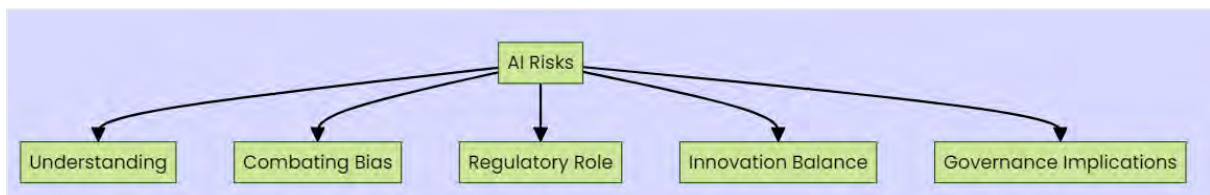


Mitigate Biased Decision-Making in AI Algorithms

Jiaming Zuo, FSA, CERA, FCAA, FASHK

Any views and ideas expressed in the essays are the author's alone and may not reflect the views and ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, nor the author's employer.

The widespread use of artificial intelligence (AI) in the insurance sector brings several risks and challenges that companies need to navigate to ensure the responsible and ethical deployment of AI technologies. Here are some key risks emerging from the use of AI in the insurance industry.



Addressing algorithmic bias in AI systems requires a combination of technical approaches, ethical considerations, and regulatory oversight. Techniques such as bias detection, data preprocessing, fairness constraints, and explainable AI can help mitigate bias and promote more equitable and transparent decision-making in AI systems. By understanding how bias manifests and taking proactive steps to address it, developers and users can work towards building fairer and more inclusive AI technologies.

SEVERAL FACTORS CAN CONTRIBUTE TO ALGORITHMIC BIAS IN AI SYSTEMS

Several factors can contribute to the emergence of algorithmic bias in AI systems. These factors often intersect throughout the AI development process and can influence the presence and extent of bias in the resulting algorithms. Here are some key factors that contribute to the emergence of algorithmic bias:

Biased Training Data

The most common source of algorithmic bias is biased training data. Historical data often reflects societal biases, stereotypes, or systemic inequalities, which can be unintentionally encoded into AI systems during the training phase.

For example, an insurance company uses an AI algorithm to determine car insurance premiums for policyholders based on historical claims data. The training data predominantly consists of claims data from urban areas, leading to overrepresentation of claims from city drivers. The dataset lacks sufficient data from rural areas and under-represents low-income individuals.

Data Selection Bias

Data selection bias occurs when certain groups or perspectives are underrepresented or overrepresented in the training data, leading to skewed or incomplete datasets that do not accurately reflect the full range of real-world scenarios. The Imagin insurance company uses an AI algorithm to assess health insurance risk profiles based on historical claims data. The training data primarily consists of claims data from individuals who have regularly visited healthcare facilities and have a higher documented medical history. The dataset lacks representation of healthy individuals or those who may have had minimal healthcare needs.

Data Labeling Bias

Biases can also be introduced during the data labeling process, where human annotators may unknowingly inject their biases into the training data through subjective or culturally influenced labeling decisions. An insurance company uses an AI algorithm to assess risk profiles for home insurance policies. The algorithm relies on labeled data to identify risk factors associated with properties. The data labeling process involves human annotators who unintentionally introduce biases in determining property risks based on subjective judgments or assumptions.

Algorithm Design Choices

Algorithmic bias can be unintentionally introduced through design choices such as feature selection, model complexity, hyperparameter tuning, or optimization strategies. Biased assumptions embedded in the algorithm design can lead to biased outcomes.

Feedback Loop Effects

AI systems that interact with users and learn from feedback data can develop feedback loop bias. If the feedback data is biased, the system may reinforce or amplify existing biases over time, leading to discriminatory outcomes.

Contextual Biases

The context in which AI systems are deployed can also contribute to algorithmic bias. Biases may emerge from specific use cases, application domains, cultural norms, or social structures that influence the data collection, algorithm design, or decision-making processes.

Human Involvement

Humans involved in the AI development lifecycle, including data scientists, engineers, and designers, can introduce biases consciously or unconsciously. Their subjective judgment, prior beliefs, assumptions, or cultural influences can shape the AI system's behavior.

Lack of Diversity

Lack of diversity in AI development teams or insufficient representation of diverse perspectives and voices can contribute to the perpetuation of biases in AI systems. Diverse teams can bring different viewpoints and experiences to identify and address bias effectively.

Addressing algorithmic bias requires a holistic approach that involves careful data curation, transparency in algorithmic decision-making, diversity in AI teams, ongoing monitoring for bias, and the integration of fairness considerations throughout the AI development lifecycle. By understanding and mitigating the

factors that contribute to bias, developers and practitioners can work towards creating AI systems that are more equitable, accountable, and inclusive.

THE CONSEQUENCES OF ALGORITHMIC BIAS ON ACTUARIAL ANALYSIS

In actuarial analysis, AI algorithms trained on biased data may lead to discriminatory outcomes in insurance underwriting and pricing practices. One example of this is the use of historical claims data that may contain inherent biases related to factors like race, gender, or socioeconomic status. If AI algorithms are trained on this biased data, they may inadvertently perpetuate these biases in insurance risk assessments and pricing decisions, leading to discriminatory outcomes for certain demographic groups.

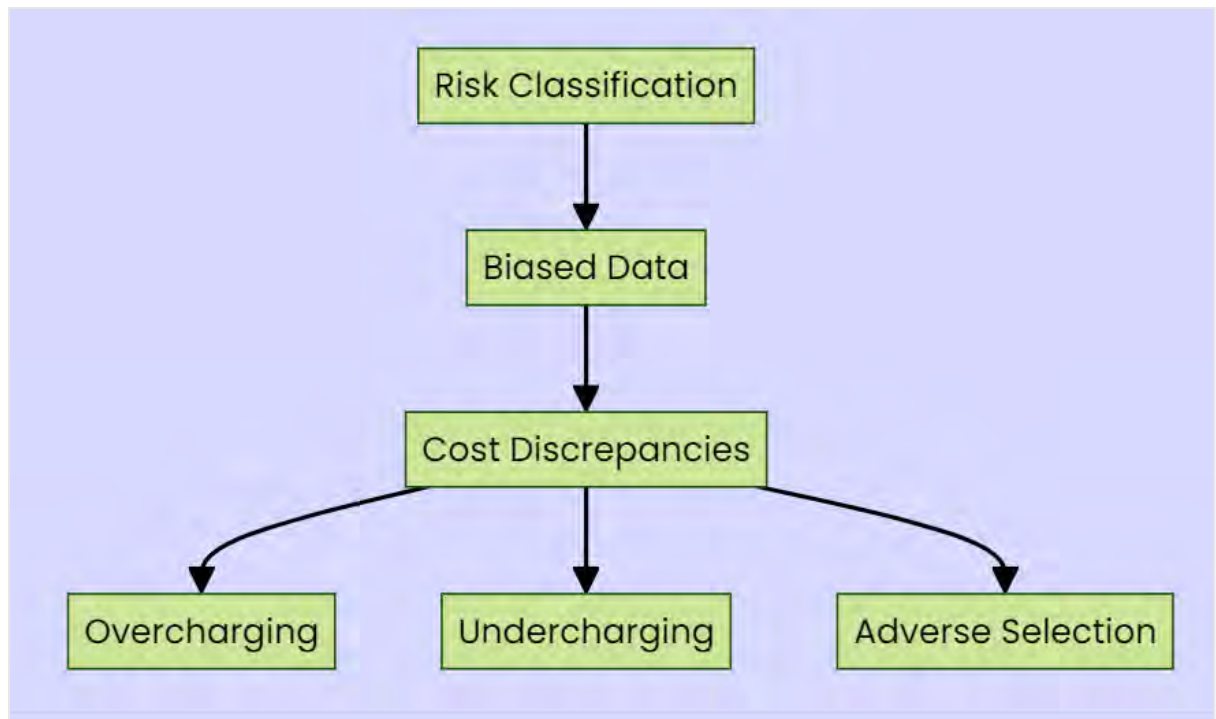
As noted above, actuarial services are data-driven; data bias, left unaddressed, can lead to incorrect conclusions, unwanted consequences, wrong policy decisions, or inadequate system performance. This section provides a few examples of actuarial services that can be impacted by data bias.

RISK CLASSIFICATION

Risk classification is the process of evaluating and estimating the future costs related to transferring risk. Biased data can introduce discrepancies between the actual future costs and the actuary's projections, potentially leading to overcharging or undercharging and adverse selection. Availability bias and historical bias are two significant factors that can impact actuarial decisions.

Using a machine learning model to predict life insurance policyholder mortality rates, an insurer inadvertently incorporates biased data that skews towards affluent applicants. As a result, the algorithm may underestimate risks for certain demographic groups, leading to improper risk assessments and potentially unsustainable pricing strategies. An additional instance of bias is historical bias, where differences in homeownership by race are overlooked in a personal auto rating plan. This omission can lead actuaries to base results on this bias rather than the genuine driver of future loss performance.

The Imagin insurance company uses an AI algorithm to assess risk profiles and determine premiums for auto insurance. If the algorithm is trained on biased data that correlates accidents with a specific demographic group rather than driving behavior, it may unfairly penalize individuals in that group with higher premiums, leading to discrimination and perpetuating unfair practices.



By paying more attention to possible historical influences in the data, the actuary can focus on the true drivers of future expected costs such as experience and driving record.

EXPERIENCE STUDIES

Experience studies are instrumental for life insurers in establishing accurate assumptions for life and annuity policy premiums. By aggregating mortality data from life companies to generate industry mortality tables, insurers inform their premium calculations. This process is mirrored in lapse assumptions, aiding insurers in comprehending policyholder decrements beyond mortality factors. A comprehensive understanding of unbiased historical data is paramount to mitigate the risks of poor assumption development and underpricing life insurance policies.

An AI system is used to automate claims processing for health insurance. If the algorithm is biased to associate certain medical conditions with higher costs or lower validity, it may systematically deny or delay legitimate claims from individuals with those conditions, resulting in unfair treatment and negative financial impacts for affected policyholders.

One more example, imagine an insurance company using an AI algorithm to analyze historical claims data for setting insurance premiums. The AI algorithm is trained on past claims data that inadvertently reflects biases against certain demographic groups, such as age or gender. Due to this biased training data, the algorithm may learn patterns that unfairly penalize older policyholders by overestimating their risk levels compared to younger policyholders.

RESERVING

The reserving actuary's core task is to estimate future claims and expenses by analyzing claims data and other experiences to establish crucial assumptions. In the domain of property and casualty insurance, the

claims department is responsible for managing loss payments and reserves for future loss and expense payments. However, reserve analyses may encounter aggregation bias, particularly when attempting to generalize development patterns across diverse data subsets such as long-tailed liability and short-tailed property data.

MODELING

Actuaries employ modeling and advanced analytical techniques to refine decision-making in insurance operations. Omitted variable bias poses a substantial threat to risk classification models, as the exclusion of critical variables can introduce spurious correlations or signal loss, resulting in less comprehensive models and potentially skewed coefficient estimates. Similarly, confirmation bias can hinder predictive modeling efforts, prompting actuaries to manipulate models until they align with preconceived expectations, influencing assumption selection in reserving practices.

For instance, an insurance company implements AI-powered dynamic pricing for home insurance based on property features and location. If the algorithm incorporates biased assumptions about neighborhood characteristics or housing types, it may inadvertently undervalue or overvalue certain properties, leading to price disparities that disproportionately affect specific groups of policyholders.

REDUCING AND MITIGATING ALGORITHMIC BIAS IN AI SYSTEMS

Reducing and mitigating algorithmic bias in AI systems for actuarial analysis is essential to ensure fair and accurate decision-making processes. Here are some approaches and techniques that can help address algorithmic bias in actuarial analysis:

Diverse and Representative Data

To enhance fairness in model predictions, it is imperative to utilize diverse, inclusive, and representative training data for AI models, minimizing biases and ensuring equitable outcomes

Data Preprocessing

To fix biases in training data, use data preprocessing techniques like de-biasing, cleaning, feature tweaking, and balancing to ensure fair model training.

Fairness Constraints

Ensure fairness in AI algorithms by adding constraints to prevent discrimination during training, stopping biased patterns, and ensuring fair outcomes for everyone.

Explainable AI (XAI)

Use explainable AI (XAI) techniques to make AI models clearer and easier to understand. XAI shows how algorithms make decisions, helping to find and fix biased patterns.

Bias Audits

Conduct regular bias audits to assess and identify biases in AI models. Evaluate model performance across different demographic groups and identify disparities that may indicate algorithmic bias. Adjust the model parameters as needed to mitigate bias.

Human Oversight

Incorporate human oversight into AI systems to review and interpret model decisions, especially in sensitive or high-stakes applications such as actuarial analysis. Human intervention is critical for ensuring ethical and transparent AI-driven decision-making.

Sensitive Feature Removal

Remove or de-emphasize sensitive attributes (such as race, gender, or age) from the input data used for training AI models to prevent unwanted biases from influencing model predictions.

Regular Monitoring and Evaluation

Continuously monitor and evaluate AI systems for bias post-deployment. Implement feedback mechanisms, conduct bias testing, and solicit diverse perspectives to ensure fair and equitable outcomes over time.

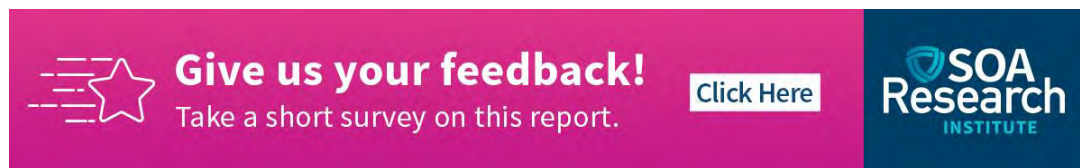
Diverse Teams and Stakeholder Engagement


Encourage diversity and inclusion in AI teams to bring different perspectives to AI systems. Involve stakeholders, including those affected, to get feedback and ensure AI systems are ethical and fair.

By implementing these approaches and techniques, insurance companies and actuarial teams can work towards reducing algorithmic bias in AI systems used for actuarial analysis. Prioritizing fairness, transparency, and inclusivity in AI development processes can help build more reliable, ethical, and equitable AI-driven decision-making in the insurance industry.

* * * * *

Jiaming Zuo is a Senior Partner for EverBright Actuarial Consulting Limited. EverBright integrated AI into their digital platform for customizing and managing group health and insurance policies for clients. She can be reached at jzuo@ebactuary.com.



 **Give us your feedback!**
Take a short survey on this report. [Click Here](#) 