Draft: 7/5/22

**Home AI/ML Survey Definitions**

**Content:**

1. **Artificial Intelligence/Machine Learning (AI/ML) Definition:** Definition of AI/ML for the purpose of this survey. (Pages 1-2)
2. **Regression/Static/Pre-2000 Models Definition:** The questions about regression/static/pre-2000 models are on the bottom of the General, Governance, and Insurer Operations tabs. This section defines which models to evaluate to answer these questions. (Page 2)
3. **Governance** – Questions about governance are contained on the “Governance” tab in the survey. (Pages 2-3)
4. **AI/ML Use Descriptions and/or Explanations** – The first section/rows on each of the company operations’ pages (rating, underwriting, etc.) are questions about AI/ML “Uses.” This section defines or explains the uses, **levels of AI/ML deployment, and levels of decisions influenced by AI/ML.** (Pages 4-10)
5. **AI/ML Model Category Types –** When listing each model under the appropriate insurer operation use, columns to the right ask for categorization of the model types. This section defines the columns. All appropriate columns should be selected for each model. (Pages 10-13)
6. **Data Elements Definitions**: The data elements are located in columns at the bottom of each of the company operations’ (rating, underwriting, etc.) pages. This section defines or explains the data elements. (Page 13-14)
7. **Artificial Intelligence/Machine Learning (AI/ML)**

AI/ML describes an automated process in which a system begins recognizing patterns without being specifically programmed to achieve a pre-determined result.  This is different from a standard algorithm in that an algorithm is a process or set of rules executed to solve an equation or problem in a pre-determined fashion.  Evolving algorithms are considered a subset of AI/ML.

**Artificial Intelligence / Machine Learning Systems include:**

* Systems that adapt and adjust to new data and experience without manual human intervention.
* Systems that arrive at results for which the outcomes and the stepwise approach toward the outcomes were not configured in advance by a human programmer.
* Systems that dynamically respond to conditions in the external environment without the specific nature of such responses being known in advance to the designers of the systems.
* Systems that utilize neural networks or deep-learning algorithms under a supervised, semi-supervised, unsupervised, or reinforcement-learning style. These learning styles are also applied to other machine learning techniques.
* Systems that engage in automatic speech recognition, facial recognition, image recognition, text recognition, natural language processing, generation of customer-specific recommendations, automated customer communications (e.g., chatbots with non-preprogrammed prompts), autonomous or semi-autonomous vehicle operation or data gathering, or any other approach that does not require either preprogramming or a manual human intervention in every instance of an action or decision.
* Systems that automatically generate adaptive responses based on interactions with a consumer or third party.
* Systems that determine which data elements to rely upon, in a non-preprogrammed fashion, among a variety of possible alternatives.

**Artificial Intelligence / Machine Learning Systems are not:**

1. Static “scorecards” that deterministically map consumer or other risk characteristics to treatments or decisions. (However, an AI/ML system may use the output of such static “scorecards” as input data for the AI/ML system to consider.)
2. Systems with solely preprogrammed decision rules (e.g., “If A, then B” applied invariably in all situations).
3. Tables of point or factor assignments in rating plans.
4. **Static ratemaking and/or predictive-modeling methodologies, including linear regression, generalized linear modeling (GLM), or generalized additive modeling (GAM).**
5. Purely informational static databases, such as databases used to obtain reference amounts for claim settlements, or static databases pertaining to consumer characteristics or experience, regardless of the amount of information in the database.  However, if AI/ML is used to create a static predictive model, that AI/ML system is considered within the scope of this survey.
6. Deterministic “phone trees” that navigate consumers through pre-recorded voice prompts.
7. Any approach that an insurer could have realistically utilized in the year 2000 or prior.

**2. Regression, Static, or Pre-2000 Models** – For those questions asking about “regression, static, or pre-2000 models,” answer the questions for models excluded from the AI/ML definition for this survey in reasons #4 and #6. Do not report any models that are already reported under the AI/ML definition (e.g., a model that uses both GLM and a Neural Network). The regression, static, or pre-2000 models are defined to be excluded from the AI/ML definition for this survey.

# 3. Governance – Model Governance Questions

The purpose of the questions related to model governance is to obtain a better understanding of a company’s awareness of specific risk areas tied to the NAIC Artificial Intelligence Principles. In addition, the survey seeks information to understand if guidelines and/or best practices are documented. Specifically, if the company is involved in using AI/ML models, **does the company have a documented process in place** that addresses:

* Fairness and Ethics Considerations: Ensuring responsible adherence to fairness and ethical considerations. It is clear there is debate regarding the definition of “fairness and ethics”, so for the purposes of this survey, and assuming a general understanding of the terms, the response should be consistent with how the company defines those terms. Common principles that fall under this category include Transparency, Justice and Fairness, Non-Maleficence, and Responsibility and Privacy. Generally, respect the rule of law and implement trustworthy solutions designed to benefit consumers in a manner that avoids harmful or unintended consequences including unfair or proxy discrimination.
* Accountability for Data Algorithms' Compliance with Laws as well as Intended and Unintended Impacts: Ensuring the data used and the algorithms/models within the scope of the AI/ML system are delivering the intended benefit, and there are proactive processes in place to ensure there is no unacceptable unintended impact. Simply put, be responsible for the creation, implementation, and impacts of any AI system.
* Appropriate Resources and Knowledge Involved to Ensure Compliance with Laws Including those Related to Unfair Discrimination.
* Ensure Transparency with Appropriate Disclosures Including Notice to Consumers Specific to Data Being Used and Methods for Appeal and Recourse Related to Inaccurate Data: Ensuring documented processes and best practices are in place that govern and actively address the issue

of transparency, ensuring adequate and complete/understandable consumer disclosure regarding the data being used and how the data are used, as well as providing a way for consumers to appeal or correct inaccurate data. This is intended to be specific for data not already protected by legislation such as the Fair Credit Reporting Act (FCRA), as the assumption is all companies would be compliant with that law. This pertains to consumer data NOT specified in the FCRA.

* AI Systems are Secure, Safe, and Robust including Decision Traceability and Security and Privacy Risk Protections: Ensuring an appropriate governance process is in place and documented specific to the company’s AI/ML activity or program that focuses on protecting security, in terms of its data and intellectual property, from potentially compromising interference or risk and relevant and necessary privacy protections are in place; and ensuring the data and the AI/ML models are Ensuring the requisite and appropriate resources, skillsets and knowledge needed to ensure compliance with laws, including those related to unfair discrimination, are actively involved in these programs and decision-making – including oversight of third parties understanding and competence related to compliance with relevant laws and the issue of unfair discrimination. Sufficiently transparent and explainable so that they can be reviewed for compliance with laws and best practices and proven to not be unfairly discriminatory or used for an unethical purpose.

It is understood that governance models vary in terms of components and terms used to describe these risk areas. However, there is a common thread across most governance models, and this language was specifically used in this survey as it ties directly to the NAIC’s adopted AI Principles. Where there may be concerns about overlap, the intention is for this additional information to clarify the unique intent of each. The company should reply to each component as specifically as possible.

Governance, for purpose of this survey, includes both controls within the data science group as well as controls at the higher level of Enterprise Risk Management (ERM). Governance should include situations where 3rd parties are used (e.g., audits).

**4. AI/ML Use Descriptions and/or Explanations**

(in order of appearance on the Excel survey)

* **Rating:** AI/ML Uses
  + **Rating Class Determination**: Decisions regarding which insureds to place within which rating category and which criteria to use to establish a given rating category.
  + **Price Optimization**: As defined in the NAIC Casualty and Actuarial Statistical (C) Task Force white paper: <https://www.naic.org/documents/committees_c_catf_related_price_optimization_white_paper.pdf>
  + **Retention Modeling**: Estimation of the effects of a particular insurer-initiated rate change on the decisions of existing insureds to remain with the insurer.
  + **Numerical Relativity Determination**: Decisions regarding which quantitative rating factor to assign to a particular rating category.
* **Underwriting:** AI/ML Uses
  + **Automated Approval**: Approving an application without human intervention on that particular application.
  + **Automated Denial**: Denying an application without human intervention on that particular application.
  + **Underwriting Tier Determination**: Decisions regarding the criteria to use to establish specific named or numbered categories (called tiers) which utilize combinations of attributes that affect an insurer’s underwriting decision.
  + **Company Placement**: Decisions regarding which of several affiliated companies within an insurance group will accept an individual risk.
  + **Input into Non-Automated Approval Decision**: Providing data, analysis, or recommendations regarding a decision to approve an application in a situation where a human decision-maker still has the ability and responsibility to affirmatively consider this information and make a decision independently of the AI/ML system. In this situation, the AI/ML system cannot automatically approve the application, and protocols exist that ensure that each recommendation from the AI/ML system is actively reviewed and not adopted by default.
  + **Input into Non-Automated Denial Decision**: Providing data, analysis, or recommendations regarding a decision to deny an application in a situation where a human decision-maker still has the ability and responsibility to affirmatively consider this information and make a decision independently of the AI/ML system. In this situation, the AI/ML system cannot automatically deny the application, and protocols exist that ensure that each recommendation from the AI/ML system is actively reviewed and not adopted by default.
  + **Automate Processing Thru the Agency Channel**: Enabling agencies to receive certain information about applicants automatically without specifically requesting that information and/or to provide quotes to the applicants and/or recommend a decision regarding the application to the agent without being based on preprogrammed decision rules.
  + **Renewals and Reinstatements**: Determination of which policies would be eligible for renewal, renewal under certain conditions, or reinstatement after a lapse. This also includes determination of which properties to inspect at renewal in order to assess underwriting eligibility and/or the presence of any hazards that would be taken into account in the renewal underwriting process.
  + **Verification of Policy Characteristics**: Evaluation of whether the attributes of the customer or risk provided at the time of the application or at the time of a request for a policy modification are accurate and supported by additional data or likely to be true based on any other set of considerations used by the AI/ML system.
  + **Policy Anomaly Detection**: Identification of any features of a particular policy or risk that are atypical for the policy or risk of that general type and that may be considered by the insurer as deserving additional review and/or scrutiny.
  + **Down-Payment Requirements**: Determination of which payment plans a given insured would be eligible for and the amount(s) of minimum down payment(s) that a particular insured would be required to pay.
  + **Motor Vehicle Record (MVR) Reordering**: Determination of which policies should be subject to a reorder of the insured’s Motor Vehicle Record in order for the insurer to verify information about recent driving history, including chargeable accidents, violations, and any other attributes of driving history considered by the insurer in rating.
* **Claims:** AI/ML Uses
  + **Claim Approval**: Approving a claim without human intervention on that particular claim.
  + **Claim Denial**: Denying a claim without human intervention on that particular claim.
  + **Determine Settlement Amount**: Recommending which amount to offer to a claimant in order to resolve the insurer’s obligations on the claim.
  + **Claim Assignment Decisions**: Recommending which adjusters are assigned to which claims.
  + **Informational Resource for Adjusters**: Providing facts, data, and analysis to claim adjusters without recommending a decision or limiting the adjusters’ authority over handling the claim.
  + **Evaluation of Images of the Loss**: Analysis of photographic, video, or other visual evidence pertaining to a potentially insured loss in order to extract facts relevant to an insurer’s decision and/or provide guidance and recommendations based on the information obtained in this manner.
  + **Subrogation:** Identification of which claims have the potential for the insurer to recover amounts from the responsible party or parties or other sources of recovery, and/or determination of the potential recoverable amounts.
  + **Claims Triage:** Determination of which claims to route through which of the insurer’s internal processes, potentially including which claims could be fast-tracked, which claims to assign to which adjusters, and which claims would require more detailed review and/or scrutiny.
  + **Speech Analysis:** Analysis of spoken communications from the claimant(s) and/or insured(s) with an attempt to derive potentially relevant or predictive insights regarding the nature, circumstances, and possible outcomes of a claim.
  + **Litigation Likelihood:** Determination of which claims are more likely to result in legal action involving the insurer and any of the parties involved in such claims.
* **Fraud Detection:** AI/ML Uses
  + **Fast Tracking of Likely Non-Fraudulent Claims**: For claims that are identified to be at a low risk of fraud, establishing a rapid process for approving and paying those claims without further scrutiny or follow-up with the claimant.
  + **Referral of Claims for Further Investigation**: For claims that are identified to be at a higher risk of fraud or other potential issues that affect the legitimacy of those claims, determining that those claims should be assigned to investigators for a more intensive and human-driven review process.
  + **Detect Medical Provider Fraud**: Identification of claims where medical providers may have submitted inappropriate or questionable amounts for reimbursement.
  + **Detect First-Party Liability**: Identification of potential situations where a first-party insured may have been at fault for a claim and/or may have misrepresented information to the insurer.
  + **Detect Third-Party Liability**: Identification of potential situations where a third-party claimant may have been at fault for a claim and/or may have misrepresented information to the insurer.
  + **Fraudulent Quote Detection:** Identification of which quote requests from consumers are more likely to be based on intentionally false, inaccurate, and/or misleading information.
  + **Organized Crime Rings Identification:** Evaluation of circumstances and conditions of a policy and/or a claim which may indicate some presence of criminal activity orchestrated with the purpose of obtaining illegitimate proceeds from insurers.
  + **Social Network Analysis:** Evaluation of a claimant’s or insured’s behavior on various social-media platforms in an attempt to discern signs of potential fraud or material misrepresentation.
  + **Facial Recognition and Behavior Models**: Evaluation of a claimant’s or insured’s facial features, video-recorded excerpts, or other actions displayed by the claimant or insured in an attempt to discern signs of potential fraud or material misrepresentation.
  + **Evaluation of Potential for Intentional Infliction of Damage:** Identification of circumstances in which it is likely that an insured may have intentionally damaged the covered property and/or may have strong incentives to do so.

* **Marketing:** AI/ML Uses
  + **Targeted Online Advertising**: Determination of which individuals on the Internet should receive or see which advertisements from the insurer.
  + **Identification of Recipients of Mail or Phone Advertising**: Determination of which individuals would be desirable recipients or an insurer’s advertisements via the telephone or physical mail.
  + **Provision of Offers to Existing Customers**: Determination of which customers should be notified of new insurance products, discounts, options to be written in a different book of business, or any other benefit or favorable treatment that the insurer seeks to extend.
  + **Identification of Potential Customer Groups**: Determination regarding which consumer sub-populations could become additional likely customers of the insurer and/or benefit from the insurer’s products and services.
  + **Demand Modeling**: Identification of consumers’ needs for and interest in specific types of insurance and insurance products that the insurer is offering or whose development or sale the insurer may be considering or exploring.
  + **Direct Online Sales**: Selling insurance policies to consumers through a direct Internet-based channel in a manner that does not rely solely on preprogrammed decision rules.
  + **Customer Interactions Using Natural Language Processing (NLP):** Providing services or recommendations to potential applicants and/or current insureds through interactions that recognize the everyday text and/or speech that such potential applicants and/or current insureds utilize in their search for insurance products or attempts to receive customer service.
  + **Media Mix Modeling**: Analysis of the impact of an insurer’s marketing and advertising campaigns by marketing channel to determine how various elements contribute to the insurer’s goal.
  + **Customer Acquisition and Retention:** Analysis of which marketing approaches would be most successful in attracting new customers and retaining existing customers.
  + **Click Analysis on Third-Party Sites:** Consideration of how customers interact with websites that are unaffiliated with the insurer but which may serve as marketing channels for the insurer – for instance, through the insurer’s ads placed on the unaffiliated website, by means of which a potential customer could access a quote or other information on the insurer’s own website.
* **Loss Prevention:** AI/ML Uses: With advances in technology, the loss-prevention function is likely to grow in importance as insurers strive to improve their results not only through efforts at risk selection and matching the rate to the risk, but also through reducing the probability that the underlying insured losses would occur in the first place. Possible uses of artificial intelligence in loss prevention for private passenger automobile insurance may include, but are not limited to, the following:
  + **Identification of High-Risk Customers**: The goal of such identification in a loss-prevention context is not to make an underwriting or rating decision, but rather to recognize which specific customers may benefit most from loss-prevention advice and mitigation techniques that the insurer may be able to provide, thereby reducing such customers’ frequency and/or severity of losses. For example, an AI/ML system might determine that certain households with youthful drivers are more likely to benefit from risk-mitigation advice and other approaches.
  + **Risk-Mitigation Advice to Consumers**: Artificial intelligence systems might be used to target messaging to consumers based on specific risks identified for a given policy. For example, in a household with youthful drivers, AI/ML-targeted messaging and incentives could focus on ways those drivers could gain experience in a low-risk manner and drive more carefully in day-to-day context. For households in mountainous areas, AI/ML systems could provide targeted advice about safe driving in rugged terrain.
  + **Determination of Advance Payments**: In many situations, small payments issued at or shortly after the time of loss, prior to the full adjustment of the claim, can help the insured or third-party claimant prevent much larger amounts of damage that would otherwise greatly raise the costs of the claim for the insurer. In a home insurance context, examples could include, but are not limited to, (i) making a payment for minor repairs that prevent further damage and/or enable the insured to continue residing in the damaged home, instead of needing to seek alternative living arrangements; (ii) making a payment for prompt, inexpensive medical treatment of a third-party claimant under the Liability and/or Medical Payments coverages of a home insurance policy, which could prevent the emergence of a longer-term, chronic, and much more costly health condition; or (iii) making a payment for expenses related to rebuilding a home even in advance of repairs beginning, with the expectation that such a payment will enable repairs to proceed more rapidly and effectively, with fewer delays related to payment issues with the contractor(s) performing the repairs.
  + **Guidance for Loss-Control Inspections:** Providing recommendations regarding which risks should receive an inspection to identify and/or reduce the probability and/or severity of potential insured losses. This may also include recommendations on which aspects of an insured risk an inspection should focus on.

**On each company operations tab (e.g., rating, underwriting) there is the following question: “If yes, what is the current level of AI/ML Deployment? (Select the highest level of deployment).”** Two of the options for answers are “Proof of Concept (POC)” and “Prototype.” The difference between a Proof of Concept (POC) and a Prototype is discussed below.

* **Proof of Concept (POC):** A small exercise to test the design idea or assumption. The main purpose of developing a POC is to demonstrate the functionality and to verify a certain concept or theory that can be achieved in development.
* **Prototype**: Prototyping is a valuable exercise that allows the innovator to visualize how the product will function. A prototype is a working interactive model of the end-product that gives an idea of the design, navigation, and layout.
* **Difference between POC and Prototype**: While a POC shows that a product or feature can be developed, a prototype shows how it will be developed.

**On each company operations tab (e.g., rating, underwriting), there is the following question: “Indicate the Level of Decisions Influenced by AI/ML (anticipated or already implemented).”** The following are the potential answers to this question:

* + **Automation:** Model requires no human intervention on execution.
  + **Augmentation:** Model advises the human who makes a decision. The model suggests an answer.
  + **Support:** Model provides information but does not suggest a decision or action.

**5. AI/ML Model Category Types**

For each of the AI/ML operational areas, there is a question asking the respondent to list each AI/ML model used. These AI/ML models should meet the definition of AI/ML provided for this survey. To the right of each model name are columns to document the type of ML being employed and/or whether an AI or another model is being employed. Mark all that apply for the named model.

When selecting an appropriate category(ies) to describe a model, use the taxonomy provided below to determine which category(ies) applies. If the method being used is not specifically listed in the taxonomy, use expert judgment to select the best category(ies). If no category applies, enter your method in the “Other” column. You may enter more than one method in that column.

The column headings are defined as follows:

1. DL – Deep Learning

* + Deep Boltzmann Machine (DBM)
  + Deep Belief Network (DBN)
  + Convolutional Neural Network (CNN)
  + Stacked Auto-Encoder

2. ENS – Ensemble

* + Random Forest
  + Gradient Boosting Machine (GBM)
  + Bootstrapped Aggregation (Bagging)
  + AdaBoost
  + Stacked Generalization (Blending)
  + Gradient Boosted Regression Trees

3. NN – Neural Network –

* + Radial Basis Function Network (RBFN)
  + Perceptron
  + Back-propagation
  + Hopfield Network

4. REG – Regularization

* + Ridge Regression
  + Least Absolute Shrinkage and Selection Operator (LASSO)
  + Elastic Net
  + Least Angle Regression (LARS)

5. RS – Rule System

* + Cubist
  + One Rule (OneR)
  + Zero Rule (ZeroR)
  + Repeated Incremental Pruning to Produce Error Reduction (RIPPER)

6. RGS – Regression (Note: Only applies if used in conjunction with a method defined as “AI/ML” for purposes of this survey.)

* + Linear Regression
  + Ordinary Least Squares Regression (OLSR)
  + Stepwise Regression
  + Multivariate Adaptive Regression Splines (MARS)
  + Locally Estimated Scatterplot Smoothing (LOESS) Logistic Regression

7. BAY – Bayesian Methods

* + Naïve Bayes
  + Averaged One-Dependence Estimators (AODE)
  + Bayesian Belief Network (BBN)
  + Gaussian Naïve Bayes
  + Multinomial Naïve Bayes
  + Bayesian Network (BN)

8. DT – Decision Trees

* + Classification and Regression Tree (CART)
  + Iterative Dichotomiser (ID3)
  + C4.5
  + C5.0
  + Chi-square Automatic Interaction Detection (CHAID)
  + Decision Stump
  + Conditional Decision Trees
  + M5

9. DR – Dimensionality Reduction

* + Principal Component Analysis (PCA)
  + Partial Least Square Regression (PLSR)
  + Sammon Mapping
  + Multidimensional Scaling (MDS)
  + Project Pursuit
  + Principal Component Regression (PCR)
  + Partial Least Squares Discriminant Analysis
  + Mixture Discriminant Analysis (MDA)
  + Quadratic Discriminant Analysis (QDA)
  + Regularized Discriminant Analysis (RDA)
  + Flexible Discriminate Analysis (FDA)
  + Linear Discriminant Analysis (LDA)

10. IB – Instance-Based

* + k-Nearest Neighbor (KNN)
  + Learning Vector Quantization (LVQ)
  + Self-Organizing Map (SOM)
  + Locally Weighted Learning (LWL)

11. CLU – Clustering

* + k-Means
  + k-Medians
  + Expectation Maximization
  + Hierarchical Clustering

12. AI – AI that is not categorized as ML

13. Any Other that meets the definition of AI/ML selected for this survey.

**Note:** Please make sure that any model supplied by an external vendor is also appropriately identified as one or more of the model category types above.

**6. Data Elements Definitions**

The data elements are located in columns at the bottom of each of the company operations’ (rating, underwriting, etc.) pages.

1. **Consumer or Other Type of “Score”:** A numeric value generated based on a combination of any underlying attributes or behaviors of the consumer, insured risk, or any items considered by the insurer to be relevant to the consumer or insured risk. Scores are computed using deterministic algorithms or models which are not themselves considered to be AI / ML systems. Inquiries in this survey regarding such scores seek to understand whether these scores are used as input data elements within AI / ML systems.
2. **Geodemographic Data (Non-Insurance Statistics by ZIP Code, Census Block, etc.)**
3. **Education**
4. **Facial Detection / Recognition / Analysis**: A picture to confirm identity, estimate biological age or gender of the consumer
5. **Geocoding:** Latitude and longitude coordinates of a physical address
6. **Topography (Land Slope, Elevation, etc.)**:
7. **Historical Weather Information (Temperature, Precipitation, etc.)**
8. **Hurricane Model Output (AAL, PML, etc.)**
9. **Excess Wind/Hail Model Output (AAL, PML, etc.)**
10. **Flood Model Output (AAL, PML, etc.)**
11. **Earthquake Model Output (AAL, PML, etc.)**
12. **Wildfire Wind/Hail Model Output (AAL, PML, etc.)**
13. **Job Stability:** Current employment, length of employment at prior employers, unemployment
14. **Income:** Annual income, income source
15. **Occupation**
16. **Personal Financial Information:** Net worth, type of bank account or credit account, number of bank accounts or credit accounts, available credit, payment history data
17. **Insured Claim Experience--Home**
18. **Insured Claim Experience--Auto**
19. **Industry Territorial Loss Statistics**
20. **Territorial Crime Rates**
21. **Territorial Tax Rates**
22. **Medical**: Medical history, medical condition, prescription data, lab data
23. **Online Media:** Web searches, online purchases, social media activities
24. **Smart Home Devices**
25. **Security Systems**
26. **Roof Data**
27. **Defect Identification in Images (Inherent Risk in the Property)**
28. **Hazard Detection in Images (Risk Due to Surrounding Area)**
29. **Potential Loss Estimates in Images (When Writing the Policy)**
30. **Claims Estimates in Images (When Settling or Adjusting a Claim)**
31. **Other**