NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

MEMORANDUM

To: Superintendent Elizabeth Kelleher Dwyer,
Chair of the Big Data and Artificial Intelligence (H) Working Group

From: Commissioner Kevin Gaffney, Chair of Workstream One (Surveys) of the Big Data and Artificial Intelligence (H) Working Group

Cc: Ten-State Subject Matter Expert Group; Kris DeFrain (NAIC)

Date: August 10, 2023

Re: 2022-23 Home Artificial Intelligence/Machine Learning Survey Analysis

---

The 2022-23 Home Artificial Intelligence/Machine Learning Survey (Home AI/ML Survey)[1] was conducted to inform the work of the Big Data and Artificial Intelligence (H) Working Group in support of its charge to:

> *Research the use of big data and artificial intelligence (AI) in the business of insurance, and evaluate existing regulatory frameworks for overseeing and monitoring their use. Present findings and recommended next steps, if any, to the Innovation and Technology (EX) Task Force, which may include model governance for the use of big data and AI for the insurance industry.*

The survey was conducted under the market examination authorities of 10 requesting states (Connecticut, Illinois, Iowa, Louisiana, Nevada, North Dakota, Pennsylvania, Rhode Island, Vermont, and Wisconsin) and completed by insurers who write Home insurance in one of the ten participating states and have at least $50 million in national Homeowners insurance premium for 2020. The following subject matter experts (SMEs) represented the ten states:

CT: George Bradner
IL: Erica Weyhenmeyer and CJ Metcalf
IA: Jared Kirby
LA: Nichole Torblaa and Arthur Schwartz

[1] The 2022-2023 Home AI/ML survey was conducted under the market conduct examination authority of ten states: Connecticut, Illinois, Iowa, Louisiana, Nevada, North Dakota, Pennsylvania, Rhode Island, Vermont, and Wisconsin. Subject matter experts (SMEs) from these states opted to limit the survey request to only larger companies, defined as those Home (NAIC Type of Insurance--TOI 4.0) writers with more than $50 million in 2020 direct premium written. Type of Insurance (TOI) 4.0 includes policies for residential owners of stand-alone homes, tenants, and condominium unit-owners. The SMEs also limited the scope to only “advanced” AI/ML models but still asked companies to estimate the number of excluded models such as generalized linear models (GLMs). A total of 194 responses were received.

ND: Mike Andring and Chris Aufenthie
NV: Gennady Stolyarov
PA: Michael McKenney
RI: Matt Gendron
VT: Commissioner Kevin Gaffney
WI: Timothy Cornelius

This memorandum contains the SMEs' summary of the survey analysis, key takeaways, and some recommendations for next steps. The SMEs also approved public distribution of the attached NAIC staff's survey analysis, which provides more detail about the survey results.

SURVEY ANALYSIS SUMMARY

"More Advanced" Artificial Intelligence/Machine Learning Model Use by Companies

The survey was intentionally limited to the reporting of "more advanced" types of AI models, so the data should be interpreted as applying to that subset of insurers' predictive models. Out of 194 companies completing the survey, 136 companies currently use, plan to use, or plan to explore using AI/machine learning (ML) as defined for this survey. This equates to approximately 70% of reporting companies. For comparison, approximately 88% of the companies responding to the PPA Survey reported they currently use, plan to use, or plan to explore using AI/machine learning (ML).

Among insurer operations areas, companies reported varying levels of AI/ML use, from 14% in the loss prevention area to 54% in claims operations[2]. In order from maximum to minimum use, the percentage of companies using AI/ML by insurer function were: claims, 54%; underwriting and marketing (both at 47%); fraud detection, 42%; rating, 35%; and loss prevention, 14%.

The top two most popular reasons reported for not using, not planning to use, and not exploring use of AI/ML were "no compelling business reason", and "waiting for regulatory guidance".

The following shows the predominant insurer uses of AI/ML, the levels of decision-making when there are AI/ML models in place, and how often models are developed in-house or externally for each insurer operation.

MODELS BY INSURER OPERATIONS

Claims Models

Uses: In insurance claims operations, companies reported currently using AI/ML claims models mostly for subrogation and claims triage (44 companies each) and evaluation of images of loss (36). Few companies are using AI/ML claims models to determine the settlement amount (2) and a couple are researching how they might use them for claim approval and claim denial. Companies noted some additional uses of claims models in their write-in comments: identification of contractor outcomes, workforce management, legal bill review or automatic payment decisions, method of inspection, DocBot document classification, customer satisfaction, and NLP analysis of complaint intensity. In comparison, the main uses of AI/ML models for claims reported in the PPA Surveys were for informational resources for adjusters (96 companies), claim assignment decisions (58), evaluation of images to estimate losses (55), and to determine settlement amount (50).

[2] 'Other' areas included: routing or triage of customer service calls, customer service chat for property, social media sentiment analysis, document processing/identification/routing, and chatbots.

Level of decision-making: Over 60% of the AI/ML models used to help determine subrogation and claims triage *augment* human decision-making. In contrast, over 60% of the models used to evaluate images to estimate loss amounts *support* human decision-making.

In-house or third-party: Models were developed in-house for most of the claims applications (subrogation, claims triage, adjuster information, claim assignment, litigation, approval and denial). Models used to evaluate images to determine loss and for speech analysis tend to be developed by third parties.

Types of models: About a third of the types of AI/ML models used for Claims consisted of machine learning ensembles/automated machine learning techniques which are effective to estimate claim amounts and probability of litigation, and another third were neural networks/deep learning which are effective for image and speech analysis.

Fraud Models

Uses (identified by at least 50 companies): In insurance fraud detection, companies reported currently using AI/ML models mostly to refer claims for further investigation. Interestingly, 26 companies are using AI/ML to detect organized crime rings, as compared to very few respondents reporting this use in PPA Survey. Further, 24 companies reported using social network data on the Home Survey, but this was not a listed data source for those responding to the PPA Survey.

Level of decision-making: For the referral of claims for further investigation, most of the models *augmented* human decisions. Other fraud-detection models were a mixture of *augmentation* and *support*. This is consistent with the findings reported in the PPA AI/ML Surveys.

In-house or third-party: Fraud detection models are mostly from third-party sources but for some uses, the models are just as frequently in-house as they are from third parties.

Types of Models: A variety of models were used to help determine Fraud, including deep learning/neural networks, machine learning ensembles and other techniques.

Marketing Models

Uses: Companies use marketing models mainly for customer acquisition and retention, and targeted online advertising.

Level of decision-making: Many of the marketing models were used to *augment* human decision-making. Marketing models are mostly used as *augmentation* when used for customer acquisition and retention and provision of offers to existing customers. *Automated* models, with no human intervention on execution, were most often used for targeted online advertising and customer interactions using NLP. For comparison, the PPA AI/ML Surveys reported a greater degree of *automated* Marketing model implementations than the Home AI/ML Surveys.

In-house or third-party: Marketing models being used by insurance companies combine in-house and third-party sources in all but click analysis on third-party sales. In-house sources are predominately used for customer acquisition and retention, provision of offers to existing customers, and identification of potential customer groups. Third-party sources are prevalent in targeting online advertising, identification of recipients of mail or phone advertising, and media mix modeling.

Types of Models: A majority of the models used for Marketing were machine learning ensembles.

Rating, Underwriting, and Loss-Prevention Models

Similar to the PPA AI/ML Surveys, there were fewer "more advanced" AI/ML models used for rating, underwriting, and loss prevention. Therefore, the data for some detailed questions is less credible. This may, however, simply be a reflection of the limited extent of the deployment of such more advanced models to date due to the transparency required by state departments of insurance. This corroborates the understanding of the SMEs that the majority of rating approaches that Home insurers use today continue to involve more traditional ratemaking techniques and older-generation static predictive models. The more advanced AI/ML models currently constitute a minority of the models used by insurers in rating, underwriting, and loss prevention.

Only 68 companies reported current rating model uses, and the majority of those were for rating class determination, retention modeling, or other rate-related functions. The levels of decision were mixed between *automated* and *support*; and as similarly reported in the PPA Surveys, almost all rating models were developed in-house mainly using machine learning ensembles.

A greater number of companies than reported in the PPA Surveys (91 vs. 34) reported current underwriting model uses, and the majority of those were for automated denial/inputs into non-automated denial decisions, and for verification of policy characteristics. About half of the models were used for *automation*, with a third used for *support* and a sixth of the models were used for *augmentation*. Most of the models used in underwriting were developed in-house, using deep learning and machine learning ensembling techniques, consistent with the PPA Surveys.

Only 28 companies reported current loss-prevention uses (compared to 3 companies using models for loss prevention in the PPA Surveys), mainly for guidance on loss control inspections, and with 12 companies reporting active research in providing risk-mitigation advice, where most of the models were reported used to *augment* or *support* human decisions. Models were split between being developed internally and by third parties. A variety of neural network and other machine learning models were reported being used to develop models for loss prevention.

DATA ELEMENTS BY INSURER OPERATIONS

The following are the most frequent data elements used in the different insurer operations.

- Claims
  - Home loss Experience (internal), Roof data (from external source), and Geocoding (mix of internal and external sources)
- Fraud Detection:
  - Mainly home loss experience (split between internal and externally-sourced)
- Marketing:
  - Insured demographic/Geo-demographic data, Consumer or other type of Score, Occupation, Education, Personal Finance Information, Home loss experience, geo-coding, and Income; these data elements were mixed between internally and externally sourced.
- Rating:
  - Insured demographic/Geo-demographic data, Consumer or other type of Score, Geo-coding, and Home loss experience; these data elements were mixed between internally and externally sourced.

- Underwriting:
    - Roof data, Consumer or other type of Score, Defect identification in images, and Insured demographic data; these data elements were mixed between internally and externally sourced.
- Loss Prevention:
    - Of the data elements listed, very few data elements were reported being used in AI/ML loss prevention models, implying that other data elements were used but not reported in the Surveys.

CUSTOMER DATA CORRECTION

The response rate to the consumer data correction questions in the Home Surveys was low. Less than 10% of the companies responded they provide information regarding the data elements used and the purposes of use to consumers beyond what is required by law. Of these companies, half responded they had more consumer data correction processes than required by the federal Fair Credit Reporting Act (FCRA).

Data Element Information Provided to Consumers
Insurers were asked to identify if they were providing additional information about data elements to consumers *other than what is required by law*. The answer, although the number of reporting companies was lower than expected, was almost unanimously "no" for each of the insurer operations. The second question was similar but asked whether consumers are told the purposes of data elements beyond what is required by law. For this question, the answer was almost unanimously "no." This response was consistent with the PPA Surveys.

Consumer Opportunity to Challenge or Correct Data
For the question on whether consumers have the opportunity to challenge or correct their specific data outside of processes for the federal Fair Credit Reporting Act (FCRA), many did not answer. Of those who answered this question, 60% responded "yes" for rating and underwriting; a little less than half responded "yes" for claims; over 40% responded "yes" for marketing and fraud detection; and 35% responded "yes" for loss prevention.

GOVERNANCE

The purpose of the model governance questions is to obtain a better understanding regarding a company's awareness of specific risk areas tied to selected categories in the NAIC Artificial Intelligence Principles.

Insurers were asked if the following are *documented* in a governance program:

- Fairness and ethics considerations.
- Accountability for data algorithms' compliance with laws, as well as intended and unintended impacts.
- 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.
- AI systems are secure, safe, and robust, including decision traceability and security and privacy risk protections.

Insurers' answers to the fairness and ethics consideration, accountability, transparency, and secure/safe/robust questions were fairly consistent between each question, where approximately 50% of insurers reported "yes" in the Home Surveys, as compared to approximately 25% of insurers reporting "yes" in the PPA Surveys. The answers for the appropriate resources question tended to be higher percentages of "yes" (approximately 90%) than for the other questions.

THIRD-PARTY DATA SOURCES AND MODELS

Insurers identified third-party vendors they use to purchase models and/or data. There were 2,413 models listed in the survey (with some models being counted more than once because of separate uses for the same model); 1,006 (42%) are developed by a third party, and 1,407 (58%) are developed internally. The proportions of models developed internally vs. a third party reported in the PPA Surveys were nearly identical. After grouping the similarly named third parties, there are 95 unique third-party companies listed in the survey whose models are being used by insurers (vs. 76 third parties reported in the PPA Surveys). Marketing has 48 different third parties listed as providing models, followed by claims with 37. There were 76 third-party companies listed as data providers used by insurers (vs. 104 unique third parties reported in the PPA Surveys).

CONCLUSION/NEXT STEPS

The insight gained from the survey will be used to supplement state insurance regulators' knowledge of the current regulatory framework around AI/ML, governance, consumers, and third parties and to evaluate whether any changes should be made to the regulatory frameworks.

Following are some potential next steps, including many activities already in progress. This list is not intended to be complete, but it may be helpful as a starting point for discussions and decision-making about what next steps to take at the NAIC:

- Evaluate the survey analysis and determine whether to further explore the following subjects:
    - Insurer AI/ML model usage and the level of decision-making (i.e., the amount of human involvement in decision-making).
    - Insurer data elements.
    - Insurers' governance frameworks and the documentation of such.
    - Consumer data recourse.
    - Third-party regulatory framework.
- Create a risk hierarchy to prioritize the need for more model governance and insurer oversight. The general concept is that more oversight of a model will be needed as the consumer risk or impact increases from the modeling or models.
- Evaluate consumer data recourse. Insurers report a wide variety of methods for consumers to evaluate and correct data used by insurers. Some methods are short and easy, such as using an app to correct data, and other methods are more time-consuming and require personal contact with the agent or company. In some cases, consumers may not know their data is being used, so consumer transparency is a priority. *(Privacy Protections (D) Working Group)*
- Evaluate the regulatory framework around the use of third-party models and third-party data. Evaluate the ability of insurers and state insurance regulators to obtain needed information from third parties and for regulators to oversee this work either through the insurers or third parties in some way. *(Workstream Two of the Big Data and Artificial Intelligence (H) Working Group)*
- Evaluate concerns about third-party concentration by insurer use. *(Workstream Two of the Big Data and Artificial Intelligence (H) Working Group)*

- Determine whether additional white papers on best practices would be useful on subjects in the AI/ML space.

Additional information was collected but not documented due to the confidential nature. Regulators may contact Kris DeFrain, kdefrain@naic.org to seek additional, but non-company identifying information. This report is confidential because data was collected in a market conduct examination of the ten states and agreed confidentiality protections were applied.

August 10, 2023

Home Insurance
Artificial Intelligence/Machine Learning
Survey Results

NAIC Staff Report

NAIC SURVEY TECHNICAL TEAM

Kris DeFrain, Director of Research and Actuarial Services
Justin Cox, Data Analyst III
Nancy Beydler, Senior Administrative Assistant
Roberto Perez Santiago, Property/Casualty Rate Modeling Actuary and Data Scientist

Table of Contents

Index of Tables

INTRODUCTION

Purpose of the Home AI/ML Survey

At the outset of the Artificial Intelligence (AI)/Machine Learning (ML) surveys, the predecessor to the Big Data and Artificial Intelligence (H) Working Group defined five key objectives. Regulators want to: 1) learn directly from the industry about what is happening in this space; 2) get a sense of the current level of risk and exposure and whether or how the industry is managing or mitigating that risk; 3) develop information for trending, such as how the risk is evolving over time, and the industry's responsive actions; 4) inform a meaningful and useful regulatory approach, framework, and/or strategy for overseeing and monitoring this activity; and 5) learn from prior surveys to inform and improve future surveys.

Goals of the Home Survey
1. Analyze industry use of artificial intelligence (AI)/machine learning (ML). 2. Identify industry's risk and exposure and mitigation of model risk. 3. Calculate trends. 4. Gather background for regulatory approach/framework. 5. Inform/improve future surveys.

This Home survey is expected to help regulators in terms of 1) consumer protections and 2) areas that regulators might expect companies involved in this type of activity to be, actively and with intention, ensuring that they are putting processes and procedures in place to meet, or at least consider, the expectations laid out in the NAIC's AI Principles.

The requesting states agreed the collected data will <u>not</u> be used to evaluate or determine the company's compliance with applicable laws and regulations.

Purpose of This NAIC Staff Report

With the large volume of data submitted for this survey, the subject matter expert (SME) group asked NAIC technical staff to assist in conducting a thorough analysis. NAIC staff were asked to evaluate the results, provide data analysis, and investigate potential inaccuracies in the data. The team was specifically asked to investigate what types of data are being used by companies in their AI/ML models; evaluate third-party AI/ML model and data use; explore levels of governance; and evaluate transparency, consumer disclosures, and potential consumer actions to correct data.

BACKGROUND

The Home AI/ML survey was conducted under market conduct authority of ten states: Connecticut, Illinois, Iowa, Louisiana, Nevada, North Dakota, Pennsylvania, Rhode Island, Vermont, and Wisconsin ("Requesting States"). The Requesting States conducted the survey to:

- Gain a better understanding of the insurance industry's use and governance of big data and AI/ML.
- Seek information that could aid in the development of guidance or a potential regulatory framework to support the insurance industry's use of big data and AI/ML.
- Inform as to the current and planned business practices of the company.

The survey call letter from the Requesting States was distributed on Nov. 1, 2022, to larger home insurance companies with survey responses requested by Dec. 1, 2022. The due date was extended to Dec. 15, 2022, for all companies. The survey officially closed on April 3, 2023. A total of 194 responses were received, and 70% of those indicated they are doing something pertaining to Home insurance AI/ML.

Survey Web Page

The survey template, filing documentation, frequently asked questions (FAQ), definitions, and other information can be found on the **Home AI/ML Survey** web page.

Surveyed Companies

Home insurance was defined as insurance described under the National Association of Insurance Commissioners (NAIC) Type of Insurance (TOI) 4.0. This includes policies for residential owners of stand-alone homes, tenants, and condominium unit-owners.

The Home insurance companies required to respond to the survey were those 1) reporting at least $50 million in national home insurance premium for 2020 and 2) transacting ongoing business in at least one of the Requesting States.

Regulatory Subject Matter Experts

For each of the requesting states, the following SMEs created the survey and will communicate the survey responses to the Big Data and Artificial Intelligence (H) Working Group.

CT: George Bradner
IL: Erica Weyhenmeyer and CJ Metcalf
IA: Jared Kirby
LA: Nichole Torblaa and Arthur Schwartz
ND: Mike Andring and Chris Aufenthie
NV: Gennady Stolyarov
PA: Michael McKenney
RI: Matt Gendron
VT: Commissioner Kevin Gaffney
WI: Timothy Cornelius

The following NAIC staff assisted the SMEs with survey development, survey distribution, and data collection: Tim Mullen, Teresa Cooper, Justin Cox, Sam Kloese, and Kris DeFrain.

Artificial Intelligence/Machine Learning Definition

The definition of AI/ML was provided on the Home AI/ML survey web site with the following link: **Home AI/ML Filing Guidance & Definitions**. The AI/ML definition was written for this survey only and is not intended to be used for other NAIC projects.

> 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:

- 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.)
- Systems with solely preprogrammed decision rules (e.g., "If A, then B" applied invariably in all situations).
- Tables of point or factor assignments in rating plans.
- Static ratemaking and/or predictive-modeling methodologies, including linear regression, generalized linear modeling (GLM), or generalized additive modeling (GAM).
- 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.
- Deterministic "phone trees" that navigate consumers through pre-recorded voice prompts.
- Any approach that an insurer could have realistically utilized in the year 2000 or prior.

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).

A key decision affecting interpretation of results was the definition of AI/ML for purposes of the survey. The SME group drafted the AI/ML definition to exclude some methods, such as linear regression, commonly used models such as GLM and GAMs, and any approach that a company could have realistically used in the year 2000 or prior. The SMEs developed the AI/ML definition to focus on the "more advanced" models.

Regulators noted they have extensive experience reviewing the older models used for rating, having completed the NAIC's 2020 white paper *Regulatory Review of Predictive Models* and having conducted numerous training and educational events.

Unlike the private passenger auto AI/ML survey, regulators asked for estimates of the number of regression, static, or pre-2000 models. Regulators also asked for the types of advanced models in the survey. More information is included in this report.

<u>Confidentiality</u>

The individual company results are confidential. Some combined results will be publicly presented at Big Data and Artificial Intelligence (H) Working Group meetings and are presented in this report.

GENERAL SECTION OF THE SURVEY

Out of 194 companies that completed the survey, 136 companies currently use, plan to use, or plan to explore using AI / ML as defined for this survey. This equates to 70.1% of reporting companies. (Refer to Table 1.)

Table 1: Companies Using or Exploring the Use of AI/ML

Number of Companies Using, Planning to Use, or Exploring Use of AI/ML		
Yes/No?	#	%
Yes	136	70.1
No	58	29.9
Total	194	100.0

The 58 companies that indicated they had no plan to use or explore use of AI/ML also provided their reason(s) why, with the most often selected reason being "no compelling business reason." In addition to the options listed in the survey and shown in Table 2, a few companies wrote in additional reasons such as was not a priority, current systems meet existing needs, and models are used, but the models do not meet the definition of AI/ML in the survey.

Table 2: Companies' Reasons for Not Using AI/ML

If not using, planning to use, or exploring use of AI/ML for any company operation, why?	
Options listed in the survey:	Number of Companies
No compelling business reason	38
Lack of resources and expertise	23
Waiting for regulatory guidance	30
Lack of reliable data and associated security risk	31
Reliance on legacy systems requiring IT (Information Technology), data, and technology system upgrade before starting AI/ML initiatives	29
Waiting on the availability of a third-party vendor product/service	23
Risk not commensurate with current strategy or appetite	25

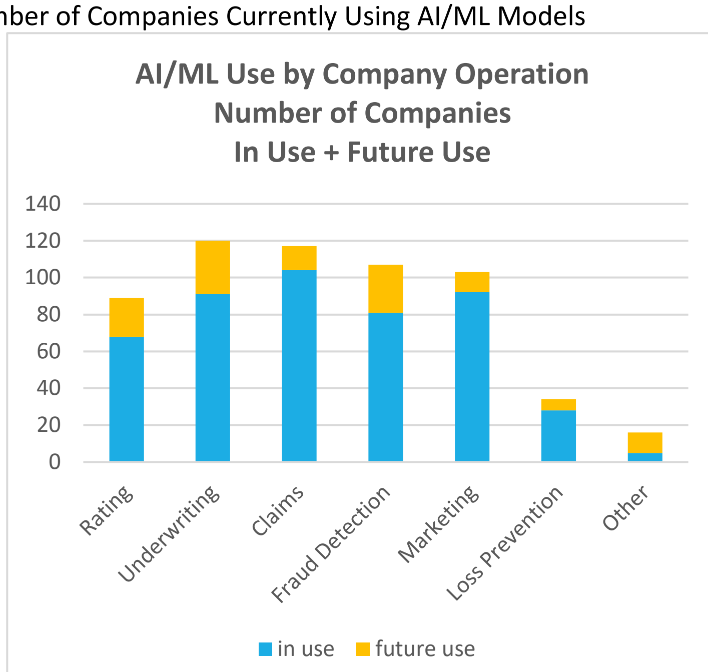
For companies using AI/ML, the use varies by company operation area with expectations the underwriting, fraud detection, and rating will grow the most in the next few years. Current uses are: Claims, 54%; **underwriting, 47%,** marketing, 47%; **fraud detection, 42%; rating, 35%;** loss prevention 14%; and other, 3%. Adding in the companies with models under construction (the combination of those being researched, in proof of concept, and prototypes), the percentages of use in the future would be predicted to be: **Underwriting, 62%;** claims, 61%; **fraud detection, 55%;** marketing, 53%; **rating, 46%;** loss prevention 17%; and other, 9%. Expectations are the usage of AI/ML models in underwriting, fraud detection, and rating will grow more than 10 percentage points over the next few years. (Refer to Table 3.)

Table 3: Companies with Models (in Use or Under Construction) by Company Operation Area

Company Operation Area[1]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Rating	68	35	21	11	105	54	194	100
Underwriting	91	47	29	15	74	38	194	100
Claims	104	54	13	7	77	40	194	100
Fraud Detection	81	42	26	13	87	45	194	100
Marketing	92	47	11	6	91	47	194	100
Loss Prevention	28	14	6	3	160	82	194	100
Other	5	3	11	6	178	92	194	100

The same information is shown pictorially in Figure 1.

Figure 1: Number of Companies Currently Using AI/ML Models



[1] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

In addition to the company operations areas listed in the survey template, companies provided numerous "other" AI/ML uses. The following are additional uses of AI/ML: routing or triage of customer service calls, customer service chat for property, social media sentiment analysis, document processing / identification/ routing, and chatbots.

COMPANY OPERATION: CLAIMS[2]

Out of 194 reporting companies, 104 reported using AI/ML for claims operations and 13 reported having models under construction (i.e., in a research, proof of concept, or prototype stage). (See Table 3.)

Excerpt from Table 3:

Company Operation Area[3]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Claims	104	54	13	7	77	40	194	100

Claims Model Uses

In insurance claims operations, companies reported currently using AI/ML claims models mostly for subrogation (44 companies), claims triage (44), and evaluation of images of loss (36). Few companies are using AI/ML claims models to determine the settlement amount (2) and a couple are researching how they might use them for claim approval and claim denial. The uses of claims models identified in Table 4 were options that could be selected in the survey template. Companies noted some additional uses of claims models in their write-in comments: identification of contractor outcomes, workforce management, legal bill review or automatic payment decisions, method of inspection, DocBot document classification, customer satisfaction, and NLP analysis of complaint intensity.

Once models under construction begin to be used, companies will most often be using AI/ML claims models for subrogation (74 companies), other claim-related functions (63), and as an informational resource for adjusters (60).

Table 4: Companies' Use of Claims Models

Claims Model Uses[4]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None/Null
Subrogation	44	11	--	19	124
Claims Triage	44	3	1	--	146
Evaluation of Images of Loss	36	6	1	--	151
Other Claim-Related Functions	31	11	6	15	131

[2] For definitions, refer to Appendix B: Definitions Specific to Claims.
[3] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
[4] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Claims Model Uses[4]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None/Null
Informational Resource for Adjuster	30	29	1	--	134
Claim Assignment Decisions	28	2	--	--	164
Speech Analysis	14	3	--	8	169
Litigation Likelihood	6	1	1	11	175
Determine Settlement Amount	2	--	--	--	192
Claim Approval	--	1	--	1	192
Claim Denial	--	1	--	--	193

The level of insurance company employee decisions influenced by AI/ML varies by model use. For most uses of claims models, the model is used for support, while subrogation and claims triage are mainly used as augmentation. Models for claim assignment decisions are evenly used as automation and augmentation. (Refer to Table 5.) Note that Table 5 differs from the previous tables because the data represents the number of models instead of the number of companies.

Table 5: Level of Decision-Making by Use of Claims Models

Claims Model Uses[5]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Subrogation	11	54	23	--
Claims Triage	6	47	25	1
Other Claim-Related Functions	9	13	80	--
Evaluation of Images of Loss	11	25	71	13
Informational Resource for Adjuster	18	7	81	--
Claim Assignment Decisions	27	27	--	1
Speech Analysis	3	2	48	3
Litigation Likelihood	--	--	19	--
Determine Settlement Amount	--	5	1	--
Claim Approval	2	--	--	--
Claim Denial	--	1	--	--

*"Automation" was defined as no human intervention on execution. "Augmentation" was defined as a model that suggests an answer and advises the human who is making a decision. "Support" was defined as a model that provides information but does not suggest a decision or action.

Models being used by insurance companies are developed in-house (with or without third-party assistance) or purchased from a third party. Most claims models are developed in-house. Models used to evaluate images of the loss, speech analysis, and determination of claim settlement amounts tend to be developed by third parties. (Refer to Table 6.)

[5] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Table 6: Claims Model Sources by Model Use

Claims Model Uses[6]	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Subrogation	79	90	9	10	88	100
Claims Triage	77	97	2	3	79	100
Other Claim-Related Functions	85	83	17	17	102	100
Evaluation of Images of Loss	31	26	89	74	120	100
Informational Resource for Adjuster	64	60	42	40	106	100
Claim Assignment Decisions	55	100	--	--	55	100
Speech Analysis	18	32	38	68	56	100
Litigation Likelihood	13	68	6	32	19	100
Determine Settlement Amount	--	--	6	100	6	100
Claim Approval	2	100	--	--	2	100
Claim Denial	1	100	--	--	1	100

For each claim model, the type of model by category was identified. When a company used more than one category type for the same model, the model will be counted in all applicable categories. This leads to overcounting the models in total; but provides a good comparison of how often each category type is being used by insurers. This table shows a total of 693 models, whereas the actual number of models is 634. The most frequently used categories of claims models are DL = Deep Learning (137), ENS = Ensemble (227), and NN = Neural Network (86). (Refer to Table 7.)

Table 7: Number of Models by Type of AI/ML Model

Claim Model Uses	Number of Models by Type of AI/ML Model[7]												
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Subrogation	1	48	--	1	1	1	18	13	--	--	1	1	8	93
Claims Triage	19	43	7	--	--	--	9	7	1	--	--	--	2	91
Other Claim-Related Functions	2	59	--	--	8	--	--	5	--	--	--	17	11	102
Evaluation of Images of Loss	56	7	25	2	1	2	--	2	1	--	16	3	8	138
Informational Resource for Adjuster	16	35	15	--	2	--	--	5	--	--	--	12	19	104
Claim Assignment Decisions	--	21	19	--	--	--	--	12	--	--	--	--	--	55
Speech Analysis	33		16	1	1	1	--	--	--	--	--	10	13	75
Litigation Likelihood	4	14	4	--	--	--	--	1	--	--	3	--	--	26
Determine Settlement Amount	--	--	6	--	--	--	--	--	--	--	--	--	--	6
Claim Approval	--	--	--	--	--	--	--	2	--	--	--	--	--	2

6 For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

7 For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Claim Model Uses	Number of Models by Type of AI/ML Model[7]													
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Claim Denial	--	--	--	--	--	--	--	1	--	--	--	--	--	1
TOTAL	137	227	86	4	13	4	27	48	2	--	20	45	59	693

*DL = Deep Learning, ENS = Ensemble, NN = Neural Network, REG = Regularization, RS = Rule System, RGS = Regression, BAY = Bayesian Methods, DT = Decision Trees, DR = Dimensionality Reduction, IB = Instance Based, CLU = Clustering, AI = Artificial Intelligence that is not ML; and AML – Automated Machine Learning (a third-party company provides the answer), TOTAL includes "Other" models.

Data Elements

It is well known that insurers use big data for many purposes and models. Given this survey is focused on the use of the "more advanced" AI/ML, the data element information here is similarly focused on the use of data elements only when used in "more advanced" AI/ML models.

For claims, the following five data elements were the most frequently reported as being used in AI/ML models:

- Insured Claim Experience-Home (48 companies)
- Roof Data (38)
- Geocoding (32)
- Other-Non-Traditional (20)
- Historical Weather Information (15)

There are at least some companies using additional data elements as shown in Table8. (Refer to Table 8.)

Table 8: Companies' Use of Claims Data Elements

Claims Data Elements[8]	Number of Companies Using/Not Using the Data Element in a Claims AI/ML Model*	
	Yes	No
Insured Claim Experience – Home	48	146
Roof Data	38	156
Geocoding	32	162
Other: Non-Traditional	20	175
Historical Weather Information	15	179
Delay in Reporting	11	183
Defect Identification in Images	4	190
Modeled Behavioral Characteristics	3	191
Hazard Detection in Images	3	191
Aerial Imagery	3	191
Loss Description	2	192
Smart Phone Devices	1	193
Security Systems	1	193
Potential Loss Estimates in Images	1	193
Personal Financial Information	1	193

[8] For definitions, refer to Appendix H: Data Use Table Definitions.

Claims Data Elements[8]	Number of Companies Using/Not Using the Data Element in a Claims AI/ML Model*	
	Yes	No
National Recall Database	1	193
Insured Demographic Data	1	193
Geodemographic Data	1	193
Claims Estimates in Images	1	193

*The question is not whether the data element is used, but only whether the data element is used in an AI/ML model.

The data elements used in claims models are typically external except for insured claim experience which is most often internal data. Geocoding data and other non-traditional data elements are evenly split between internal and external. (Refer to Table 9.)

Table 9: Claims Model Sources (Internal vs. Third Party) by Data Elements

Claims Data Elements[9]	# of Companies Using the Data Element in a Claims AI/ML model*		
	Internal Data Source	External Data Source	Both Internal and External Data Sources
Insured Claim Experience – Home	33	1	14
Roof Data	2	36	--
Geocoding	12	16	3
Other: Non-Traditional	3	6	11
Historical Weather Information	--	15	--
Delay in Reporting	--	--	--
Defect Identification in Images	3	1	--
Modeled Behavioral Characteristics	--	--	--
Hazard Detection in Images	3	--	--
Aerial Imagery	--	--	--
Loss Description	--	--	--
Smart Phone Devices	1	--	--
Security Systems	1	--	--
Potential Loss Estimates in Images	--	1	--
Personal Financial Information	--	1	--
National Recall Database	--	--	--
Insured Demographic Data	1	--	--
Geodemographic Data	--	1	--
Claims Estimates in Images	--	1	--

Only one company reported using a consumer or other type of "score" as an input for claims models. (Refer to Table 10.)

[9] For definitions, refer to Appendix H: Data Use Table Definitions.

Table 10: Companies' Use of Consumer or Other Type of "Score" as an Input for Claims Data Elements

Claims Data Elements[10]		
Number of Companies Using a Consumer or Other Type of "Score" as an Input		
Yes	No	Null
1	126	67

Refer to the "Customer Data Correction," "Governance," and "Third-Party" sections of this report for additional data analysis regarding company operations areas.

Companies were asked to estimate their use of data elements in their regression, static, or pre-2000 models. These are the models defined to be excluded from the AI/ML definition for this survey. There is limited use of the data elements in claims models. (Refer to Table 11.)

Table 11: Companies' Estimated Use of **Regression, Static, or Pre-2000 Models f**or Claims

Claims Data Elements[11]	Estimated # of Regression, Static, or Pre-2000 Models
Auto	1
Crime Rates	1
Criminal Convictions	--
Defect ID	1
Earthquake	--
Education	--
Excess Wind Hail	--
Facial Detection	--
Flood	1
Geocoding	3
Geodemo Data	1
Hazard	1
Historical Weather	2
Home	5
Hurricane	1
Income	--
Insured Demographic Data	3
Job Stability	1
Loss Statistics	--
Medical	1
Occupation	--
Online Media	--
Other	1
Personal Financial Info	--
Potential Loss	--
Roof	2
Score	2
Security System	1

[10] For definitions, refer to Appendix H: Data Use Table Definitions.
[11] For definitions, refer to Appendix H: Data Use Table Definitions.

Claims Data Elements[11]	Estimated # of Regression, Static, or Pre-2000 Models
Smart Home	--
Tax Rates	--
Topography	1
Wildfire	--

COMPANY OPERATION: FRAUD DETECTION[12]

Out of 194 reporting companies, 81 companies reported using AI/ML for fraud-detection operations and 26 reported having models under construction (i.e., in a research, proof of concept, or prototype stage). Excerpt from Table 3:

Company Operation Area[13]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Fraud Detection	81	42	26	13	87	45	194	100

Fraud-Detection Model Uses

In insurance fraud detection, companies reported using AI/ML models mostly to refer claims for further investigation (81). The uses of fraud-detection models identified in Table 12 were options that could be selected in the survey template. Companies noted some additional uses of fraud detection models in their write-in comments: identification of fraudulent documents and agent behavior, claims watch list, text analysis, entity resolution, document analysis, photo analysis, network graph visualization, internal rules-based scoring, financial motive, and roof damages.

Some models are being researched, in proof-of-concept status, or in prototype for fraud detection. The most growth is expected in three of the top four current frequent uses: for organized crime rings identification, for social network analysis, and to detect medical producer fraud.

Table 12: Companies' Use of Fraud-Detection Models

Fraud-Detection Model Uses[14]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None (N/A)
Referral of Claims for Further Investigation	81	3	5	9	96
Organized Crime Rings Identification	26	9	2	--	157
Social Network Analysis	24	11	2	--	157

[12] For definitions, refer to Appendix C: Definitions Specific to Fraud Detection.

[13] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

[14] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Fraud-Detection Model Uses[14]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None (N/A)
Detect Medical Producer Fraud	23	12	2	--	157
Fast-tracking of Likely Non-Fraudulent Claims	19	14	--	9	152
Detect First-Party Liability	16	--	--	--	178
Detect Third-Party Liability	15	--	--	--	19
Other Fraud Detection-Related Functions	7	3	4	4	176
Evaluation of Potential for Intentional Infliction of Damage	7	--	--	--	189
Fraudulent Quote Detection	1	1	--	1	191
Facial Recognition & Behavior Models	--	--	--	--	194

The level of decisions influenced by AI/ML varies by model use. For most uses, the models are used for augmentation and support and rarely automation. (Refer to Table 13.)
(Note that Table 13 differs from the previous tables because the data represents the number of models instead of the number of companies.)

Table 13: Level of Decision-Making by Use of Fraud-Detection Models

Fraud-Detection Model Uses[15]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Referral of Claims for Further Investigation	15	82	63	--
Organized Crime Rings Identification	3	26	16	--
Social Network Analysis	5	10	25	--
Detect Medical Producer Fraud	--	20	23	--
Fast-tracking of Likely Non-Fraudulent Claims	17	10	18	--
Detect First-Party Liability	1	20	5	--
Detect Third-Party Liability	1	22	5	--
Other Fraud Detection-Related Functions	--	7	23	--
Evaluation of Potential for Intentional Infliction of Damage	--	3	4	--
Fraudulent Quote Detection	1	1	1	1
Facial Recognition & Behavior Models	--	--	--	--

*"Automation" was defined as no human intervention on execution. "Augmentation" was defined as a model that suggests an answer and advises the human making a decision. "Support" was defined as a model that provides information but does not suggest a decision or action.

[15] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Fraud detection models are mostly from third-party sources but for some uses, the models are just as frequently in-house as they are from third parties. (Refer to Table 14.)

Table 14: Fraud-Detection Model Sources by Model Use

Fraud-Detection Model Uses	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Referral of Claims for Further Investigation	85	53	75	47	160	100
Organized Crime Rings Identification	14	31	31	69	45	100
Social Network Analysis	10	25	30	75	40	100
Detect Medical Producer Fraud	15	35	28	65	43	100
Fast-tracking of Likely Non-Fraudulent Claims	24	53	21	47	45	100
Detect First-Party Liability	--	--	26	100	26	100
Detect Third-Party Liability	3	11	25	89	28	100
Other Fraud Detection-Related Functions	14	47	16	53	30	100
Evaluation of Potential for Intentional Infliction of Damage	--	--	7	100	7	100
Fraudulent Quote Detection	2	50	2	50	4	100
Facial Recognition & Behavior Models	--	--	--	--	--	--

For each fraud detection model, the type of model by category was identified. When a company used more than one category type for the same model, the model will be counted in all applicable categories. This leads to overcounting the models in total; but provides a good comparison of how often each category type is being used by insurers. This table shows a total of 704 models, whereas the actual number of models is 428. The most frequently used categories of fraud detection models are ENS = Ensemble (185), and RS = Rule System (87), DT = Decision Trees (76), and AML – Automated Machine Learning (70). (Refer to Table 15.)

Table 15: Number of Models (in Use or Under Construction) by Use of AI/ML Model

Fraud Detection Model Uses	Number of Models (In Use or Under Construction) by Use of AI/ML Model												
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Referral of Claims for Further Investigation	28	74	1	17	21	17	2	22	9	2	2	9	6	234
Organized Crime Rings Identification	--	10	--	1	15	1	--	2	--	6	1	1	19	62
Social Network Analysis	7	10	6	1	9	1	--	2	--	4	1	5	15	63
Detect Medical Producer Fraud	--	17	--	5	11	5	--	9	--	2	1	1	19	75
Fast-tracking of Likely Non-Fraudulent Claims	14	30	--	4	8	4	--	14	--	--	--	--	--	77

Fraud Detection Model Uses	Number of Models (In Use or Under Construction) by Use of AI/ML Model													
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Detect First-Party Liability	2	14	--	9	11	9	--	8	--	--	1	1	--	55
Detect Third-Party Liability	5	16	---	9	10	9	--	8	3	--	1	1	--	62
Other Fraud Detection-Related Functions	6	10	-	4	1	4	--	5	--	--	--	--	10	48
Evaluation of Potential for Intentional Infliction of Damage	--	4	--	4	--	4	--	4	--	--	--	--	--	19
Fraudulent Quote Detection	--	--	--	1	1	1	--	2	--	--	1	1	1	9
Facial Recognition & Behavior Models	--	--	--	--	--	--	--	--	--	--	--	--	--	--
TOTAL	62	185	7	55	87	55	2	76	12	14	8	19	70	704

*DL = Deep Learning, ENS = Ensemble, NN = Neural Network, REG = Regularization, RS = Rule System, RGS = Regression, BAY = Bayesian Methods, DT = Decision Trees, DR = Dimensionality Reduction, IB = Instance Based, CLU = Clustering, AI = Artificial Intelligence that is not ML; AML = Automated Machine Learning (a third-party company provides the answer), and TOTAL includes "Other" models.

Data Elements

The survey was limited to the use of the "more advanced" AI/ML. Therefore, the data element information here does not represent the industry's entire use of big data (which would require adding in the data element information from excluded models (e.g., regression-type models, etc.).

For fraud detection, the following three data elements were the most frequently reported as being used for AI/ML:

- Insured Claim Experience (66 companies)
- Personal Financial Information (22)
- Criminal Convictions (20)

There are at least some companies using other data elements in their fraud detection AI/ML models. (Refer to Table 16.)

Table 16: Companies' Use of Fraud-Detection Data Elements

Fraud-Detection Data Elements[16]	Number of Companies Using/Not Using the Data Element in a Fraud-Detection AI/ML Model*	
	Yes	No
Insured Claim Experience	66	128
Personal Financial Information	22	172
Criminal Convictions	20	174

[16] For definitions, refer to Appendix H: Data Use Table Definitions.

Fraud-Detection Data Elements[16]	Number of Companies Using/Not Using the Data Element in a Fraud-Detection AI/ML Model*	
	Yes	No
Others	17	177
Insured Demographic Data	12	182
Geodemographic Data	12	182
Other: Non-Traditional	11	183
Insured Claim Experience – Auto	11	183
Consumer or Other Type of "Score"	9	185
Online Media	8	186
Geocoding	8	186
Historical Weather Information	7	187
Medical	5	189
Industry Territorial Loss Statistics	5	189
Hurricane Weather Information	4	190
Excess Wind/Hail Model Output	4	190
Un-Structured Claims Adjuster Notes	3	191
Roof Data	3	191
Address	1	193

*The question is not whether the data element is used, but only whether the data element is used in an AI/ML model.

There are differences in data sources for the data elements. The data elements used in fraud-detection models are most often from external data sources although when looking for insured claim experience, companies most often use their own data. (Refer to Table 17.)

Table 17: Fraud-Detection Model Sources (Internal vs. Third Party) by Data Elements

Fraud-Detection Data Elements[17]	Internal Data Source	External Data Source	Both Internal and External Data Sources
Insured Claim Experience	25	13	24
Personal Financial Information	--	14	--
Criminal Convictions	--	12	--
Others: Non-Traditional	--	13	--
Insured Demographic Data	4	--	4
Geodemographic Data	1	3	4
Other: Non-Traditional	--	11	--
Insured Claim Experience – Auto	2	2	3
Consumer or Other Type of "Score"	--	5	--
Online Media	--	3	5
Geocoding	--	4	4
Historical Weather Information	--	3	4
Medical	--	--	1
Industry Territorial Loss Statistics	--	5	--
Hurricane Weather Information	--	--	4
Excess Wind/Hail Model Output	--	--	4

[17] For definitions, refer to Appendix H: Data Use Table Definitions.

Fraud-Detection Data Elements[17]	Internal Data Source	External Data Source	Both Internal and External Data Sources
Un-Structured Claims Adjuster Notes	--	--	--
Roof Data	3	--	--
Address	--	--	--

Two companies reported using a consumer or other type of "score" as an input for fraud-detection models. (Refer to Table 18.)

Table 18: Companies' Use of Consumer or Other Type of "Score" as an Input for Fraud-Detection Data Elements

Fraud-Detection Data Elements[18]		
Number of Companies Using a Consumer or Other Type of "Score" as an Input		
Yes	No	Null
2	117	76

Refer to the "Customer Data Correction", "Governance," and "Third-Party" sections of this report for additional data analysis regarding company operations areas.

Companies were asked to estimate their use of data elements in their regression, static, or pre-2000 models. These are the models defined to be excluded from the AI/ML definition for this survey. There is very limited use of the data elements in these types of fraud detection models.

Table 19: Companies' Estimated Use of **Regression, Static, or Pre-2000 models** for Fraud Detection

Fraud-Detection Data Elements[19]	Estimated # of Regression, Static, or Pre-2000 Models
Auto	16
Crime Rates	--
Criminal Convictions	7
Defect ID	1
Earthquake	--
Education	--
Excess Wind Hail	--
Facial Detection	--
Flood	--
Geocoding	3
Geodemo Data	3
Hazard	--
Historical Weather	1
Home	19
Hurricane	--
Income	--
Insured Demographic Data	3

[18] For definitions, refer to Appendix H: Data Use Table Definitions.
[19] For definitions, refer to Appendix H: Data Use Table Definitions.

Fraud-Detection Data Elements[19]	Estimated # of Regression, Static, or Pre-2000 Models
Job Stability	--
Loss Statistics	1
Medical	1
Occupation	--
Online Media	--
Other	--
Personal Financial Info	7
Potential Loss	--
Roof	1
Score	7
Security System	--
Smart Home	1
Tax Rates	--
Topography	--
Wildfire	--

COMPANY OPERATION: MARKETING[20]

0ut of 194 reporting companies, 92 companies reported using AI/ML for marketing operations and 11 reported having models under construction (i.e., in a research, proof of concept, or prototype stage). Approximately half of the companies are using AI/ML for marketing. (See Table 3.)

Excerpt from Table 3:

Company Operation Area[21]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Marketing	92	47	11	6	91	47	194	100

<u>Marketing Model Uses</u>

Companies are using many marketing models for multiple uses. Companies use marketing models for customer acquisition and retention (54 companies), targeted online advertising (46), provision of offers to existing customers (37), customer interactions using NLP (28), identification of potential customer groups (24), identification of recipients of mail and phone advertising (24), media mix marketing (16), and direct online sales (5). Only four companies are currently using models for demand modeling, three for other marketing-related functions and one for click analysis on third-party sales.

The uses of marketing models identified in Table 20 were options that could be selected in the survey template. Companies noted some additional uses of marketing models in their write-in comments: multi-touch attribution, program effectiveness, agent productivity, sales lead prioritization, and brand safety.

---

[20] For definitions, refer to Appendix D: Definitions Specific to Marketing.
[21] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Table 20: Companies' Use of Marketing Models

Marketing Model Uses[22]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None (N/A)
Customer Acquisition and Retention	54	--	--	--	141
Targeted Online Advertising	46	3	3	--	143
Provision of Offers to Existing Customers	37	3	4	--	151
Identification of Potential Customer Groups	24	12	8	--	--
Media Mix Modeling	16	16	3	1	--
Customer Interactions Using NLP	28	3	--	--	--
Identification of Recipients of Mail or Phone Advertising	24	--	4	--	--
Other Marketing-Related Functions	3	12	--	1	--
Demand Modeling	4	2	--	--	--
Direct Online Sales	5	--	--	--	--
Click Analysis on Third-party Sales	1	--	--	--	--

Many of the marketing models were augmented, where a model provides an answer and advises the human who is making the decision. Marketing models are mostly augmented when used for customer acquisition and retention (137) and provision of offers to existing customers (78). Automated models, with no human intervention on execution, were most often used for targeted online advertising (69) and customer interactions using NLP (51). (Refer to Table 21.)

Table 21: Level of Decision-Making by Use of Marketing Models

Marketing Model Uses[23]	Number of Models (In Use or Under Construction) by Level of Decisions influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Targeted Online Advertising	69	49	23	--
Customer Acquisition and Retention	21	137	3	--
Provision of Offers to Existing Customers	--	78	1	--
Identification of Potential Customer Groups	22	17	63	--
Media Mix Modeling	4	18	30	--
Customer Interactions Using NLP	51	8	8	1
Identification of Recipients of Mail or Phone Advertising	10	68	1	--
Other Marketing-Related Functions	5	12	4	--

[22] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
[23] For definitions, See Appendix A: "Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention"

Marketing Model Uses[23]	Number of Models (In Use or Under Construction) by Level of Decisions influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Demand Modeling	2	4	4	--
Direct Online Sales	3	2	--	--
Click Analysis on Third-party Sales	--	--	1	--

*"Automation" was defined as no human intervention on execution. "Augmentation" was defined as a model that suggests an answer and advises the human making a decision. "Support" was defined as a model that provides information but does not suggest a decision or action.

Marketing models being used by insurance companies combine in-house and third-party sources in all but click analysis on third-party sales. In-house sources are predominately used for customer acquisition and retention (138), provision of offers to existing customers (69), and identification of potential customer groups (62). Third-party sources are prevalent in targeting online advertising (137), identification of recipients of mail or phone advertising (53), media mix modeling (51) (Refer to Table 22.)

Table 22: Marketing Model Sources by Model Use

Marketing Model Uses[24]	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Targeted Online Advertising	4	3	137	97	141	100
Customer Acquisition and Retention	138	86	23	14	161	100
Provision of Offers to Existing Customers	69	87	10	13	79	100
Identification of Potential Customer Groups	62	61	40	39	102	100
Media Mix Modeling	1	2	51	98	52	100
Customer Interactions Using NLP	30	44	38	56	68	100
Identification of Recipients of Mail or Phone Advertising	27	34	53	66	80	100
Other Marketing-Related Functions	6	29	15	71	21	100
Demand Modeling	3	30	7	70	10	100
Direct Online Sales	3	60	2	40	5	100
Click Analysis on Third-party Sales	--	--	1	100	1	100

For each marketing model, the type of model by category was identified. When a company used more than one category type for the same model, the model will be counted in all applicable categories. This leads to overcounting the models in total; but provides a good comparison of how often each category type is being used by insurers. This table shows a total of 786 models, whereas the actual number of models is 720. The most frequently used categories of fraud detection models are ENS = Ensemble (185), and RS = Rule System (87), DT = Decision Trees (76), and AML – Automated Machine Learning (70). (Refer to Table 23.)

[24] For definitions, See Appendix A: "Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention"

Table 23: Number of Models (In Use or Under Construction) by Use of AI/ML Model

Marketing Model Uses	Number of Models (In Use or Under Construction) by Use of AI/ML Model													
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Targeted Online Advertising	1	16	--	3	6	3	28	6	--	--	--	9	37	163
Customer Acquisition and Retention	1	105	10	--	3	--	--	17	--	--	--	--	3	155
Provision of Offers to Existing Customers	1	63	--	--	--	--	8	6	--	--	--	--	2	80
Identification of Potential Customer Groups	--	51	--	--	4	--	--	2	--	--	29	--	17	113
Media Mix Modeling	--	12	--	6	1	6	15	--	--	--	--	--	14	70
Customer Interactions Using NLP	31	--	8	--	--	--	--	2	--	--	--	1	--	76
Identification of Recipients of Mail or Phone Advertising	--	59	--	3	1	3	5	8	--	--	--	--	1	90
Other Marketing-Related Functions	--	3	--	--	--	--	10	--	--	--	--	3	6	22
Demand Modeling	--	2	--	--	--	--	1	--	--	--	--	--	7	10
Direct Online Sales	--	3	--	--	--	--	--	--	--	--	--	--	2	5
Click Analysis on Third-party Sales	--	--	--	--	--	--	--	--	--	--	--	1	--	2
TOTAL	34	314	18	12	15	12	67	41	--	--	29	14	89	786

*DL = Deep Learning, ENS = Ensemble, NN = Neural Network, REG = Regularization, RS = Rule System, RGS = Regression, BAY = Bayesian Methods, DT = Decision Trees, DR = Dimensionality Reduction, IB = Instance Based, CLU = Clustering, AI = Artificial Intelligence that is not ML; AML = Automated Machine Learning (a third-party company provides the answer), and TOTAL includes "Other" models.

Data Elements

The survey was limited to the use of the "more advanced" AI/ML. Therefore, the data element information here does not represent the industry's entire use of big data (which would require adding in the data element information from excluded models (e.g., regression-type models, etc.).

For marketing, the following five data elements were the most frequently reported as being used:

- Insured Demographic Data (71 companies)
- Geodemographic Data (66)
- Consumer or Other Type of "Score" (64)
- Occupation (55)
- Education (54)

There are at least some companies using additional data elements as shown in Table 24.

Table 24: Companies' Use of Marketing Data Elements

Marketing Data Elements[25]	Number of Companies Using/Not Using the Data Element in a Marketing AI/ML Model*	
	Yes	No
Insured Demographic Data	71	123
Geodemographic Data	66	128
Consumer or Other Type of "Score"	64	130
Occupation	55	139
Education	54	140
Personal Financial Information	52	142
Insured Claim Experience – Home	38	156
Geocoding	35	159
Income	34	160
Roof Data	25	169
Insured Claim Experience – Auto	24	170
Job Stability	24	170
Historical Weather Information	24	170
Territorial Crime Rates	22	172
Other: Non-Traditional	16	178
Quote Stage	10	184
Product Ownership	10	184
Lapse in Coverage	10	184
EFT Enrollment	10	184
Change in Policy Premium Amount	10	184
Online Media	7	187
Wildfire Wind/Hail Output	5	189
Hurricane Model Output	5	189
Earthquake Model Output	5	189
Industry Territorial Loss Statistics	4	190
Territorial Tax Rates	3	191
Excess Wind/Hail Model Output	3	191
New/Existing Home Buyer	3	191
Defect Identification in Images	3	191
Year Home Built	3	191
Stories in Home	3	191
Flood Model Output	2	192
Web Visit Data	1	193
Potential Loss Estimates in Images	1	193
Summarized Vehicle Data	1	193
Partnership-specific Information	1	193
Modeled Customer Behavior	1	193
Home Property Data	1	193

[25] For definitions, see Appendix H: Data Use Table Definitions.

*The question is not whether the data element is used, but only whether the data element is used in an AI/ML model.

There are differences in data sources for the data elements. On average, data sources are evenly split between internal and external sources. (Refer to Table 20.)

Table 25: Marketing Model Sources (Internal vs. Third Party) by Data Elements

Marketing Data Elements[26]	Internal Data Source	External Data Source	Both Internal and External Data Sources
Insured Demographic Data	35	22	14
Geodemographic Data	28	25	13
Occupation	19	34	2
Education	20	32	2
Personal Financial Information	20	21	11
Insured Claim Experience – Home	37	1	--
Geocoding	16	19	--
Income	--	32	2
Roof Data	20	5	--
Insured Claim Experience – Auto	24	--	--
Job Stability	19	5	--
Historical Weather Information	4	20	--
Territorial Crime Rates	3	19	--
Other: Non-Traditional	14	2	--
Quote Stage	--	--	--
Product Ownership	--	--	--
Lapse in Coverage	--	--	--
EFT Enrollment	--	--	--
Change in Policy Premium Amount	--	--	--
Online Media	--	7	--
Wildfire Wind/Hail Output	4	1	--
Hurricane Model Output	4	1	--
Earthquake Model Output	4	1	--
Industry Territorial Loss Statistics	4	--	--
Territorial Tax Rates	3	--	--
Excess Wind/Hail Model Output	3	--	--
New/Existing Home Buyer	--	--	--
Defect Identification in Images	--	3	--
Year Home Built	--	--	--
Stories in Home	--	--	--
Flood Model Output	1	1	--
Web Visit Data	--	--	--
Potential Loss Estimates in Images	--	1	--
Summarized Vehicle Data	--	--	--
Partnership-specific Information	--	--	--
Modeled Customer Behavior	--	--	--
Home Property Data	--	--	--

[26] For definitions, see Appendix H: Data Use Table Definitions.

Approximately 20% of companies use a consumer or other type of "score" as an input for marketing data elements. (Refer to Table 26.)

Table 26: Companies' Use of Consumer or Other Type of "Score" as an Input for Marketing Data Elements

Marketing Data Elements[27]		
Number of Companies Using a Consumer or Other Type of "Score" as an Input		
Yes	No	Null
43	75	76

Companies were asked to estimate their use of data elements in their regression, static, or pre-2000 models. These are the models defined to be excluded from the AI/ML definition for this survey. The data element used most often in these types of marketing models is a score followed by insured demographic data. (Refer to Table 27.)

Table 27: Companies' Estimated Use of **Regression, Static, or Pre-2000 models** for Marketing

Marketing Data Elements[28]	Estimated # of Regression, Static, or Pre-2000 Models
Auto	4
Crime Rates	--
Criminal Convictions	--
Defect ID	--
Earthquake	--
Education	7
Excess Wind Hail	--
Facial Detection	--
Flood	--
Geocoding	4
Geodemo Data	19
Hazard	--
Historical Weather	5
Home	2
Hurricane	--
Income	11
Insured Demographic Data	49
Job Stability	1
Loss Statistics	3
Medical	--
Occupation	8
Online Media	2
Other	--
Personal Financial Info	15
Potential Loss	--
Roof	--
Score	62

[27] For definitions, see Appendix H: Data Use Table Definitions.
[28] For definitions, refer to Appendix H: Data Use Table Definitions.

Marketing Data Elements[28]	Estimated # of Regression, Static, or Pre-2000 Models
Security System	--
Smart Home	--
Tax Rates	--
Topography	--
Wildfire	--

Refer to the "Customer Data Correction," "Governance," and "Third-Party" sections of this report for additional data analysis regarding company operations areas.

COMPANY OPERATION: RATING[29]

Out of 194 reporting companies, 68 companies reported using AI/ML for rating operations and 21 reported having models under construction.

Excerpt from Table 3:

Company Operation Area[30]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Rating	68	35	21	11	105	54	194	100

<u>Rating Model Uses</u>

The most common use cases within the rating area of operations are rating class determination (48 companies) and retention modeling (42). (Refer to Table 28.)

Table 28: Companies' Interest in AI/ML for Rating Model Use

Rating Model Uses [31]	Number of Companies Using or Investigating AI/ML		
	Yes	No	Blank
Rating Class Determination	48	121	25
Price Optimization	1	164	29
Retention Modeling	42	122	30
Numerical Relativity Determination	2	162	30
Other Rate-Related Functions	23	141	30

The uses of rating models identified in Table 28 were options that could be selected in the survey template. Companies noted some additional uses of rating models in their write-in comments: close rate evaluation,

[29] For definitions, See Appendix E: Definitions Specific to Rating
[30] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
[31] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

GLM variable selection, tiering, monitoring business mix, subsidy analyses, roof scoring, water risk, fire risk, causes of estimate price changes, and excess loss loading variable recommendations.

The number of companies using these more advanced AI/ML models for rating class determination is expected to double given the reported number of companies conducting research and building prototypes. (Refer to Table 29.)

Table 29: Companies' Use of Rating Models

Rating Model Uses[32]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None (N/A)
Rating Class Determination	21	25	--	2	146
Price Optimization	--	1	--	--	193
Retention Modeling	36	--	--	6	152
Numerical Relativity Determination	1	--	--	1	192
Other Rate-Related Functions	18	3	--	2	171

Rating models are mostly automated and/or provide support. A few rating models provide augmentation. (Refer to Table 30.)

Table 30: Level of Decision-Making by Use of Rating Models

Rating Model Uses[33]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Rating Class Determination	35	6	19	4
Price Optimization	--	--	1	1
Retention Modeling	30	--	38	--
Numerical Relativity Determination	--	1	1	--
Other Rate-Related Functions	2	4	16	--

*"Automation" was defined as no human intervention on execution. "Augmentation" was defined as a model that suggests an answer and advises the human making a decision. "Support" was defined as a model that provides information but does not suggest a decision or action.

Rating models tend to be developed in-house by companies and not third parties. About 90% of the rating models are developed by companies in-house. (Refer to Table 31.)

Table 31: Rating Model Sources by Model Use

Rating Model Uses[34]	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Rating Class Determination	54	84	10	16	64	100

[32] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

[33] For definitions, See Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

[34] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Rating Model Uses[34]	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Price Optimization	2	100	--	--	2	100
Retention Modeling	67	99	1	1	68	100
Numerical Relativity Determination	2	100	--	--	2	100
Other Rate-Related Functions	17	77	5	23	22	100

For each rating model, the type of model by category was identified. When a company used more than one category type for the same model, the model will be counted in all applicable categories. This leads to overcounting the models in total; but provides a good comparison of how often each category type is being used by insurers. This table shows a total of 176 models, whereas the actual number of models is 158. The most frequently used categories of rating models are ENS = Ensemble (87) and DL = Deep Learning (36). (Refer to Table 32)

Table 32: Number of Models (In Use or Under Construction) by Use of AI/ML Model

Rating Model Uses	Number of Models (In Use or Under Construction) by Use of AI/ML Model												
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Rating Class Determination	32	12	2	5	--	5	--	2	9	--	6	2	--	75
Price Optimization	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Retention Modeling	--	57	--	--	--	--	--	4	3	--	--	--	--	64
Numerical Relativity Determination	--	1	--	--	--	--	--	1	--	--	0	--	--	2
Other Rate-Related Functions	4	17	--	4	--	4	--	--	1	--	5	--	--	35
TOTAL	36	87	2	9	--	9	--	7	13	--	11	2	--	176

*DL = Deep Learning, ENS = Ensemble, NN = Neural Network, REG = Regularization, RS = Rule System, RGS = Regression, BAY = Bayesian Methods, DT = Decision Trees, DR = Dimensionality Reduction, IB = Instance Based, CLU = Clustering, AI = Artificial Intelligence that is not ML; AML = Automated Machine Learning (a third-party company provides the answer), and TOTAL includes "Other" models.

Data Elements

The survey was limited to the use of the "more advanced" AI/ML. Therefore, the data element information here does not represent the industry's entire use of big data (which would require adding in the data element information from excluded models (e.g., regression-type models, etc.).

For rating, the following five data elements were the most frequently reported as being used for AI/ML:

- Insured Demographic Data (46)
- Geodemographic Data (46)
- Consumer or Other Type of "Score" (45)
- Geocoding (43)
- Insured Claim Experience – Home (37)

There are at least some companies using other data elements. (Refer to Table 33.)

Table 33: Companies' Use of Rating Data Elements

Rating Data Elements[35]	Number of Companies Using/Not Using the Data Element in a Rating AI/ML Model*	
	Yes	No
Insured Demographic Data	46	148
Geodemographic Data	46	148
Consumer or Other Type of "Score"	45	149
Geocoding	43	151
Insured Claim Experience--Home	37	157
Historical Weather Information	32	162
Roof Data	31	163
Hazard Detection in Images	31	163
Defect Identification in Images	31	163
Personal Financial Information	10	184
Other: Non-Traditional	9	185
Insured Claim Experience--Auto	9	185
Parcel Information	8	186
Industry Territorial Loss Statistics	7	187
Excess Wind/Hail Model Output	7	187
Wildfire Wind/Hail Model Output	5	189
Hurricane Model Output	4	190
Territorial Crime Rates	3	191
Potential Loss Estimates in Images	3	191
Topography	2	192
Security Systems	2	192
Smart Home Devices	1	193
Geospatial Imaging	1	193

*The question is not whether the data element is used, but only whether the data element is used in an AI/ML model.

There are differences in data sources for the data elements. Data sources for rating models are about 60% external and 40% internal but vary by data element. (Refer to Table 34.)

Table 34: Rating Model Sources (Internal vs. Third Party) by Data Elements

Rating Data Elements[36]	Internal Data Source	External Data Source	Both Internal and External Data Sources	External Non-Traditional Data Source
Insured Demographic Data	46	--	--	--
Geodemographic Data	3	42	1	--
Consumer or Other Type of "Score"	14	25	6	--
Geocoding	1	31	11	--
Insured Claim Experience--Home	6	5	26	--
Historical Weather Information	7	25	--	--
Roof Data	4	6	21	--

[35] For definitions, refer to Appendix H: Data Use Table Definitions.
[36] For definitions, refer to Appendix H: Data Use Table Definitions.

Rating Data Elements[36]	Internal Data Source	External Data Source	Both Internal and External Data Sources	External Non-Traditional Data Source
Hazard Detection in Images	--	6	25	--
Defect Identification in Images	5	7	19	--
Personal Financial Information	1	9	--	--
Other: Non-Traditional	--	9	--	--
Insured Claim Experience--Auto	--	3	6	--
Parcel Information	--	--	--	8
Industry Territorial Loss Statistics	--	7	--	--
Excess Wind/Hail Model Output	1	3	3	--
Wildfire Wind/Hail Model Output	--	1	4	--
Hurricane Model Output	--	1	3	--
Territorial Crime Rates	--	3	--	--
Potential Loss Estimates in Images	--	3	--	--
Topography	1	1	--	--
Security Systems	2	--	--	--
Smart Home Devices	1	--	--	--
Geospatial Imaging	--	--	--	1

Less than half of the companies use a consumer or other type of score as a data element in their advanced AI/ML models. (Refer to table 35.)

Table 35: Companies' Use of Consumer or Other Type of "Score" as an Input for Rating Data Elements

Rating Data Elements[37]		
Number of Companies Using a Consumer or Other Type of "Score" as an Input		
Yes	No	Blank
48	69	77

Companies were asked to estimate their use of data elements in their regression, static, or pre-2000 models. These are the models defined to be excluded from the AI/ML definition for this survey. Compared to other company operation areas, the use of the data elements in these types of rating models is substantially higher. (Refer to Table 36.)

Table 36: Companies' Estimated Use of **Regression, Static, or Pre-2000 models** for Rating

Rating Data Elements[38]	Estimated # of Regression, Static, or Pre-2000 Models
Auto	306
Crime Rates	42
Criminal Convictions	1
Defect ID	12
Earthquake	45
Education	10
Excess Wind Hail	157

[37] For definitions, refer to Appendix H: Data Use Table Definitions.
[38] For definitions, refer to Appendix H: Data Use Table Definitions.

Rating Data Elements[38]	Estimated # of Regression, Static, or Pre-2000 Models
Facial Detection	--
Flood	12
Geocoding	270
Geodemo Data	534
Hazard	36
Historical Weather	458
Home	620
Hurricane	102
Income	5
Insured Demographic Data	656
Job Stability	--
Loss Statistics	364
Medical	--
Occupation	96
Online Media	--
Other	--
Personal Financial Info	241
Potential Loss	--
Roof	427
Score	554
Security System	136
Smart Home	23
Tax Rates	3
Topography	48
Wildfire	58

Refer to the "Customer Data Correction", "Governance," and "Third-Party" sections of this report for additional data analysis regarding company operations areas.

COMPANY OPERATION: UNDERWRITING[39]

Out of 194 reporting companies, 91 companies reported using AI/ML for fraud-detection operations and 29 reported having models under construction.

Excerpt from Table 3:

Company Operation Area[40]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Underwriting	91	47	29	15	74	38	194	100

[39] For definitions, refer to Appendix F: Definitions Specific to Underwriting.
[40] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Underwriting Model Uses

Most current underwriting models are used for verification of policy characteristics (37 companies), automated denial (30), or input into non-automated denial decisions (26). The use of underwriting models for renewals and reinstatements should grow significantly over the next few years as there are 28 companies researching and 19 companies at the stage of proof of concept. No companies reported using or having plans to use AI/ML models for company placement.

The uses of underwriting models identified in Table 37 were options that could be selected in the survey template. Companies noted some additional uses of underwriting models in their write-in comments: work triage, underwriting flags/referrals, wildfire risk, guidance for commercial lines premium adjustments, and inspections/surveys.

Table 37: Companies' Use of Underwriting Models

Underwriting Model Uses[41]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None (N/A)
Automate Processing Through the Agency Channel	7	3	--	--	184
Automated Approval	11	5	1	--	177
Automated Denial	30	1	--	2	161
Company Placement	--	--	--	--	194
Down-payment Requirements	--	1	--	--	193
Input into Non-Automated Approval Decision	8	5	3	--	178
Input into Non-Automated Denial Decision	26	3	9	1	155
Motor Vehicle Record Reordering	3	3	--	--	188
Other	26	31	2	--	135
Policy Anomaly Detection	2	1	3	2	186
Renewals and Reinstatements	13	28	19	2	133
Underwriting Tier Determination	3	3	--	--	188
Verification of Policy Characteristics	37	3	3	2	149

Underwriting models are predominantly automated and used for support. (Refer to Table 38.)

Table 38: Level of Decision-Making by Use of Underwriting Models

Underwriting Model Uses[42]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Automate Processing Through the Agency Channel	6	6	1	--

[41] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
[42] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Underwriting Model Uses[42]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Automated Approval	11	2	1	--
Automated Denial	39	7	2	--
Down-payment Requirements	--	--	--	1
Input into Non-Automated Approval Decision	--	2	15	--
Input into Non-Automated Denial Decision	6	19	21	--
Motor Vehicle Record Reordering	3	0	0	--
Other	67	2	21	--
Policy Anomaly Detection	--	4	5	--
Renewals and Reinstatements	25	8	35	--
Underwriting Tier Determination	5	--	3	--
Verification of Policy Characteristics	39	21	36	--

*"Automation" was defined as no human intervention on execution. "Augmentation" was defined as a model that suggests an answer and advises the human making a decision. "Support" was defined as a model that provides information but does not suggest a decision or action.

Most underwriting models (approximately 70%) are developed by companies. (Refer to Table 39.)

Table 39: Underwriting Model Sources by Model Use

Underwriting Model Uses[43]	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Automate Processing Through the Agency Channel	12	92	1	8	13	100
Automated Approval	12	86	2	14	14	100
Automated Denial	37	77	11	23	48	100
Down-payment Requirements	1	100	--	--	1	100
Input into Non-Automated Approval Decision	4	24	13	76	17	100
Input into Non-Automated Denial Decision	26	57	20	43	46	100
Motor Vehicle Record Reordering	--	--	3	100	3	100
Other	84	93	6	7	90	100
Policy Anomaly Detection	7	78	2	22	9	100
Renewals and Reinstatements	38	56	30	44	68	100
Underwriting Tier Determination	3	38	5	63	8	100
Verification of Policy Characteristics	73	76	23	24	96	100

For each underwriting model, the type of model by category was identified. When a company used more than one category type for the same model, the model will be counted in all applicable categories. This leads to overcounting the models in total; but provides a good comparison of how often each category

[43] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

type is being used by insurers. This table shows a total of 638 models, whereas the actual number of models is 413. The most frequently used categories of underwriting models are DL = Deep Learning (189), ENS = Ensemble (142); and BAY = Bayesian Methods (94). (Refer to Table 40.)

Table 40: Number of Models (In Use or Under Construction) by Use of AI/ML Model

Underwriting Model Uses	Number of Models (In Use or Under Construction) by Use of AI/ML Model													
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AML*	TOTAL*
Automate Processing Through the Agency Channel	7	7	--	1	--	1	--	--	--	--	---	-	--	16
Automated Approval	2	11	--	--	--	--	74	1	--	--	--	--	--	88
Automated Denial	4	36	4	--	--	--	1	2	--	--	--	6	1	55
Down-payment Requirements	--	--	--	--	--	--	--	1	--	--	--	--	--	1
Input into Non-Automated Approval Decision	12	2	1	3	1	3	1	2	--	1	1	1	1	32
Input into Non-Automated Denial Decision	26	12	1	4	1	4	1	5	--	--	1	1	1	64
Motor Vehicle Record Reordering	--	--	--	--	3	--	--	--	--	--	--	--	--	3
Other	22	50	13	2	2	2	1	3	3	--	2	2	1	103
Policy Anomaly Detection	1	--	2	--	1	--	--	--	--	3	3	2	--	14
Renewals and Reinstatements	44	16	11	3	1	3	7	9	--	--	7	5	7	117
Underwriting Tier Determination	--	3	--	3	--	3	--	3	--	--	--	5	--	17
Verification of Policy Characteristics	70	5	8	--	7	--	9	6	--	--	--	2	6	128
TOTAL	189	142	40	16	16	16	94	32	3	4	14	24	17	638

*DL = Deep Learning, ENS = Ensemble, NN = Neural Network, REG = Regularization, RS = Rule System, RGS = Regression, BAY = Bayesian Methods, DT = Decision Trees, DR = Dimensionality Reduction, IB = Instance Based, CLU = Clustering, AI = Artificial Intelligence that is not ML, AML = Automated Machine Learning (a third-party company provides the answer), and TOTAL includes "Other" models.

<u>Data Elements</u>

The survey was limited to the use of the "more advanced" AI/ML. Therefore, the data element information here does not represent the industry's entire use of big data (which would require adding in the data element information from excluded models (e.g., regression-type models, etc.).

The following four data elements were the most frequently reported as being used for AI/ML underwriting systems:

- Roof Data (59 companies)
- Consumer or Other Type of "Score" (49)
- Defect Identification in Images (47)

- Insured Demographic Data (37)

There are companies using other data elements in AI/underwriting models. (See Table 41.)

Table 41: Companies' Use of Underwriting Data Elements

Underwriting Data Elements[44]	Number of Companies Using/Not Using the Data Element in an Underwriting AI/ML Model*	
	Yes	No
Roof Data	59	135
Consumer or Other Type of "Score"	49	145
Defect Identification in Images	47	147
Insured Demographic Data	37	157
Geocoding	27	167
Hazard Detection in Images	21	173
Insured Claim Experience – Home	19	175
Other: Non-Traditional	9	185
Historical Weather Information	8	186
Geodemographic Data	8	186
Topography	7	187
Insured Claim Experience – Auto	7	187
Industry Territorial Loss Statistics	7	187
Potential Loss Estimates in Images	5	189
Vehicle Condition	4	190
Territorial Crime Rates	3	191
Excess Wind/Hail Model Output	3	191
Aerial Imagery	3	191
Wildfire Wind/Hail Model Output	2	192
Personal Financial Information	2	192
Parcel Information	2	192
Valuation of Artwork/Collections	1	193

*The question is not whether the data element is used, but only whether the data element is used in an AI/ML model.

Underwriting data elements use almost the same amount of internal and external data sources. There are differences in data sources for the data elements. (Refer to Table 42.)

Table 42: Underwriting Model Sources (Internal vs. Third Party) by Data Elements

Underwriting Data Elements[45]	Internal Data Source	External Data Source	Both Internal and External Data Sources
Roof Data	43	10	6
Consumer or Other Type of "Score"	11	38	--
Defect Identification in Images	6	20	21
Insured Demographic Data	37	--	--

[44] For definitions, refer to Appendix H: Data Use Table Definitions.
[45] For definitions, refer to Appendix H: Data Use Table Definitions.

Underwriting Data Elements[45]	Internal Data Source	External Data Source	Both Internal and External Data Sources
Geocoding	3	17	7
Hazard Detection in Images	5	16	--
Insured Claim Experience – Home	15	4	--
Other: Non-Traditional	--	9	--
Historical Weather Information	3	4	1
Geodemographic Data	--	6	2
Topography	--	7	--
Insured Claim Experience – Auto	4	3	--
Industry Territorial Loss Statistics	--	7	--
Potential Loss Estimates in Images	--	5	--
Vehicle Condition	--	--	--
Territorial Crime Rates	--	3	--
Excess Wind/Hail Model Output	--	3	--
Aerial Imagery	--	--	--
Wildfire Wind/Hail Model Output	--	2	--
Personal Financial Information	--	2	--
Parcel Information	--	--	--
Valuation of Artwork/Collections	--	--	--

Of those companies who answered, approximately 36% use a consumer or other type of score as a data element. (Refer to Table 43.)

Table 43: Companies' Use of Consumer or Other Type of "Score" as an Input for Underwriting Data Elements

Underwriting Data Elements[46]		
Number of Companies Using a Consumer or Other Type of "Score" as an Input		
Yes	No	Null
43	77	75

Companies were asked to estimate their use of data elements in their regression, static, or pre-2000 models. These are the models defined to be excluded from the AI/ML definition for this survey. The data elements used most often in these types of underwriting models are the home, geodemo data, and a score. (Refer to Table 44.)

Table 44: Companies' Estimated Use of **Regression, Static, or Pre-2000 models** for Underwriting

Underwriting Data Elements[47]	Estimated # of Regression, Static, or Pre-2000 Models
Auto	33
Crime Rates	4
Criminal Convictions	--
Defect ID	6

[46] For definitions, refer to Appendix H: Data Use Table Definitions.
[47] For definitions, refer to Appendix H: Data Use Table Definitions.

Underwriting Data Elements[47]	Estimated # of Regression, Static, or Pre-2000 Models
Earthquake	14
Education	--
Excess Wind Hail	24
Facial Detection	--
Flood	11
Geocoding	17
Geodemo Data	56
Hazard	5
Historical Weather	38
Home	87
Hurricane	25
Income	--
Insured Demographic Data	47
Job Stability	--
Loss Statistics	4
Medical	--
Occupation	--
Online Media	--
Other	--
Personal Financial Info	3
Potential Loss	--
Roof	41
Score	64
Security System	10
Smart Home	--
Tax Rates	--
Topography	6
Wildfire	20

Refer to the "Customer Data Correction," "Governance," and "Third-Party" sections of this report for additional data analysis regarding company operations areas.

COMPANY OPERATION: LOSS PREVENTION[48]

Out of 194 reporting companies, 28 companies reported using AI/ML for loss prevention operations and 6 reported having models under construction.

[48] For definitions, refer to Appendix G: Definitions Specific to Loss Prevention.

Excerpt from Table 3:

Company Operation Area[49]	Number and Percentage of Companies							
	In Use		Research, Proof of Concept, Prototype		Not Using		Total	
	#	%	#	%	#	%	#	%
Loss Prevention	28	14	6	3	160	82	194	100

Loss Prevention Model Uses

Out of the 6 main company operations in this study, the least number of companies use loss prevention models. It is notable that Table 3 and Table 45 are not consistent in the number of "in use" loss prevention models, with Table 45 having fewer reported "in use" models. It is possible this is an input error by the survey takers, but we do not know which table's data is more accurate. (See Table 45.)

There were no write-ins for "other" loss prevention model uses.

Table 45: Companies' Use of Loss Prevention Models

Loss Prevention Model Uses[50]	Number of Companies				
	In Use	Research	Proof of Concept	Prototype	None (N/A)
Guidance for Loss Control Inspections	12	3	--	3	176
Identification of High-Risk Customers	2	4	--	3	185
Risk-Mitigation Advice to Consumers	4	12	--	2	176
Determination of Advance Payments	3	--	1	--	190
Other Loss Prevention-Related Functions	0	--	--	--	194

Loss prevention models are used mostly for support. (Refer to Table 46.)

Table 46: Level of Decision-Making by Use of Loss Prevention Models

Loss Prevention Model Uses[51]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Guidance for Loss Control Inspections	3	16	7	--
Identification of High-Risk Customers	1	2	9	--

49 For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
50 For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
51 For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Loss Prevention Model Uses[51]	Number of Models (In Use or Under Construction) by Level of Decisions Influenced by AI/ML			
	Automation*	Augmentation*	Support*	Other
Risk-Mitigation Advice to Consumers	3	--	15	--
Determination of Advance Payments	3	--	--	1

*"Automation" was defined as no human intervention on execution. "Augmentation" was defined as a model that suggests an answer and advises the human making a decision. "Support" was defined as a model that provides information but does not suggest a decision or action.

Of the few reported loss prevention models, about half are developed by companies in-house and half are developed by a third party. (Refer to Table 47.)

Table 47: Loss Prevention Model Sources by Model Use

Loss Prevention Model Uses[52]	Model Source					
	In-House	In-House	Third-Party	Third-Party	Total	Total
	#	%	#	%	#	%
Guidance for Loss Control Inspections	11	42	15	58	26	100
Identification of High-Risk Customers	5	42	7	58	12	100
Risk-Mitigation Advice to Consumers	13	72	5	28	18	100
Determination of Advance Payments	4	100	--	--	4	100

For each loss prevention model, the type of model by category was identified. When a company used more than one category type for the same model, the model will be counted in all applicable categories. This leads to overcounting the models in total; but provides a good comparison of how often each category type is being used by insurers. This table shows a total of 88 models, whereas the actual number of models is 60. The most frequently used categories of loss prevention models are AI = Artificial Intelligence that is not ML (23); ENS = Ensemble (18); and NN = Neural Network (15). (Refer to Table 48.)

Table 48: Number of Models (In Use or Under Construction) by Use of AI/ML Model

Loss Prevention Model Uses	Number of Models (In Use or Under Construction) by Use of AI/ML Model												
	DL*	ENS*	NN*	REG*	RS*	RGS*	BAY*	DT*	DR*	IB*	CLU*	AI*	AMI*	TOTAL*
Guidance for Loss Control Inspections	3	6	7	2	1	2	--	5	--	--	--	12	3	41
Identification of High-Risk Customers	--	1	2	2	--	2	1	3	--	--	1	9	--	21
Risk-Mitigation Advice to Consumers	5	11	3	--	1	--	--	--	--	--	--	2	--	22
Determination of Advance Payments	--	--	3	--	--	--	--	1	--	--	--	--	--	4
Total	8	18	15	4	2	4	1	9	--	--	1	23	3	88

*DL = Deep Learning, ENS = Ensemble, NN = Neural Network, REG = Regularization, RS = Rule System, RGS = Regression, BAY = Bayesian Methods, DT = Decision Trees, DR = Dimensionality Reduction, IB = Instance Based, CLU = Clustering, AI = Artificial

[52] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Intelligence that is not ML, AML = Automated Machine Learning (a third-party company provides the answer), and TOTAL includes "Other" models.

Data Elements

The survey was limited to the use of the "more advanced" AI/ML. Therefore, the data element information here does not represent the industry's entire use of big data (which would require adding in the data element information from excluded models (e.g., regression-type models, etc.).

The following four data elements were the most frequently reported as being used for AI/ML loss prevention:

- Consumer or Other Type of "Score" (7 companies)
- Roof Data (5)
- Insured Demographic Data (3)
- Insured Claim Experience – Home (3)

Other data elements are used in loss prevention models. (Refer to Table 49.)

Table 49: Companies' Use of Loss Prevention Data Elements

Loss Prevention Data Elements[53]	Number of Companies Using/Not Using the Data Element in a Loss Prevention AI/ML Model*	
	Yes	No
Consumer or Other Type of "Score"	7	187
Roof Data	5	189
Insured Demographic Data	3	191
Insured Claim Experience – Home	3	191
Geocoding	2	192
Geodemographic Data	2	192
Wildfire Wind/Hail Model Output	1	193
Hazard Detection in Images	1	193
Detect Identification in Images	1	193

*The question is not whether the data element is used, but only whether the data element is used in an AI/ML model.

Almost all loss prevention data is either internally or externally sourced. (Refer to Table 50.)

Table 50: Loss Prevention Model Sources (Internal vs. Third Party) by Data Elements

Loss Prevention Data Elements[54]	Number of Companies Using the Data Element in a Loss Prevention AI/ML Model*			
	Internal Data Source	External Data Source	Both Internal and External Data Sources	Blank
Consumer or Other Type of "Score"	3	4	--	--
Roof Data	1	4	--	--
Insured Demographic Data	3	--	--	--

[53] For definitions, refer to Appendix H: Data Use Table Definitions.
[54] For definitions, refer to Appendix H: Data Use Table Definitions.

Loss Prevention Data Elements[54]	Number of Companies Using the Data Element in a Loss Prevention AI/ML Model*			
	Internal Data Source	External Data Source	Both Internal and External Data Sources	Blank
Insured Claim Experience – Home	2	1	--	--
Geocoding	--	2	--	--
Geodemographic Data	--	2	--	--
Wildfire Wind/Hail Model Output	--	1	--	--
Hazard Detection in Images	--	1	--	--
Detect Identification in Images	--	1	--	--

Only 6 of the 28 companies using loss prevention models indicated they are using a consumer or other type of "score" as an input for any of the data elements. (Refer to Table 51.)

Table 51: Companies' Use of Consumer or Other Type of "Score" as an Input for Loss Prevention Data Elements

Loss Prevention Data Elements[55]		
Number of Companies Using a Consumer or Other Type of "Score" as an Input		
Yes	No	Null
6	100	88

Companies were asked to estimate their use of data elements in their regression, static, or pre-2000 models. These are the models defined to be excluded from the AI/ML definition for this survey. There is very limited use of the named data elements in these types of loss prevention models. (Refer to Table 52.)

Table 52: Companies' Estimated Use of **Regression, Static, or Pre-2000 models** for Loss Prevention

Loss Prevention Data Elements[56]	Estimated # of Regression, Static, or Pre-2000 Models
Auto	--
Crime Rates	1
Criminal Convictions	1
Defect ID	6
Earthquake	--
Education	--
Excess Wind Hail	--
Facial Detection	--
Flood	1
Geocoding	2
Geodemo Data	12
Hazard	6
Historical Weather	10
Home	13

[55] For definitions, refer to Appendix H: Data Use Table Definitions.
[56] For definitions, refer to Appendix H: Data Use Table Definitions.

Loss Prevention Data Elements[56]	Estimated # of Regression, Static, or Pre-2000 Models
Hurricane	1
Income	--
Insured Demographic Data	10
Job Stability	--
Loss Statistics	--
Medical	--
Occupation	--
Online Media	--
Other	--
Personal Financial Info	1
Potential Loss	--
Roof	11
Score	15
Security System	2
Smart Home	--
Tax Rates	--
Topography	--
Wildfire	--

Refer to the "Customer Data Correction," "Governance," and "Third-Party" sections of this report for additional data analysis regarding company operations areas.

CUSTOMER DATA CORRECTION

The following consumer data correction questions ask if consumers are provided information about data elements—other than what is required by law. The response rate to these questions is low. For the companies that did answer, few said "yes." (Refer to Table 53 and Table 54.)

Table 53: Companies' Disclosure to Consumers About the Data Elements by Company Operation Area

Are consumers provided information regarding the **data elements being used**? *(Answer should be no if not disclosing any information other than what is required by law.)*	
Company Operation Area[57]	Number of Companies
	Yes
Rating	6
Underwriting	4
Claims	--
Fraud Detection	1
Marketing	--
Loss Prevention	--
Other	3
None of the Above	67

[57] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Are consumers provided information regarding the **data elements being used**? *(Answer should be no if not disclosing any information other than what is required by law.)*	
Company Operation Area[57]	Number of Companies
	Yes
Null	118

Table 54: Companies' Disclosure to Consumers About the Purposes of Data Elements by Company Operation Area

Are consumers provided information regarding the purposes for which data elements are being used? *(Answer should be no if not disclosing any information other than what is required by law.)*	
Company Operation Area[58]	Number of Companies
	Yes
Rating	10
Underwriting	4
Claims	--
Fraud Detection	1
Marketing	--
Loss Prevention	--
Other	3
None of the Above	67
Null	114

*Three of the "yes" responses for rating are models in progress and not yet implemented. The answer is interpreted as: "When the model is implemented, the answer will be 'yes.'"

For those companies that answered the first two consumer questions, half said they had more consumer data correction processes than required by the federal Fair Credit Reporting Act (FCRA). As might be expected, rating and underwriting lead the list for which areas have more protections than required under FCRA. (Refer to Table 55.)

Table 55: Consumers' Ability to Correct Data by Company Operation Area

Outside of processes required because of FCRA, do consumers have an opportunity to challenge or correct their specific data?			
Company Operation Area[59]	Number of Companies		
	Yes	No	Blank
Rating	42	28	124
Underwriting	44	28	122

[58] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
[59] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Outside of processes required because of FCRA, do consumers have an opportunity to challenge or correct their specific data?			
Company Operation Area[59]	Number of Companies		
	Yes	No	Blank
Claims	34	39	121
Fraud Detection	30	41	123
Marketing	29	39	126
Loss Prevention	24	44	126
Other	24	42	128

GOVERNANCE[60]

The purpose of the model governance questions is to obtain a better understanding regarding a company's awareness of specific risk areas tied to selected categories in the NAIC's AI Principles. While companies may consider a principle, the governance responses represent whether the company has the principle "documented" within its governance program. (Refer to Tables 56–65.)

Table 56: Governance Documentation of NAIC AI Principle: AI/ML Systems Provide Disclosure and Transparency to Regulators Reviewing a Filing Related to Rating.

If using data, scores and/or AI/ML models aggregated or developed by a third-party, do those contracts include any conditions that would limit disclosure or otherwise limit transparency to regulators reviewing a filing related to Rating?		
Number of Companies		
Yes	No	Null
31	123	40

For the next few tables, an answer of "companywide" can be interpreted as applying to all company operation areas. (An insurer who entered data may have intended it to mean something else.) If answering how many companies have the governance topic documented in their governance plan for a particular company operation, add the number for the operational area + the number of companies that answered "companywide." For rating, 7 + 65 = 72 companies said "fairness and ethics considerations" are documented in the governance program.

Table 57: Governance Documentation of NAIC AI Principle: Fairness and Ethics Considerations

Are "Fairness and Ethics Considerations" documented in the governance program?		
Company Operation Area[61]	Number of Companies	
	Yes	No
Rating	7	187
Underwriting	5	189
Claims	20	174
Fraud Detection	14	180

[60] For definitions, refer to Appendix I: Model Governance Definitions.
[61] For definitions, refer to Appendix I: Model Governance Definitions.

Are "Fairness and Ethics Considerations" documented in the governance program?		
Company Operation Area[61]	Number of Companies	
	Yes	No
Marketing	6	188
Loss Prevention	1	193
Companywide	65	129
Other	4	190

Table 58: Governance Documentation of NAIC AI Principle: Accountability for Data Algorithms' Compliance with Laws, as Well as Intended and Unintended Impacts

Are "Accountability for Data Algorithms' Compliance with Laws, as Well as Intended and Unintended Impacts" documented in the governance program?		
Company Operation Area[62]	Number of Companies	
	Yes	No
Rating	7	187
Underwriting	5	189
Claims	24	170
Fraud Detection	14	180
Marketing	10	184
Loss Prevention	1	193
Companywide	65	129
Other	5	189

Table 59: Governance Documentation of NAIC AI Principle: Appropriate Resources and Knowledge Involved to Ensure Compliance with Laws, Including Those Related to Unfair Discrimination

Are "Appropriate Resources and Knowledge Involved to Ensure Compliance with Laws, Including Those Related to Unfair Discrimination" documented in the governance program?		
Company Operation Area[63]	Number of Companies	
	Yes	No
Rating	29	165
Underwriting	27	167
Claims	46	150
Fraud Detection	36	158
Marketing	32	162
Loss Prevention	23	171
Companywide	92	102
Other	5	189

[62] For definitions, refer to Appendix I: Model Governance Definitions.
[63] For definitions, refer to Appendix I: Model Governance Definitions.

Table 60: Governance Documentation of NAIC AI Principle: Ensure Transparency with Appropriate Disclosures, Including Notice to Consumers Specific to Data Being Used and Methods for Appeal and Recourse Related to Inaccurate Data

Are "Ensure Transparency with Appropriate Disclosures, Including Notice to Consumers Specific to Data Being Used and Methods for Appeal and Recourse Related to Inaccurate Data" documented in the governance program?		
Company Operation Area[64]	Number of Companies	
	Yes	No
Rating	29	165
Underwriting	5	189
Claims	24	170
Fraud Detection	14	180
Marketing	10	184
Loss Prevention	1	193
Companywide	52	142
Other	7	187

Table 61: Governance Documentation of NAIC AI Principle: AI Systems Are Secure, Safe, and Robust Including Decision Traceability and Security and Privacy Risk Protections

Are "AI Systems Are Secure, Safe, and Robust Including Decision Traceability and Security and Privacy Risk Protections" documented in the governance program?		
Company Operation Area[65]	Number of Companies	
	Yes	No
Rating	13	181
Underwriting	9	185
Claims	24	170
Fraud Detection	14	180
Marketing	10	184
Loss Prevention	1	193
Companywide	61	133
Other	5	189

Table 62: Companies Following "Other" Existing Standards or Guidance in Regard to a Governance Framework.

Do You Follow Some Other Existing Standards or Guidance in Regard to Governance Framework?		
Number of Companies		
Yes	No	Null
114	22	58

[64] For definitions, refer to Appendix I: Model Governance Definitions.
[65] For definitions, refer to Appendix I: Model Governance Definitions.

Table 63: Source (Internal or External) of "Other" Existing Standards or Guidance in Regard to a Governance Framework

If the company cited it uses "some other existing standards or guidance in regard to a governance framework," are the standards developed internally, provided by a third party, or both?			
Number of Companies			
Internal	External	Hybrid	Null
101	1	16	76

Table 64: Third Parties Used for Standards or Governance Provided in Regard to a Governance Framework

Third Party
Acquisition
Business Roundtable
Casualty Actuarial Society
CCC Intelligence Solutions and Smart Red Flags
COSO Enterprise Risk Management Framework
Ernst & Young
Hover/ PLNAR
IAB, State & Federal
Information Assets
SkopeNow
Verisk Claim Direction

Table 65: Components of Governance Framework Provided by Third Parties

If Any Standards or Guidance are Provided by Third Party, List the Components of the Governance Framework								
Components of Government Framework	Number of Companies							
	Rating	Under-writing	Claims	Fraud Detection	Marketing	Loss Prevention	Company-wide	Other
Accountability, Inventory, Model Lifecycle, Governance, Model Monitoring and Validation, Model Risk Classification	--	--	--	--	--	--	2	--
Actuarial Standards of Practice (ASOPS)	4	--	--	--	--	--	1	--
CCC Intelligence Solutions	--	--	1	--	--	--	--	--

If Any Standards or Guidance are Provided by Third Party, List the Components of the Governance Framework								
Components of Government Framework	Number of Companies							
	Rating	Under-writing	Claims	Fraud Detection	Marketing	Loss Prevention	Company-wide	Other
Governance & Culture, Strategy & Objective Setting, Performance	--	--	--	--	--	--	5	--
https://s3.amazonaws .com/ brt.org/ Business Roundtable_ Artificial Intelligence…	2	2	2	2	2	2	2	2
Support	--	--	--	4	4	--	--	--

Most companies report the governance of models apply to all category types (ML, AI, regression, etc.) of models. Three explanations were provided for why they might differ: 1) “Risk assessment is performed on all models. Model type is a factor considered in the risk assessment. The level of governance is dictated by the risk level assigned to the model.” 2) “Analysis based on loss history.” 3) “Enterprise MRM formally established in 2021. Enterprise MRM Policy including AI/ML MRM approved in 2022.” (Refer to Table 66.)

Table 66: Governance Differs Depending on Model Type

Does Governance Differ Substantially Depending on Model Type?			
Model Type	Number of Companies		
	Yes	No	Null
Rating	12	127	55
Underwriting	13	127	54
Claims	9	132	53
Fraud Detection	8	132	54
Marketing	9	131	54
Loss Prevention	8	132	54
Companywide	19	139	36
Other	9	132	53

THIRD-PARTY DATA SOURCES AND MODELS

Some AI/ML models being used by companies are developed by third parties. Many of these products are used by multiple companies. Risks exist that some “off-the-shelf” tools may not be fully understood by companies and may pose risks to consumers when data is inaccurate. In addition to using third-party models, companies are using big data from third-party data sources.

Third-Party Models Used in Claims

Insurers purchased claims models from the following third parties noted in Table 67.

Table 67: Third Parties' Claims Models Used by Companies

Claims Model Uses[66]	If Model is Developed by a Third Party, List the Third Party Third-Party Name
Evaluation of Images of the Loss	Loveland Innovations
	Hover
	ITEL
	CCC Intelligent Solutions
	PLNAR
	ISO/Verisk
	Shift Technology
	Loveland
	Hosta
	Eagle View
	Betterview
	Adjust Square
Information Resources for Adjusters	Ecopia
	Â
	Loveland Innovations
	Hover
	PLNAR
	ISO/Verisk
	AccuWeather
	NextGear
	NearMap
	Hover Solutions LLC
	Genesys DX
Speech Analysis	Mpathic
	Verint
	Tethr
	Hi Marley
	Five9
	Eleveo
	CX1
	Amazon
	Qualtrics and Genesys
	Qualtrics
	NICE
Other Functions	Â
	Skense
	Verisk
	SkopeNow
	CCC
Subrogation	Â

[66] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Claims Model Uses[66]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
	Optum
	National Subrogation Services (NSS)
	Shift Technologies
	Arrowhead General Insurance Agency, Inc.
Claims Triage	Â
	Verisk
	Assured
Determine Settlement Amount	Hover
	Planar
	Hosta
	Eagle View
	Bees360 (POC)
Litigation Likelihood	Pinpoint Predictive
	Infinilytics

Third-Party Data Sources used in Claims

Insurers purchased data from the third parties in Table 68.

Table 68: Third-Party Claims Data Element Sources Used by Companies

Claims Data Elements[67] If External or Both, List Each Data Vendor
Third-Party Name
Adjust Square
AWS
Betterview
Core Logic
Eagle View
EASI Census Data
Ecopia
Geospatial Intelligence Center (GIC)
Hosta
Hover
HRRR
IMGING
Insurance Score
ISO/Verisk
Loveland Innovations
NICB
NOAA
Pinpoint Predictive
PLNAR

[67] For definitions, refer to Appendix H: Data Use Table Definitions.

Claims Data Elements[67] If External or Both, List Each Data Vendor
Third-Party Name
PLRB
Shift Technology

Third-Party Models Used in Fraud Detection

Insurers purchased fraud detection models from the third parties listed in Table 69.

Table 69: Third Parties' Fraud-Detection Models Used by Companies

Fraud-Detection Model Uses[18]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
Fast-Tracking of Likely Non-Fraudulent Claims	Â
	FRISS
	ISO/Verisk
	Shift Technology
Referral of Claims for Further Investigation	Â
	CCC
	FRISS
	IBM
	ISO/Verisk
	Shift Technology
Fraudulent Quote Detection	Shift Technologies
Evaluation of Potential for Intentional Infliction of Damage	Shift Technology
Detect Medical Provider Fraud	Â
	ISO
	Shift Technology
Detect First-Party Liability	IBM
	Shift Technology
Social Network Analysis	Â
	Shift Technology
	SkopeNow
	Social Discovery
Organized Crime Rings Detection	Â
	IBM
	ISO
	Shift Technology
Detect Third-Party Liability	IBM
	Shift Technology
Other Fraud Detection-Related Functions	Cape Analytics
	IBM

Fraud-Detection Model Uses[18]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
	ISO/Verisk
	Shift Technology

Third-Party Data Sources Used in Fraud Detection

Insurers purchased data from the third parties in Table 70.

Table 70: Third-Party Fraud-Detection Data Element Sources Used by Companies

Fraud-Detection Data Elements[68] If External or Both, List Each Data Vendor
Third-Party Name
ISO/Verisk
NICB
Shift Technology
RealtyTrac
ISO Claim Director
TransUnion
Weatherstack
Geocoder
Social Discovery
ScopeNow
Precisely
LexisNexis
US Census
Realtor
Insurance Score

Third-Party Models Used in Marketing

Insurers purchased marketing models from the third-party vendors listed in Table 71.

Table 71: Third Parties' Marketing Models Used by Companies

Marketing Model Uses[69]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
Targeted Online Advertising	Google
	Facebook
	The Trade Desk
	Universal McCann

[68] For definitions, refer to Appendix H: Data Use Table Definitions.
[69] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Marketing Model Uses[69]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
	Microsoft
	Meta
	LinkedIn
	NextDoor
	PegaSystems
	IBM
	AT&T
	Instagram
	Bing
	Twitter
	Seismic
	Salesforce
	Genus AI
Identification of Recipients of Mail or Phone Advertising	DataLab
	EXL
	Pega Systems
	Ameriprise
Media Mix Modeling	IPSOS MMA
	Neustar
	AARP
	Xandr
	The Trade Desk
	Meta
	Google
	Nielsen
	Marketing Evolution
	Kantar
Identification of Potential Customer Groups	Facebook
	Google
	Nextdoor
	Â
	Acxicom
	Instagram
	Deloitte
	AT&T
	PegaSystems
	EXL
	Chadwick Martin Bailey (CMB)
	Microsoft
Customer Acquisition and Retention	Â
	Acxicom
	PegaSystems
	Rocket Referrals
	Kantar
Direct Online Sales	Marketing Evolution
Fraud Detection	Yext, Inc.

Marketing Model Uses[69]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
Provision of Offers to Existing Customers	PegaSystems
	IBM
Demand Modeling	Google
	Marketing Evolution
	Microsoft
	Elsy
Interactions Using NLP	Voci
	Qualtrics
	Nuance
	Invoca
	RepRIsk AG
	Clarabridge
	Drips
	Verint
	Genesys
	https://www.ada.cx/
Other Marketing-Related Functions	IPSOS
	Human
	Zefr
	iSpot/Blockgraph
	Fenestra

Third-Party Data Sources Used in Marketing

Insurers purchased data from the third parties listed in Table 72.

Table 72: Third-Party Marketing Data Element Sources Used by Companies

Marketing Data Elements[70] If External or Both, List Each Data Vendor
Third-Party Name
Acxiom
Amerprise Advisor Information
ATTOM
Cape Analysis
Claritas
LexisNexis
Costco
Datalab
EASI
Experian
Google
Guy Carpenter
Insurance Bureau

[70] For definitions, refer to Appendix H: Data Use Table Definitions.

Marketing Data Elements[70] If External or Both, List Each Data Vendor
Third-Party Name
Insurance Score
IPSOS MMA
LinkedIn
Merkle
Meta
Neustar
Realtor
Rocket Referrals
Seismic
The Trade Desk
TransUnion
US Census
Verisk
Xandr

Third-Party Models Used in Rating

Insurers purchased "more advanced AI/ML" rating models from the third-party vendors listed in Table 73.

Table 73: Third Parties' Rating Models Used by Companies

Rating Model Uses[71]	If Model is Developed by a Third Party, List the Third Party
	Third-Party Name
Rating Class Determination	Cape Analytics
	Precisely
Other Functions	Cape Analytics
	CoreLogic
Retention Modeling	Willis Towers Watson

Third-Party Data Sources Used in Rating

Insurers purchased data from the third parties listed in Table 74.

Table 74: Third-Party Rating Data Element Sources Used by Companies

Rating Data Elements[72] If External or Both, List Each Data Vendor
Third-Party Name
AIR
Applied Geographic Solutions
CAPE Analytics

[71] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.
[72] For definitions, refer to Appendix H: Data Use Table Definitions.

Rating Data Elements[72] If External or Both, List Each Data Vendor
Third-Party Name
Census Bureau
CLUE
CoreLogic
EASI
Ecopia
Epsilon
Equifax
GIC
ISO/Verisk
LexisNexis
MSB
National Agriculture Imagery Program
National Oceanic and Atmospheric Administration
NearMap
Oak Ridge National Laboratory
Pitney-Bowes
Precisely
Property and Liability Resource Bureau (formerly AIR)
Realtor
US Department of Agriculture
US Geological Survey
Zesty

Third-Party Models Used in Underwriting

Insurers purchased underwriting models from third-party vendors listed in Table 75.

Table 75: Third Parties' Underwriting Models Used by Companies

Underwriting Model Uses[73]	If Model is Developed by a Third-Party, List the Third Party
	Third-Party Name
Automated Approval	Shift Technology
	Pinpoint Predictive
Automated Denial	Cape Analytics
	CoreLogic
	Zesty.ai
	Shift Technology
Motor Vehicle Record Reordering	Explore
Policy Anomaly Detection	Pinpoint Predictive
	Cape Analytics

[73] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

Underwriting Model Uses[73]	If Model is Developed by a Third-Party, List the Third Party
	Third-Party Name
Underwriting Tier Determination	Cape Analytics
Input Into Non-Automated Approval Decision	Precisely
	Cape Analytics
	Nearmap
	Zesty.ai
	Shift Technology
Input Into Non-Automated Denial Decision	Cape Analytics
	Precisely
	Verisk
	Nearmap
	CoreLogic
	Zesty.ai
	Shift Technology
	Experion
Automate Processing Through the Agency Channel	Cape Analytics
Renewal Reinstatement	Cape Analytics
	Zesty
	Nearmap
	Gic/Vexcel
	Zesty.ai
	Shift Technology
Verification of Policy Characteristics	Verisk
	Nearmap
	Cape Analytics
	Gic/Vexcel
	Flyreel
Other Underwriting-Related Functions	Shift Technology
	Eagleview
	Zesty.ai
	Better View

Third-Party Data Sources Used in Underwriting

Insurers purchased data from the third parties listed in Table 76. (See Table 76.)

Table 76: Third-Party Underwriting Data Element Sources Used by Companies

Underwriting Data Elements[74] If External or Both, List Each Data Vendor
Third-Party Name
AIR
Arturo
AWS

[74] For definitions, refer to Appendix H: Data Use Table Definitions.

Underwriting Data Elements[74] If External or Both, List Each Data Vendor
Third-Party Name
Better View
Cape Analytics
Carfax
CLUE
Core Logic
EASI Census
Exopia
Explore
Flyreel Inc.
Geospatial Intelligence Center (GIC)
HLDI
LexisNexis
Near Map
NOAA
Pinpoint Predictive
Precisely
Realtor
Shift Technology
Transunion
US Census
Verisk
Wondour
Zesty

Third-Party Models Used in Loss Prevention

Insurers purchased loss prevention models from third-party vendors listed in Table 77.

Table 77: Third Parties' Loss Prevention Models Used by Companies

Loss Prevention Model Uses[75]	If Model is Developed by a Third-Party, List the Third Party
	Third-Party Name
Identification of High-Risk Customers	Cape Analytics
	Shift Technology
Risk-Mitigation Advice to Consumers	Flyreel
Guidance of Loss Control Inspections	CCC
	Ernst & Young
	Cape Analytics
	Verisk
	Flyreel
	Betterview

[75] For definitions, refer to Appendix A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention.

<u>Third-Party Data Sources Used in Loss Prevention</u>

Insurers purchased data from the third parties listed in Table 78.

Table 78: Third-Party Loss Prevention Data Element Sources Used by Companies

Loss Prevention Data Elements[76] If External or Both, List Each Data Vendor
Third-Party Name
AWS
Cape Analytics
CoreLogic
EASI Census
GIC
ISO/Verisk
Lexis Nexis
Realtor
TransUnion
US Census

CONCLUSION/NEXT STEPS

As requested by the SME group, the NAIC's technical team completed an analysis of the data submitted in the Home AI/ML survey. Insight was gained around the general use of AI/ML by insurance companies, uses of AI/ML in insurance company operations, data elements and sources used in insurance company operations, governance frameworks and documentation, consumer data recourse, and third-party sources for AI/ML models and/or data.

The insight gained from the survey will be used to supplement regulators' knowledge of the current regulatory framework around AI/ML, governance, consumers, and third parties and to evaluate whether any changes should be made to the regulatory frameworks.

The SME group, other regulators, and NAIC staff have identified some potential next steps, including many activities already in progress. The following list of next steps is not intended to be complete, but it may be helpful as a starting point for discussions and decision-making about what next steps to take at the NAIC:

- Evaluate the survey analysis and determine whether to further explore the following subjects:
    - Company AI/ML model usage and the level of decision-making (i.e., the amount of human involvement in decision-making).
    - Company data elements.
    - Companies' governance frameworks and the documentation of such.
    - Consumer data recourse.
    - Third-party regulatory framework.
- Create a risk hierarchy to prioritize the need for more model governance and company oversight. The general concept is that more oversight of a model will be needed as the consumer risk or impact increases from the modeling or models.

[76] For definitions, refer to Appendix H: Data Use Table Definitions.

- Evaluate consumer data recourse. Companies report a wide variety of methods for consumers to evaluate and correct data used by companies. Some methods are short and easy, such as using an app to correct data, and other methods are more time consuming and require personal contact with the agent or company. In some cases, consumers may not even know about their data being used, so consumer transparency is a priority. *(Privacy Protections (H) Working Group)*
- Evaluate the regulatory framework around the use of third-party models and third-party data. Evaluate the ability of companies and regulators to obtain needed information from third parties and for regulators to oversee this work either through the companies or third parties in some way. *(Workstream Two of the Big Data and Artificial Intelligence (H) Working Group)*
- Evaluate concerns about third-party concentration by company use. *(Workstream Two of the Big Data and Artificial Intelligence (H) Working Group)*
- Determine whether additional best-practices white papers would be useful on subjects in the AI/ML space.

APPENDIX A: Guidance for Questions in Each Operational Area: Rating, Underwriting, Claims, Fraud Detection, Marketing, Loss Prevention

For the purposes of this survey, the insurer operational areas are rating, underwriting, claims, fraud detection, marketing, and loss prevention. The respondent can add other operational areas in parts of the survey.

Level of AI/ML Deployment

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.

Level of Decisions Influenced by AI/ML

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.

AI/ML Model Category Types

For each of the AI/ML operational areas, there is a question asking the respondent to select whether a listed model is AI or a specific type of ML. If the model employs more than one type, mark all types 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 select more than one method.

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 above model category types.

APPENDIX B: Definitions Specific to Claims

The following are the definitions specific to claims:

- **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.

APPENDIX C: Definitions Specific to Fraud Detection

The following are the definitions specific to fraud detection:

- **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.

APPENDIX D: Definitions Specific to Marketing

The following are the definitions specific to Marketing:

- **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.

APPENDIX E: Definitions Specific to Rating

The following are the definitions specific to rating:

- **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.

APPENDIX F: Definitions Specific to Underwriting

The following are the definitions specific to underwriting:

- **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.

APPENDIX G: Definitions Specific to Loss Prevention

The following are the definitions specific to loss prevention:

- **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.

APPENDIX H: Data Use Table ("Data Elements") Definitions

The data elements are located 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**

APPENDIX I: Model Governance Definitions

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 the 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).