Proposed Guidance / Model Bulletin

Use of Algorithms, Predictive Models and Artificial Intelligence Systems by Regulated Entities

Recommendations of the Center for Economic Justice

August 12, 2023

I. Purpose of the Bulletin

The use of new, non-traditional data types and data sources, combined with the ability to process huge amounts of information and deploy the results of algorithms and artificial intelligence applications in real time is the most significant change in insurance in a generation. For purposes of this bulletin, big data and artificial intelligence (AI) are the terms used to describe the vast new data and new technologies, respectively. The term AI Systems encompasses both big data and AI.

Insurers' (and other regulated entities') use of AI Systems hold tremendous promise to reduce the cost of insurance, increase the availability and affordability of insurance, more quickly bring products to market, improve risk management, close the protection gap, create greater transparency of insurance products and processes, create risk prevention and loss mitigation opportunities and partnerships and reduce the impact of structural racism in insurance. But, these outcomes are not guaranteed. For nearly every potential benefit, there is a potential downside. The purpose of this bulletin is to alert regulated entities to your responsibilities regarding the use of AI Systems, how to ensure compliance with relevant laws and regulation, provide guidance for what the Department expects of you regarding such use and to alert you to changes in the Department's regulatory practices to align with current and emerging technology.

II. Basis for the Guidance

The basis for this guidance are the Principles of Artificial Intelligence adopted by the National Association of Insurance Commissioners (NAIC AI Principles) in 2020 and current legislative authority. This bulletin provides the guidance to implement the NAIC AI Principles,[1] which are consistent with and informed by current legislative authority. Regulated entities' use of AI Systems must comply with the letter and spirit of insurance laws and regulations.

[1] The NAIC AI Principles state, "This document should be used to assist regulators and NAIC committees addressing insurance-specific AI applications."

The regulatory guidance relies upon current laws and regulations, including:

- Unfair Trade Practices Model Act (#880)
- Unfair Claims Settlement Practices Model Act (#900)
- Corporate Governance Annual Disclosure Model Act (#305)
- Property and Casualty Model Rating Law (#1780): The Property and Casualty Model Rating Law, [insert citation to state statute or regulation corresponding to the Model #1780], requires that property/casualty (P/C) insurance rates not be excessive, inadequate, or unfairly discriminatory and provides the regulatory framework for licensing and oversight of advisory organizations.
- Market Conduct Surveillance Model Law (#693)

III. Nature of the Guidance

The guidance provided in this bulletin includes the requirement for a regulated entity to establish and document a governance program to manage its AI Systems applications. While the bulletin offers resources for regulated entities regarding approaches to AI governance, the method of AI Systems risk management and governance is left to the regulated entity as long as that governance system produces the outcomes set out in the NAIC AI Principles and further developed in this bulletin. The guidance is, for the most part, outcomes-based to guide the implementation of the AI Principles.

The outcomes-based guidance focuses on the consumer-facing AI Systems applications used by the regulated entity. Insurers also utilize AI Systems applications for other aspects of their operations, including investment decisions, enterprise risk management and establishing reserves, among others. Application of AI Systems governance and risk management is also essential for these non-consumer facing AI Systems tools.

Regulated entities' use of AI Systems is rapidly evolving. The intent of this guidance is highlight the guardrails of greatest importance to the Department and have regulated entities report their experience implementing and using the guardrails. The Department expects that the guidance will develop further over time. However, the consumer protection issues are sufficiently important for regulated entities to start addressing the potential harms of AI Systems applications as set out in this bulletin. The Department will update this guidance as needed.

This bulletin does not address cybersecurity because cybersecurity guidance has previously been provided to regulated entities. [insert relevant reference]

IV. All Consumer-Facing AI Systems applications are High Risk

> Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. [2]

The Department views all consumer-facing AI applications are high risk – whether for marketing, underwriting, pricing, claims settlement, antifraud, consumer relations/consumer information or loss prevention and risk mitigation. A flawed algorithm can unfairly limit product offerings, deny coverage, charge unfair prices, unfairly settle claims, incorrectly label a claim as suspicious, provide false or misleading information or prevent effective risk mitigation and loss prevention. A flawed consumer-facing algorithm can deny a consumer essential insurance coverage or the benefits of purchased coverage resulting in catastrophic consequences for the consumer. All of the following potential harms represent this high risk to consumers:

- A marketing algorithm that systematically denies product options on the basis of race;
- A policy form algorithm that generates policy language and provisions but produces unclear, misleading, deceptive, unfair or prohibited provisions;
- A pricing algorithm that systematically charges people on the basis of race;
- A claims settlement algorithm that systematically offers lower claims settlements on the basis of race;
- An antifraud algorithm that reflects and perpetuates historic racial discrimination in policing and criminal justice;
- A chatbot that provides misleading or false information to consumers that causes consumers to not get the benefits of their purchase; or
- An algorithm designed to provide relevant loss prevention tools to policyholders that systematically that systematically offers less opportunity to communities of color.

---

[2] Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *Columbia Law Review* at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

V. Algorithm vs. Artificial Intelligence vs. Machine Learning

Artificial Intelligence is a broad term that refers to the use of technologies to build machines and computers that have the ability to mimic cognitive functions associated with human intelligence.

An algorithm is a formula or computer code that rapidly executes decision rules set by programmers, or in the case of machine learning, revises decision rules based on ongoing ingestion and analysis of data. With machine learning AI applications, the algorithm can change without human intervention.

An algorithm can be as simple as the premium calculation formula in a rate filing. A machine learning AI application might be a learning algorithm that analyzes consumer characteristics and the nature of the consumer's inquiry to provide automated response (chatbot) or to route the consumer to a consumer service representative most likely to meet the insurer's outcome goals. Another example of machine learning AI applications might be claim settlement anti-fraud algorithms that change as new data are received during the claim settlement process for individual claims or in aggregate.

This bulletin utilize the term "algorithm" broadly to refer to AI Systems and AI applications.

VI. Definitions

Unfair Discrimination Actuarial Basis is one of two types of unfair discrimination in insurance. A practice is unfairly discriminatory if there is not actuarial basis for different treatment of consumers in underwriting, pricing, claims settlement, antifraud, customer relations, risk prevention and loss mitigation practices. Unfair Discrimination on the Actuarial Basis occurs when similarly situated consumers are treated differently – there is no distinction in the cost of the transfer of risk to justify different treatment of the consumers.

Unfair Discrimination Protected Class Basis is the second type of unfair discrimination in insurance and means that insurers are prohibited from treating consumers differently on the basis of a protected class characteristic. The protected classes in this state include race, religion, national origin [insert others]. There are two types of protected class unfair discrimination – proxy discrimination and disparate impact.

Proxy Discrimination means that a data type or algorithm or AI system is predicting a protected class characteristic and not the insurance outcome. Consequently, the facial relationship between the data type, algorithm or AI system and the insurance outcome is spurious, a proxy for the protected class characteristic and, consequently, discriminating on the basis of that protected class characteristic. Proxy discrimination is a violation of both the actuarial and protected class bases for unfair discrimination.

Disparate Impact means that a data type, algorithm or AI system is producing outcomes that disproportionately affect groups of consumers as defined by protected class characteristics, but comply with the actuarial basis for fair discrimination. Disparate impact is not a violation per se, as set out in this guidance, but efforts to minimize disparate impact within the cost- and risk-based foundation of insurance is part of this guidance.

Equity Trade Off means balancing public policy goals with the efficiency and accuracy of an algorithm of AI system. An example of an equity trade-off is the prohibition on discrimination on the basis of race (or other protected class characteristic). The legislature has made the decision that, regardless of actuarial fairness, there is a public policy goal of not discriminating on the basis of race.

On the Basis Of means direct or indirect discrimination related to a protected class characteristic. A data source, algorithm or AI system that has the same or similar effect as intentional discrimination against groups of consumers with protected class characteristic is discriminating on the basis of that protected class characteristic.

Advisory Organization means a third party entity that is licensed or should be licensed pursuant to [insert statutory reference] to collect information from insurers and provide guidance to insurers for phases of the insurance life cycle. The licensing and oversight of advisory organizations by the Department represents the state oversight of collective decision-making activities that exempts those activities from federal antitrust enforcement.

Third Party Not Advisory Organization – means, for purposes of this bulletin, an entity other than the insurer that provides data, algorithms, resources or other services related to AI Systems used by the regulated entity.

Statistical Agent means an entity designated by the Department to collect information from insurers on behalf of the Department, typically pursuant to a statistical plan approved by the Department.

Data Type means a singular characteristic of the consumer, vehicle, property, built or natural environment as well as data generated by the consumer. Data types are the building blocks for AI Systems applications. An algorithm or AI application will typically utilize multiple data types in both development and deployment.

Consumer means a person or organization that applies for, obtains or uses an insurance policy or contract and includes an applicant for insurance, a policyholder and a claimant.

Data Source means the origin of the data type, including provided directly by the consumer, generated by the consumer in course of applying for, maintain or using the insurance contract, generated by the insurer in course of a consumer applying for, maintaining or using the insurance contract, third party advisory organizations, third party not advisory organizations, government records. Data origins provided directly by the consumer include data provided in the application or through interaction with the regulated entity, including data generated from telematics in the vehicle, home or wearable device. Third party not advisory organizations include data brokers, online data aggregators and social media platforms, web sites and mobile device carriers.

Phase of the Insurance Life Cycle means the consumer-facing practices related to product development, marketing, underwriting, pricing, claims settlement, antifraud, policy administration, customer relations, loss prevention and risk mitigation.

VII. AI Systems Risk Management and Governance System

The Department expects regulated entities to have in place governance and oversight of your internal data, third party-supplied data, algorithms, predictive models and artificial intelligence, including any machine learning applications.

The Department does not specify any particular approach or structure for governance and risk management of AI Systems. There are numerous resources available for insurers regarding governance programs. The National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF)[3] is one excellent example.

The Department does expect and require that whatever the governance and risk management approach utilized by the regulated entity, that governance and risk management framework produces the outcomes set out in the next section of the bulletin. The Department also expects that the regulated entity will have written documentation and procedures to implement your AI Systems governance and risk management. The Department also expects your AI Systems governance and risk management will include ongoing assessment of performance and procedures to identify and remediate poor outcomes. If the Department determines that any of your AI Systems applications are producing poor outcomes, the Department may examine the governance and risk management framework in great detail to identify the source of poor outcomes.

In addition, the Department's expectations regarding regulated entities governance and risk management of AI systems (AIS Governance) include:

---

[3] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

- The AIS Governance should be designed to ensure the use of AI Systems will not violate any laws or regulations and the chief executive officer of the regulated entity is responsible for such compliance.
- The AIS Governance should address governance, risk management controls, and internal audit functions.
- The AIS Governance should address the use of AI Systems across all phases of the insurance product life cycle.
- The AIS Governance should address all of the AI Systems used by or on behalf of the Insurer to make decisions, whether developed by the Insurer or a third party and whether used by the Insurer or by an authorized agent or representative of the Insurer.

VIII. Required Outcomes of Your Use of AI Systems Applications

Whatever type or method of AIS Governance utilized, the following outcomes are required to ensure compliance with the various statutory requirements discussed above.

1. Disputability

You must be able to identify and explain why a particular outcome occurred and, for consumer-facing AI Systems applications, trace that outcome to a particular characteristic of the consumer or data associated with the consumer and aspect or component of the algorithm or application. You must be able to permit the Department or a consumer to identify the specific information that caused the consumer's outcome, allow the Department or the consumer to correct false or incorrect information and have the outcome reviewed in light of the corrected data.

Disputability includes some degree of transparency, but is a broader requirement that simply explaining how a model or algorithm works or is intended to work. The Department recognizes that with some AI techniques, you may not be able to understand how the algorithm was created because the AI application may learn and change without human intervention. Such learning and changes may occur very frequently. The Department will not ask you record every change in such models, but requires that you be able to explain how a particular consumer outcome emerged so that the outcome is disputable.

2. Testing for Unfair Discrimination

All regulated entities are required to demonstrate the absence of unfair discrimination by testing for unfair discrimination on both the actuarial and protected class bases. This bulletin provides minimum standards for such testing, how to respond to testing results showing unfair discrimination and how to document and report the results of testing and testing responses.

Testing is required for protected class characteristics in all consumer-facing AI Systems applications. Testing is also required for certain data types for which discrimination on the basis of that data type is permitted for certain parts of the insurance life cycle, but not for others. For example, age, marital status and gender are data types used for marketing, underwriting and pricing for many types of personal insurance. Claim settlement outcomes for the same type of claims, however, should not vary based on these data types.

While there are a variety of methods and models used by regulated entities to develop algorithms and a variety of ways to test for unfair discrimination, the Department requires that insurers utilize one specific testing methodology to ensure a consistent set of metrics across regulated entities. That required testing methodology is referred to as the Control Factor Approach. If you believe that the Control Factor Approach does not accurately reflect fair and unfair discrimination of your AI Systems application, you may utilize a second methodology and report the testing outcomes of both the Control Factor Approach and your second methodology with an explanation why you believe your second methodology is a better method for assessing fair and unfair discrimination than the Control Factor Approach

The basics of the Control Factor Approach are as follows. Every AI Systems application utilize certain data types as predictors of a particular outcome sought by the insurer. In the development of an AI Systems application, the modeler will examine a variety of data types to see which data types and combinations of data types best predict the outcome sought by the insurer. Some data types are then eliminated because they are not predictive or not sufficiently predictive to include in the algorithm or model ultimately deployed by the insurer.

In developing a model or algorithm, the modeler will often employ one or more control variables – data types utilized as predictors in the model, but not intended to be used once the model is deployed. The purpose of the control variable is statistically remove certain influences that would otherwise skew or statistically bias the model. For example, an insurer developing a multi-state risk classification model for personal auto insurance might include a control variable for state to ensure the model is not biased because of state differences in age distribution or tort systems. By including state as a control variable, the modeler removes the statistical influence of significant state differences on the other predictive variables, leaving the remaining results for the other predictive variables as better estimates of the unique contribution of those predictive variables to the explanation / prediction of the outcome.

For purposes of testing for unfair discrimination, the Control Factor Approach attempts to remove the correlation between predictive variables – and the algorithm as a whole – and the protected class characteristic – thereby ensuring that the predictive variables and the algorithm are predicting the outcome and are not proxies for the protected class characteristic. The Control Factor Approach also improves the assessment for actuarial fairness by removing potentially spurious correlations.

Testing for unfair discrimination using the Control Factor approach should be part of the AI Systems application development as well as used to test the final model intended for deployment and the actual consumer outcomes that result from the deployment of the AI Systems model.

The Department expects regulated entities to document the results of the Control Factor Approach testing and provide the results of the testing to the Department as set out in the Reporting section of this bulletin. If the regulated entity utilizes a second testing methodology and seeks Department consideration of the results of that second testing methodology in place of or in addition to the results of the Control Factor Approach, the Department expects the regulated entity to document and report those results, too.

a. Testing Metrics

The basic method for Control Factor Approach testing is to perform a multi-variate analysis of the model in the following general form

$$b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e = y$$

X_1, X_2 + X_3 are the predictive variables trying to predict y.

b_0 is a constant produced by the analysis

b_1 b_2 and b_3 are the coefficients for the predictive variables – the values that will be assigned to individual consumer data values. These coefficients indicate how much the predictive variable is contributing the outcome result.

Each predictive variable will also have measures of statistical significance, indicating how statistically reliable and powerful is the predictive variable.

e is the residual, reflecting the portion of the outcome not explained by the predictive variables.

In addition to statistical measures for individual predictive factors, there are also statistical measures for the model as a whole.

The Control Factor Approach adds one or more control variables to correspond to protected or prohibited class characteristics – characteristics prohibited generally and characteristics not permitted for the particular AI Systems application.

$$b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \mathbf{b_4C_1} + e = y$$

b. Metrics for Evaluation and Action

Proxy Discrimination: If, after adding control variables for protected class or prohibited characteristics, a particular predictive data type loses 75% of its predictive power – as measured either by the factors coefficient or measure of statistical strength – that data type is considered a proxy for a protected or prohibited class characteristic and may not be used in the deployment version of the model.

Disparate impact: After adding control variables for protected class characteristics, the insurer may find a protected class characteristic is a statistically significant and sizable predictor of the outcome – in addition to other predictive factors being statistically significant and sizable predictors of the outcome. This is not a violation, but the Department expects the regulated entity will explore other predictive variables that achieve a similar predictive outcome sought by the insurer, but with less disparate impact.

Equity Trade Off Metrics: The Department understands that some protected class characteristics are strong predictors of certain insurance outcomes, yet their use – directly or indirectly through proxies – is prohibited regarded of actuarial fairness. This is a public policy that recognizes acceptable trade-offs between actuarial fairness and protected class equity. Consistent with this public policy, the Department expects that if the disparate impact as measured by the contribution of the control variable in the Control Factor Approach can be reduced by 80% or more with no greater a loss of efficiency or predictive power of the AI Systems model of 10%, the insurer will accept that equity trade off and implement that change.

c. Data for Protected Class Testing

The Department recognizes that testing requires assignment of protected class characteristics to the individual transaction data utilized by regulated entities in the development of AI Systems models. Initial testing for protected class unfair discrimination will be limited to race and any other protected class characteristic for which the insurer currently has, is able to obtain or is able to infer that protected class characteristic for the consumer transactions being analyzed. Initial testing will also include testing of data types permitted for some AI Systems applications, but not others including, for example, testing for unfair discrimination on the basis of age, marital status or gender in claim settlement.

The Department recognizes that most insurers do not currently request from applicants, policyholders or claimants their self-identified race. For purposes of testing, the Department expects regulated entities to infer the race of individual consumers utilizing the Bayesian Improved First Name Geocoding (BIFSG) methodology.[4] The Department also encourages insurers to request self-identified protected class characteristics from consumers if such information is provided on a voluntary basis by the consumer. The NAIC has developed best practices for such requests for protected class characteristics. [Insert link for Health Workstream of Special Committee on Race and Insurance]

IX. New Reporting Requirements

1. Data Types, Sources, and Uses

The Department will require insurers and advisory organizations to submit two reports and then update those reports on a quarterly basis for changes. The first report is the Report on Data Types, Sources and Uses and will include the following

- Date of Report
- Data Type – brief description of the data type
- Data Source – Consumer via Application, Consumer via Telematics, Consumer Via Interaction with Insurer, Insurer Internal, Third Party Advisory Organization, Third Party Not Advisory Organization, Public/Government Records
- Name of Third Party Provider, if applicable
- If Third Party Provider, Fair Credit Reporting Act Compliant? – Yes/No
- Use Category – Marketing, Underwriting (Eligibility/Terms), Pricing, Claims Settlement, Antifraud, Risk Prevention, Loss Mitigation, Consumer Relations, Consumer Information, Other
- Models Utilizing These Data – Which of the insurer's' models utilize this data type

The second report is the Report of Algorithms and Models and will include the following:

- Date of Report
- Name of Model or Algorithm
- Internally Developed, Third Party Advisory Organization or Third Party Not Advisory Organization Algorithm
- If Third Party, Name of Vendor
- Date First Deployed
- Date Deployment Ended, if applicable

[4] See https://www.rand.org/pubs/research_reports/RRA1853-1.html and https://www.paceanalyticsllc.com/post/cfpb-bifsg-proxy

- Date Current Version Deployed
- Current Version Number
- Purpose(s) of Algorithm – Product Development, Marketing, Underwriting (Eligibility/Terms), Pricing, Claims Settlement, Antifraud, Risk Prevention, Loss Mitigation, Consumer Relations, Consumer Information, Other)
- If Third Party Provider, Fair Credit Reporting Act Compliant? – Yes/No
- For Third Party Not Advisory Organization Algorithms, List Data Types Used in the Algorithm.

2. Testing Results

The regulated entity will report on a quarterly basis any new results of testing for unfair discrimination, including testing results for any AI Systems applications developed during the reporting quarter. The initial report (pursuant to phase-in explained below) will report the results of testing of AI Systems applications developed prior to the reporting quarter. Testing results will include pre-deployment and post deployment testing.

For pre-deployment testing, the testing results shall include a description, including quantification, of changes in algorithmic performance and individual predictive variable performance after the protected class Control Variable is added. Pre-deployment testing results shall also include any changes to the algorithm made by the regulated entity in response to test results.

For post-deployment testing, the testing results shall include a description of how the actual consumer outcomes resulting from the model compare to the expected and intended results at the time of initial deployment. The testing results shall also include the actual protected class impacts of the deployed model's actual consumer outcomes.

3. New Statistical Agent and Statistical Plan for Reporting of Granular Outcome Data / Elimination of Market Conduct Annual Statement

The Department intends to solicit interest from a vendor to serve as the Department's statistical agent for major lines of insurance pursuant to a transaction detail statistical plan with quarterly reporting of consumer outcomes by insurers to the statistical agent. The statistical plan will include reporting of final quotes information as well as other sales and policy information and claims information transactions. With these data, over time, the Department will be able to independently test for unfair discrimination at both industry and individual insurer levels as well as better monitor the marketplace for emerging issues, such as changes in availability and affordability of insurance in the face of climate-related catastrophes. Once the new statistical agents and statistical plans are in place and sufficient data have been collected, the Department will eliminate reporting of the Market Conduct Annual Statement.

XI. Phased Implementation

The Department recognizes the need to phase in the testing and reporting requirements. The following is a time-table for initial reporting of testing results for specific protected or prohibited class characteristics and phases of the insurance life cycle. Subsequent reporting shall be according to the instructions in the prior Reporting section.

Phase 1: 3 Months after Publication of This Bulletin

- Initial Report of Data Types, Sources and Uses
- Initial Report of Models and Algorithms

Phase 2: 6 months after Publication of This Bulletin

- Testing for racial bias in antifraud applications, including applications that identify a claim or claimant as suspicious or requiring additional investigation. Reporting of testing results.
- Testing for unclear, misleading, confusing or deceptive language in policy forms developed via an AI System.

Phase 3: 12 months After Publication of This Bulletin

- Testing for racial bias and prohibited characteristics in claim settlement applications. Reporting of testing results.
- Testing for racial bias in underwriting and pricing applications. Reporting of testing results.

Phase 4: 18 months After Publication of This Bulletin

- Testing of racial bias in marketing, customer relations, customer information, loss prevention and risk mitigation. Reporting of test results.

XII. Advisory Organization and Other Third Party Providers of Data and Algorithms

If a third party providing an AI Systems application is licensed as an advisory organization, the Department has some oversight of that organization and the collective decision-making aspects of the development and deployment of that organization's AI Systems algorithm. Among other things, the advisory organization must file its pricing algorithm with the Department for review and approval. This regulatory approach not only ensures avoidance of potential antitrust violations, but creates great efficiencies for insurers and the Department. In the absence of an approved advisory organization filing for a particular AI Systems application, the insurer is responsible for demonstrating to the Department that the AI Systems application complies with laws and regulations, particularly compliance with unfair discrimination laws and regulation.

By licensing itself as an advisory organization and filing its algorithms with the Department, several efficiencies are generated. First, the insurer can rely on an advisory organization's approved algorithm. Second, the third party providing the algorithm does not have to provide information sought by the Department every time a different insurer wants to use the algorithm. Third, a single review by the Department is more efficient that reviewing the algorithm each time an insurer seeks to rely on that third party algorithm.

Based on the above, the Department encourages insurers to encourage their third party providers of AI Systems applications to become licensed as advisory organizations.