, E., & Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. *Scientific reports*, *9*(1), 12820.

[78]. Md Ashraful, A., Md Fokhrul, A., & Md Fardaus, A. (2020). Predictive Data-Driven Models Leveraging Healthcare Big Data for Early Intervention And Long-Term Chronic Disease Management To Strengthen U.S. National Health Infrastructure. *American Journal of Interdisciplinary Studies*, *1*(04), 26-54. https://doi.org/10.63125/1z7b5v06

[79]. Md Fokhrul, A., Md Ashraful, A., & Md Fardaus, A. (2021). Privacy-Preserving Security Model for Early Cancer Diagnosis, Population-Level Epidemiology, And Secure Integration into U.S. Healthcare Systems. *American Journal of Scholarly Research and Innovation*, *1*(02), 01–27. https://doi.org/10.63125/q8wjee18

[80]. Md Harun-Or-Rashid, M., & Sai Praveen, K. (2022). Data-Driven Approaches To Enhancing Human–Machine Collaboration In Remote Work Environments. *International Journal of Business and Economics Insights*, *2*(3), 47-83. https://doi.org/10.63125/wt9t6w68

[81]. Md. Arifur, R., & Haque, B. M. T. (2022). Quantitative Benchmarking of Machine Learning Models for Risk Prediction: A Comparative Study Using AUC/F1 Metrics and Robustness Testing. *Review of Applied Science and Technology*, *1*(03), 32–60. https://doi.org/10.63125/9hd4e011

[82]. Md. Towhidul, I., Alifa Majumder, N., & Mst. Shahrin, S. (2022). Predictive Analytics as A Strategic Tool For Financial Forecasting and Risk Governance In U.S. Capital Markets. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 238–273. https://doi.org/10.63125/2rpyze69

[83]. Mioni, G., Cardullo, S., Ciavarelli, A., & Stablum, F. (2021). Age-related changes in time discrimination: The involvement of inhibition, working memory and speed of processing. *Current Psychology*, *40*(5), 2462-2471.

[84]. Mourmouris, J., & Poufinas, T. (2023). Multi-criteria decision-making methods applied in health-insurance underwriting. *Health Systems*, *12*(1), 52-84.

[85]. Nurse, J. R., Axon, L., Erola, A., Agrafiotis, I., Goldsmith, M., & Creese, S. (2020). The data that drives cyber insurance: A study into the underwriting and claims processes. 2020 International conference on cyber situational awareness, data analytics and assessment (CyberSA),

[86]. Nyanchoka, L., Tudur-Smith, C., Iversen, V., Tricco, A. C., & Porcher, R. (2019). A scoping review describes methods used to identify, prioritize and display gaps in health research. *Journal of clinical epidemiology*, *109*, 99-110.

[87]. Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., & Castignani, G. (2022). Explainable artificial intelligence (xai) in insurance. *Risks*, *10*(12), 230.

[88]. Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., & Grieman, K. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391-437.

[89]. Pan, T., Shu, F., Kitterlin-Lynch, M., & Beckman, E. (2021). Perceptions of cruise travel during the COVID-19 pandemic: Market recovery strategies for cruise businesses in North America. *Tourism Management*, *85*, 104275.

[90]. Panda, S., Woods, D. W., Laszka, A., Fielder, A., & Panaousis, E. (2019). Post-incident audits on cyber insurance discounts. *Computers & Security*, *87*, 101593.

[91]. Patil, S., Patil, A., Patil, V., Yadav, S., & Das, S. (2023). Cognitive Data Underwriting with Automation: A Survey. International Conference on Business Data Analytics,

[92]. Peddie, J. (2019). Ray-tracing hardware. In *Ray Tracing: A Tool for All* (pp. 129-180). Springer.

[93]. Poufinas, T., Gogas, P., Papadimitriou, T., & Zaganidis, E. (2023). Machine learning in forecasting motor insurance claims. *Risks*, *11*(9), 164.

[94]. Prokopy, L. S., Floress, K., Arbuckle, J. G., Church, S. P., Eanes, F. R., Gao, Y., Gramig, B. M., Ranjan, P., & Singh, A. S. (2019). Adoption of agricultural conservation practices in the United States: Evidence from 35 years of quantitative literature. *Journal of Soil and Water Conservation*, *74*(5), 520-534.

[95]. Pu, S., & Lam, J. S. L. (2021). Blockchain adoptions in the maritime industry: a conceptual framework. *Maritime Policy & Management*, *48*(6), 777-794.

[96]. Ratul, D., & Subrato, S. (2022). Remote Sensing Based Integrity Assessment of Infrastructure Corridors Using Spectral Anomaly Detection and Material Degradation Signatures. *American Journal of Interdisciplinary Studies*, *3*(04), 332-364. https://doi.org/10.63125/1sdhwn89

[97]. Rauf, M. A. (2018). A needs assessment approach to english for specific purposes (ESP) based syllabus design in Bangladesh vocational and technical education (BVTE). *International Journal of Educational Best Practices*, *2*(2), 18-25.

[98]. Rawat, K. (2023). Applying CRISP-DM Methodology in Developing Machine Learning Model for Credit Risk Prediction. Science and Information Conference,

[99]. Reguero, B. G., Beck, M. W., Schmid, D., Stadtmüller, D., Raepple, J., Schüssele, S., & Pfliegner, K. (2020). Financing coastal resilience by combining nature-based risk reduction with insurance. *Ecological Economics*, *169*, 106487.

[100]. Reteig, L. C., van den Brink, R. L., Prinssen, S., Cohen, M. X., & Slagter, H. A. (2019). Sustaining attention for a prolonged period of time increases temporal variability in cortical responses. *Cortex*, *117*, 16-32.

[101]. Richter, A., & Wilson, T. C. (2020). Covid-19: implications for insurer risk management and the insurability of pandemic risk. *The Geneva risk and insurance review*, *45*(2), 171-199.

[102]. Riechelmann, E., Gamer, M., Böckler, A., & Huestegge, L. (2021). How ubiquitous is the direct-gaze advantage? Evidence for an averted-gaze advantage in a gaze-discrimination task. *Attention, Perception, & Psychophysics*, *83*(1), 215-237.

[103]. Rifat, C., & Jinnat, A. (2022). Optimization Algorithms for Enhancing High Dimensional Biomedical Data Processing Efficiency. *Review of Applied Science and Technology*, *1*(04), 98–145. https://doi.org/10.63125/2zg6x055

[104]. Rifat, C., & Khairul Alam, T. (2022). Assessing The Role of Statistical Modeling Techniques in Fraud Detection Across Procurement And International Trade Systems. *American Journal of Interdisciplinary Studies*, *3*(02), 91-125. https://doi.org/10.63125/gbdq4z84

[105]. Roesmann, K., Kroker, T., Hein, S., Rehbein, M., Winker, C., Leehr, E. J., Klucken, T., & Junghöfer, M. (2022). Transcranial direct current stimulation of the ventromedial prefrontal cortex modulates perceptual and neural patterns of fear generalization. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *7*(2), 210-220.

[106]. Saja, A. A., Goonetilleke, A., Teo, M., & Ziyath, A. M. (2019). A critical review of social resilience assessment frameworks in disaster management. *International journal of disaster risk reduction*, *35*, 101096.

[107]. Salzman, J., & Ruhl, J. (2019). Currencies and the commodification of environmental law. In *Environmental Law* (pp. 3-90). Routledge.

[108]. Schallmo, M.-P., Kolodny, T., Kale, A. M., Millin, R., Flevaris, A. V., Edden, R. A., Gerdts, J., Bernier, R. A., & Murray, S. O. (2020). Weaker neural suppression in autism. *Nature communications*, *11*(1), 2675.

[109]. Schwarze, V., Marschner, C., Negrão de Figueiredo, G., Knösel, T., Rübenthaler, J., & Clevert, D.-A. (2020). Single-center study: The diagnostic performance of contrast-enhanced ultrasound (CEUS) for assessing renal oncocytoma. *Scandinavian Journal of Urology*, *54*(2), 135-140.

[110]. Siegel, R., Antony, J., Garza-Reyes, J. A., Cherrafi, A., & Lameijer, B. (2019). Integrated green lean approach and sustainability for SMEs: From literature review to a conceptual framework. *Journal of Cleaner Production*, *240*, 118205.

[111]. Smith, F. W., & Smith, M. L. (2019). Decoding the dynamic representation of facial expressions of emotion in explicit and incidental tasks. *NeuroImage*, *195*, 261-271.

[112]. Somani, A., Horsch, A., & Prasad, D. K. (2023). *Interpretability in deep learning* (Vol. 2). Springer.

[113]. Stubbs, T., Reinsberg, B., Kentikelenis, A., & King, L. (2020). How to evaluate the effects of IMF conditionality: An extension of quantitative approaches and an empirical application to public education spending. *The Review of International Organizations*, *15*(1), 29-73.

[114]. Tarr, A. A., Smith, D., Thompson, M., Chamberlain, T., Peña, A., & Golden, S. (2021). Underwriting drone insurance. In *Drone Law and Policy* (pp. 404-418). Routledge.

[115]. Taylor, R. P. (2021). The potential of biophilic fractal designs to promote health and performance: A review of experiments and applications. *Sustainability*, *13*(2), 823.

[116]. Tsohou, A., Diamantopoulou, V., Gritzalis, S., & Lambrinoudakis, C. (2023). Cyber insurance: state of the art, trends and future directions. *International Journal of Information Security*, *22*(3), 737-748.

[117]. Upreti, V., Adams, M., & Jia, Y. (2022). Risk management and the cost of equity: evidence from the United Kingdom's non-life insurance market. *The European Journal of Finance*, *28*(6), 551-570.

[118]. Valente, F., Paredes, S., Henriques, J., Rocha, T., de Carvalho, P., & Morais, J. (2022). Interpretability, personalization and reliability of a machine learning based clinical decision support system. *Data Mining and Knowledge Discovery*, *36*(3), 1140-1173.

[119]. Van Meel, C., Baeck, A., Gillebert, C. R., Wagemans, J., & de Beeck, H. P. O. (2019). The representation of symmetry in multi-voxel response patterns and functional connectivity throughout the ventral visual stream. *NeuroImage*, *191*, 216-224.

[120]. Vassilakopoulou, P., & Hustad, E. (2023). Bridging digital divides: A literature review and research agenda for information systems research. *Information Systems Frontiers*, *25*(3), 955-969.

[121]. Wang, J., Wang, Z., Chen, M., Xiao, Y., Chen, S., Wu, L., Yao, L., Jiang, X., Li, J., & Xu, M. (2022). An interpretable artificial intelligence system for detecting risk factors of gastroesophageal variceal bleeding. *NPJ Digital Medicine*, *5*(1), 183.
[122]. Wen, J., Huang, S. S., & Goh, E. (2020). Effects of perceived constraints and negotiation on learned helplessness: A study of Chinese senior outbound tourists. *Tourism Management*, *78*, 104059.
[123]. Wickham, J., Stehman, S. V., Neale, A. C., & Mehaffey, M. (2020). Accuracy assessment of NLCD 2011 percent impervious cover for selected USA metropolitan areas. *International journal of applied earth observation and geoinformation*, *84*, 101955.
[124]. Wong, E. C., Maher, A. R., Motala, A., Ross, R., Akinniranye, O., Larkin, J., & Hempel, S. (2022). Methods for identifying health research gaps, needs, and priorities: a scoping review. *Journal of General Internal Medicine*, *37*(1), 198-205.
[125]. Xu, C., Liao, Z., Li, C., Zhou, X., & Xie, R. (2022). Review on interpretable machine learning in smart grid. *Energies*, *15*(12), 4427.
[126]. Xu, S., Liu, X., Lu, Q., & Chang, S. (2022). Determination method of environmental damage liability insurance underwriting coverage of offshore factory wastewater based on OTUS. *Desalination and Water Treatment*, *245*, 54-61.
[127]. Xu, W. (2022). Legacies of institutionalized redlining: a comparison between speculative and implemented mortgage risk maps in Chicago, Illinois. *Housing Policy Debate*, *32*(2), 249-274.
[128]. Yakavenka, V., Mallidis, I., Vlachos, D., Iakovou, E., & Eleni, Z. (2020). Development of a multi-objective model for the design of sustainable supply chains: The case of perishable food products. *Annals of Operations Research*, *294*(1), 593-621.
[129]. Yan, X. (2023). Research on financial field integrating artificial intelligence: Application basis, case analysis, and SVR model-based overnight. *Applied Artificial Intelligence*, *37*(1), 2222258.
[130]. Yew, G. C. K. (2020). *Health law and medical ethics in Singapore*. Routledge.
[131]. Yu, B., Hu, W., Xu, L., Tang, J., Liu, S., & Zhu, Y. (2020). Building the computing system for autonomous micromobility vehicles: Design constraints and architectural optimizations. 2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO),
[132]. Zaman, M. A. U., Sultana, S., Raju, V., & Rauf, M. A. (2021). Factors İmpacting the Uptake of İnnovative Open and Distance Learning (ODL) Programmes in Teacher Education. *Turkish Online Journal of Qualitative Inquiry*, *12*(6).
[133]. Zeier Röschmann, A., Erny, M., & Wagner, J. (2022). On the (future) role of on-demand insurance: Market landscape, business model and customer perception. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *47*(3), 603-642.
[134]. Zhou, J., Gandomi, A. H., Chen, F., & Holzinger, A. (2021). Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics*, *10*(5), 593.