

Draft Pending Adoption

Draft: 8/28/22

Big Data and Artificial Intelligence (H) Working Group
Portland, OR
August 10, 2022

The Big Data and Artificial Intelligence (H) Working Group of the Innovation, Cybersecurity, and Technology (H) Committee met in Portland, OR, Aug. 10, 2022. The following Working Group members participated: Elizabeth Keller Dwyer, Chair (RI); Amy L. Beard, Co-Vice Chair, represented by Victoria Hastings (IN); Doug Ommen, Co-Vice Chair (IA); Adrienne A. Harris, Co-Vice Chair, represented by Sumit Sud (NY); Kevin Gaffney, Co-Vice Chair (VT); Anna Latham (AK); Ken Allen (CA); Peg Brown (CO); George Bradner and Wanchin Chou (CT); John Reilly (FL); Weston Trexler and Randy Pipal (ID); Erica Weyhenmeyer (IL); Rob Roberts (KY); Tom Travis (LA); Rachel Davison (MA); Kathleen A. Birrane and Alexander Borkowski (MD); Sandra Darby (ME); Karen Dennis (MI); Phil Vigliaturo (MN); Cynthia Amann (MO); Tracy Biehn and Angela Hatchell (NC); Jon Godfread and Chris Aufenthie (ND); Christian Citarella (NH); Barbara D. Richardson (NV); Tom Botsko (OH); Alex Cheng (OR); Shannen Logue and Michael McKenney (PA); Michael Wise (SC); Travis Jordan (SD); Carter Lawrence (TN); Leah Gillum (TX); Jon Pike (UT); Katie Johnson (VA); Molly Nollette (WA); Rebecca Rebholz (WI); and Greg Elam (WV).

1. Adopted its July 14 Minutes

The Working Group met July 14 and took the following action: 1) adopted its April 5 minutes; and 2) received reports from its four workstreams.

Commissioner Ommen made a motion, seconded by Mr. Vigliaturo, to adopt the Working Group's July 14 minutes (Attachment Two-A). The motion passed unanimously.

2. Held the Collaboration Forum on Algorithmic Bias

Superintendent Dwyer said one of the highest priority charges to the Innovation, Cybersecurity, and Technology (H) Committee is to facilitate appropriate levels of coordination and collaboration among NAIC working groups with respect to topics that relate to innovation, technology, cybersecurity, and privacy. This is being accomplished through the Collaboration Forum, which has an initial focus on algorithmic bias. Superintendent Dwyer said the Big Data and Artificial Intelligence (H) Working Group is playing a role in providing a public forum for education and discussion.

Scott Kosnoff (Faegre Drinker Biddle & Reath) talked about how insurers that are using artificial intelligence (AI) or algorithmic decision making can protect themselves from regulatory litigation and reputational risk. Mr. Kosnoff said AI tools can assist in customer engagement, rating, underwriting, claims management, and fraud detection and that insurers invest in these tools for better decision making and to remain competitive. While there are many benefits to the use of these tools, Mr. Kosnoff said state insurance regulators have expressed concerns about fairness, unintended discrimination, and lack of transparency.

Mr. Kosnoff said concerns surrounding the use of AI are not unique to the insurance industry, and he then reviewed steps insurers could take to minimize their risks. He said insurers should have a risk management framework, which includes how to identify potential bias. Mr. Kosnoff said insurers have a difficult situation because there is no clear guidance regarding what level of correlation of a data variable with a protected class is

Draft Pending Adoption

acceptable and under what circumstances. He said an insurer should take reasonable steps to identify, manage, and mitigate the risk of negative outcomes. Mr. Kosnoff said the risk management framework should be an extension of an insurer's enterprise risk management (ERM) and include the following: 1) written policies and procedures; 2) clear assignment of responsibility and accountability; 3) communication of policies; 4) training and supervision; 5) consistent use and application; 6) monitoring and corrective action, if needed; and 7) documentation.

Mr. Kosnoff said an insurer needs to inventory the organization's algorithms, understand each algorithm's objective and how it will be used, identify potential risks for each algorithm, assess the seriousness of potential harm that could occur, and the likeliness of harm occurring. After this, he said an insurer should implement appropriate safeguards, including humans having some level of oversight. Finally, Mr. Kosnoff said insurers should consider testing for bias and said he thinks this is the only part of the risk management framework that is controversial.

Mr. Kosnoff said an insurer should have a multidisciplinary team, diversity, clear assignment of roles, and board oversight for the development of AI and risk management framework. He said the risk management framework should be proportionate to the likelihood of harm and should cover every stage of the AI life cycle, which includes pre-design, design and development, testing and evaluation, deployment, and monitoring. In response to a question regarding how to monitor an algorithm that can learn over time, Mr. Kosnoff said it is important to recognize testing cannot occur only at initial deployment.

Eric Krafcheck (Milliman) said his comments today are through the lens of how to evaluate a rating plan but are applicable to other ways models are used. He said his comments focus on the testing for bias and said one needs to first identify the scope of analysis, define the fairness measure, and then collect the necessary data to test for bias, which includes protected class data. Mr. Krafcheck said detecting bias for insurance is complicated because insurance is centered on the concept of fair discrimination by treating different risks differently. Because of this, he said any methodology should use multivariate methods.

Mr. Krafcheck said the type of model used and how the model is used will dictate what methodology to use for testing of bias. For example, he said it is easy to tell if a facial recognition model is working as intended, whereas it is more difficult to tell if an insurance pricing model is working as intended. Mr. Krafcheck said the testing method may vary if one is testing for overall model prediction versus testing for disparate impact. He said the volume of data and granularity of protected class data available will also affect testing methodology. Mr. Krafcheck said it is important to establish the questions one is seeking to answer to determine the best testing methodology. He provided an overview of the following potential testing based on the question being asked.

- Is the model/variable a proxy? Mr. Krafcheck suggested the use of a control variable test, which involves adding a protected class variable as a predictor in the model to account for the predictive effect of the protected class. One can then compare the model output before and after the protected class variable is used in the model to see if there is a material difference. Mr. Krafcheck used the hypothetical example of vehicle color being correlated to risk of loss and vehicle color also being corrected to a protected class. The observable difference in how losses vary across vehicle color after accounting for the protective class would identify whether there is a proxy effect through the use of vehicle color.

Draft Pending Adoption

- Is the predictive effect consistent across protected classes? Mr. Krafcheck suggested the use of an interaction test, which involves adding the protected class variable as an interaction in the model. The output of the model will produce separate estimates for each protected class for that variable. One can then compare the effect across different protective classes to see if there is consistency. Mr. Krafcheck said this test will help determine if the predictive effect is consistent across protected classes.
- Does the inclusion of a variable disproportionately affect otherwise similar risks? Mr. Krafcheck suggested the use of a nonparametric matching test. For every policyholder of a given protected class, this test involves matching a policyholder who is not of the same protected class but has similar risk characteristics for all variables except for the variable being evaluated. Mr. Krafcheck said one would then take two matched data sets and compare average model predictions by protection class by including the evaluated variable and then excluding the evaluated variable. He said one could then add loss experience to determine whether the actual loss experience supports the relationship seen for the variable being evaluated and is within the parameters of defined fairness.
- Does the variable improve predictions across protected classes? Mr. Krafcheck suggested the use of a double life chart test, which is a test that modelers use as they are developing models. This test involves including a variable of interest and then excluding that variable to compare which model lines up better with actual experience. Mr. Krafcheck said this allows one to assess whether the model is consistently improving across protected classes.

Mr. Krafcheck said there are other tools that could be used, and one needs to balance between statistical significance (whether the result of the analysis occurred randomly) and practical significance (whether the result of the analysis has real-world relevance). He said the statistical significance will depend on whether one is using a large or small data set. Mr. Krafcheck also said the results of any analysis may be limited by the available data and that some data may have some sort of selection bias. He said one might also want to consider the intersectionality of protected classes, such as looking at race by gender, and consider looking at groups of variables because there might not be any bias at an individual variable level. In response to a question regarding the importance of having data about loss experiences for protected classes, Mr. Krafcheck said this data is important to any analysis.

3. Received Reports from Workstream One and Workstream Two

a. Workstream One

Commissioner Gaffney said the subject matter experts (SMEs) of Workstream One have been collaborating with multiple states to conduct AI/machine learning (ML) surveys for private passenger automobile (PPA) insurance, homeowners insurance, and life insurance. The goals of these surveys are to: 1) identify the current level of risk and exposure from the use of AI and whether or how the industry is managing or mitigating that risk; 2) develop information for trending, such as how the risk is evolving over time, and the industry's responsive actions; and 3) inform a meaningful and useful regulatory approach, framework, and/or strategy for overseeing and monitoring this activity.

Commissioner Gaffney said the AI/ML PPA survey was conducted last year, and NAIC technical staff produced a confidential data analysis report for the AI/ML PPA survey on June 30. The SMEs been studying this report and plan to present their public PPA survey report to this Working Group at the Fall National Meeting.

Draft Pending Adoption

Commissioner Gaffney said the homeowners survey was exposed for comment. He said no comments were received by the comment deadline of Aug. 1. However, pilot companies have provided feedback to improve the survey. Commissioner Gaffney said the survey will be completed soon, and the NAIC will then program the survey into its systems. Once programmed, the 10 states conducting the survey will formally issue their market conduct data call. Companies will likely be asked for responses within 30 days from the issuance of the formal data call.

Commissioner Gaffney said the AI/ML life insurance survey is being developed, and outreach to pilot companies has been completed. Similar to the other survey, the pilot companies have provided useful feedback to improve the survey. The life insurance survey should be exposed by the end of August for a 30-day public comment period. Commissioner Gaffney said the life insurance survey will not be issued until the homeowners survey is completed.

b. Workstream Two

Commissioner Ommen said Workstream Two is leveraging the results of the PPA survey, and the discussions for this workstream have remained regulator-to-regulator since the SMEs have been discussing specific third-party vendors identified through the PPA survey. Commissioner Ommen said recent discussions of the Workstream members have included several potential initial steps for enhanced regulatory oversight of third-party vendors and models. One potential initial approach discussed was that state insurance regulators require contracting insurance companies to obtain verification from their third-party vendor that an actuary/other professional has reviewed a model prior to using the model certifying its compliance with AI standards. Another initial regulatory option discussed was the concept that state insurance regulators create a library of third-party vendors, classified by line of insurance and operational area. This Workstream has a deadline of the Fall National Meeting to report its findings to the Big Data and Artificial Intelligence (H) Working Group with suggestions for implementation of deliverables to be considered by the Innovation, Cybersecurity, and Technology (H) Committee.

David Snyder (American Property Casualty Insurance Association—APCIA) said he supports the approach of determining appropriate definitions for AI. Mr. Snyder said a fundamental issue is how this work fits within the risk-based pricing framework and the regulatory standard of insurance rates not being excessive, inadequate, or unfairly discriminatory.

Scott Harrison (American InsurTech) said he agrees with Mr. Kosnoff's comments and the need for risk management for the use of AI and technology. Mr. Harrison said companies need to understand what they are doing, state insurance regulators need to have confidence that companies understand what they are doing, and consumers need to have confidence in the use of AI and technology. Mr. Harrison said this requires state insurance regulators establishing uniform standards and the ability of state insurance regulators to assess compliance with these standards.

Draft Pending Adoption

Birny Birnbaum (Center for Economic Justice—CEJ) suggested workstream two address antitrust issues. He said the use of big data and AI is different from past insurer practices and is a revolution in how companies operate. Mr. Birnbaum said he agrees a risk management framework is necessary but said this is not sufficient. He said testing of consumer outcomes is essential. Mr. Birnbaum said the insurance industry is not immune from the impact of structural racism and the data sources and algorithms that perpetuate structural racism. He said actuarial fairness is one standard but that state insurance regulators should also test for racial bias, proxy discrimination, and disparate impact. Mr. Birnbaum said there is a need for a uniform approach that can be measured across insurers and across lines of business.

Having no further business, the Big Data and Artificial Intelligence (H) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/H CMTE/2022_Summer/_BDAIWG/BD AI Minutes 8.10.22.docx

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

This document contains questions that regulators can ask about any models and data used by insurance companies, whether that model or data is developed internally or obtained from external sources. The questions are separated into three sections: 1) Questions to a) Insurers about their own models and b) Insurers or 3rd party (depending on your state's procedures) about the 3rd party model; 2) questions to ask insurers about the use of 3rd party models and data inputs into such models, and 3) questions to ask insurers about the use of 3rd party data.

The depth of study of the model can vary depending on the underlying reasons for the study, therefore we have separated questions into "Main General Questions" and "Detailed and Technical Questions." The "Main General Questions" section includes a list of suggested questions to obtain a high-level understanding of the model. The "Detailed and Technical Questions" section expands on the prior section by including additional details and questions to obtain a more in-depth understanding of the model. Categorizations of the general questions (where underlined) help to match the general questions to the detailed questions. These categorizations are also intended to aid different users in the selection of appropriate questions given the intent of the study. The examiner should evaluate regulatory purpose and use or modify questions for such purpose. A regulator does not need to ask every question for every regulatory interaction. A regulator may also find it helpful to use guidance in the *NAIC Market Regulation Handbook* and specifically targeted questions adopted by other committee groups.

A "Definitions" section is also included at the end of the document to provide clarification regarding some key terms used throughout the document.

I. MAIN GENERAL QUESTIONS

A. QUESTIONS TO 1) INSURERS ABOUT THEIR OWN MODELS AND 2) INSURERS OR 3rd PARTY (DEPENDING ON YOUR STATE'S PROCEDURES) ABOUT THE 3rd PARTY'S MODEL

1. Overview of the Model and Business Purpose: Describe the model and its intended purpose, including how artificial intelligence is used.
2. Data Inputs: Provide a list of all data and information used in the model with their corresponding sources. Describe any additional use of the data other than the primary purpose of using the data in the model.
3. Model Assumptions and Outcomes: Describe the material assumptions made in the construction of the models or for modeling purposes. Were any changes made to the model output, such as deviations from the model indications/output (by either a 3rd party or insurer)? If so, explain.
4. Model Testing/Validation: Describe how the model was tested and validated. Identify who conducted the testing and validation and their qualifications.

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

5. Governance: Submit a written governance framework and controls that apply to the model. (If the model is from a 3rd party, include the 3rd party's governance framework and the insurer's governance framework.)
6. Consumer Protection and Access: Explain how applicants, customers, or claimants are made aware of 1) the data and information collected and its use; 2) how they can obtain their records and correct any errors; and 3) any time limitations or parameters imposed on the ability to correct the data and information. (If this is not required for a particular model, so state.)

B. QUESTIONS TO INSURERS ABOUT USE OF 3RD PARTY MODELS

Questions in the previous section should be answered by the 3rd party and/or insurer. The following are the additional questions to ask the insurer about its use of a 3rd party's model.

1. Overview of the Model and Business Purpose: 1) Identify the entity(ies) who developed the model, their qualifications, and who the insurer contracted with for its use. 2) Based on the use identified in the previous section about the 3rd party's model, describe how the 3rd party model is used and whether that use aligns with the 3rd party's suggested appropriate uses. 3) Describe the testing performed to evaluate whether the model is appropriate for the insurer's book of business.
2. Model Testing/Validation: Describe the due diligence of testing the 3rd party's model and data for reliability and accuracy.
3. 3rd Party Contract: Provide the purchase contract with the 3rd party owner of the model and highlight the contractual terms related to the model's use.

C. QUESTIONS TO INSURERS ABOUT 3RD PARTY DATA PURCHASE

1. Overview of the Data: Provide a list of all data and information purchased to be used in a model with their corresponding sources. Describe any additional use of the data other than the primary purpose of using the data in the model.
2. Data Testing/Validation: Describe how the insurer vetted the 3rd party data for errors. Explain any significant missing data (e.g., roof condition is not available for houses with significant tree overhang; vehicle use is not included for all people in the data).
3. Governance: Submit a written data governance framework and controls for the insurer and 3rd party.
4. Consumer Protection and Access: Explain how applicants, customers, or claimants are made aware of 1) the data and information collected and its use; 2) how they can obtain their records and correct any errors; and 3) any time limitations or parameters imposed on the ability to correct the data and information.

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

5. 3rd Party Contract: Provide the purchase contract with the 3rd party owner of the data and highlight the contractual terms related to the data's use.

II. DETAILED AND TECHNICAL QUESTIONS

A. QUESTIONS TO INSURERS AND 3RD PARTY (DIRECTLY) ABOUT OWN MODELS

1. Overview of the Model and Business Purpose:

- a. To describe the model, submit the following model documentation (at a minimum):
 - i. Model Type (GLM—Generalized Linear Model, GAM—Generalized Additive Model, Neural Network, etc.).
 - ii. If applicable: Sub model Type/s (GLM, GAM, Neural Network, etc.).
 - iii. Identify the modeler(s), company affiliation, and qualifications.
 - iv. Identify the software used to develop the model.

- b. To describe the intended purpose, including how artificial intelligence is used, submit the following documentation (at a minimum):
 - i. Identify the specific use and the company operations (e.g., Claims operations: used to identify which claim's staff should be assigned the claim). See the NAIC's AI/ML surveys for some "uses of model" examples.
 - ii. Are there secondary purposes or other potential uses of the model? If so, explain.
 - Questions to Insurers only: Will you be using the model output to address other issues or areas within the Company? If so, explain. Would the model output be appropriate for the secondary purpose? If so, explain. Would the data or results be considered reliable for this secondary use? If so, explain.

- c. (Questions for Insurers only): Identify whether the model is new or an update to a model already used by the insurer. Identify who at your company inputs the data, runs the model, and checks the output and provide their qualifications.
 - i. If the model is an updated version, identify the issues addressed in the update and how the updated model addresses the related issues.

2. Data Inputs

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

- a. To describe the data and information used, submit the following documentation (at a minimum):
 - i. A data dictionary, including all variable names (plus all variables that were combined into a new variable), data sources (if external), and data types
 - Identification of the sources for the data, identify as internal/external, and include the sources' company names or other identifier. Explain how the data was collected (e.g., industry compiled, surveys, internal claims system)
 - Identify any unique technology used to collect the data (e.g., cellphones, Open Data Connectivity Devices) and issues around the use of such (e.g., may be a passenger in the car and not the driver).
 - ii. Years of data collected and locations included (e.g., states, country) with corresponding data distributions (e.g. exposures by year/state), exclusions or non-use of specific data that would otherwise fit the description (e.g. excluded catastrophe losses), and any exceptions to that description (e.g., did not have sufficient data in one state).
 - iii. Describe the process used to determine whether the data is appropriate for use and/or fit for use.
 - iv. Provide a rational explanation for any data and information that could be related to any protected class or socioeconomic status.
 - v. Explain any significant missing data (e.g., roof condition is not available for houses with significant tree overhang; vehicle use is not included for all people in the data).
 - vi. Discuss how data outliers were handled. How was the data tested for outliers (both response and predictor variables)? Were any outliers removed or capped? If so, explain why.
 - vii. Identify how frequently the data will be updated (e.g., yearly, weekly, real-time).
3. Model Assumptions and Outcomes:
- a. Describe the material assumptions made in the construction of the models or for modeling purposes.
 - b. Were any changes made to the model output, such as deviations from the model indications/output (by either a 3rd party or insurer)? If so, explain.
 - c. Identify the model outputs / target variables.
 - d. Provide model specifications (e.g., link function, distribution, final hyperparameters, tuning process).
 - e. Describe any dimensionality reduction techniques used.

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

- f. To describe deviations from the model indications/outcome, provide current indicated versus selected tables or similar type of explanation. If applicable, provide the dislocation from the current model to the new model. Were appropriate adjustments made? Describe how model results were adjusted to mitigate the largest effects of the model.
4. Model Testing/Validation:
- a. To describe testing and validation, submit the following documentation (at a minimum):
 - i. Identify the training, testing, and hold-out data (e.g., distribution of data used included in each subset).
 - ii. Describe any sensitivity testing and results.
 - iii. Explain how formula and code was checked for accuracy.
 - iv. Provide the most recent validation and audit (internal and external) reports.
5. Consumer Protection and Access:
- a. Explain how applicants, customers, or claimants are made aware of 1) the data and information collected and its use; 2) how they can obtain their records and correct any errors; and 3) any time limitations or parameters imposed on the ability to correct the data and information. (If this is not required for a particular model, so state.)
 - b. If a consumer's data was provided to another party for model design, is the consumer's data useable for other purposes or provided to other companies, or does the contract restrict such use?
6. Governance
- b. Submit a written governance framework and controls that apply to the model. (If the model is from a 3rd party, include the 3rd party's governance framework and the insurer's governance framework.). The governance framework and controls should include the following (at a minimum):
 - i. What training requirements are implemented regarding creation or use of models.
 - ii. Explain security of the model and data. How is access to the model and data controlled? Can individuals access the AI system and tamper with results or the processes? If so, explain.
 - iii. Identify the scope and process for validity testing. Describe procedures designed to reduce the risk of inaccurate or biased models.

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

- iv. Explain how the implemented model is tracked to evaluate the accuracy of results. Identify how often the model will be updated and reevaluated for effectiveness, efficiency, and appropriate use and how updates of the model or rollouts of the model will be handled.
- v. Explain how the governance policy is implemented, monitored, and audited.
- vi. Explain how the *NAIC AI Principles* (see “Governance Framework and Controls” in the definitions section) are met in the governance framework.

B. QUESTIONS TO INSURERS ABOUT USE OF 3RD PARTY MODELS

Questions in the previous section should be answered by the 3rd party and/or insurer. The following are the additional questions to ask the insurer about its use of a 3rd party’s model.

1. Overview of the Model and Business Purpose:
 - a. Identify the entity(ies) who developed the model, their qualifications, and who the insurer contracted with for its use.
 - b. Based on the use identified in the previous section about the 3rd party’s model, describe how the 3rd party model is used and whether that use aligns with the 3rd party’s suggested appropriate uses.
 - c. Describe the testing performed to evaluate whether the model is appropriate for the insurer’s book of business.
 - d. Describe the testing to evaluate any overlap with other models or double counting (e.g., using a variable in a rating model that seems to be assessing the same risk as a rating variable outside the model in another part of the rating algorithm).
2. Model Testing/Validation
 - a. Describe the due diligence of testing the 3rd party’s model and data for reliability and accuracy.
3. 3rd Party Contract
 - a. Provide the purchase contract with the 3rd party owner of the model and highlight the contractual terms related to the model’s use.

C. QUESTIONS TO INSURERS ABOUT 3RD PARTY DATA PURCHASE

1. Overview of the Data
 - a. To describe the data and information purchased, submit the following documentation (at a minimum):
 - i. Data categories (or variables) and data types

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

- ii. Description of data: should include the years of data collected, locations included (e.g., states, country), exclusions or non-use of specific data that would otherwise fit the description (e.g., excluded catastrophe losses), and any exceptions to that description (e.g., did not have sufficient data in one state).
 - iii. Identify the data sources and include the sources' company names or other identifier. Explain how the data was collected (e.g., industry compiled, surveys, internal claims system)
 - iv. Identify any unique technology used to collect the data (e.g., cellphones, Open Data Connectivity Devices) and issues around the use of such (e.g., may be a passenger in the car and not the driver).
 - v. Describe the intended use of the data. How was it determined that the data is fit for its intended use?
 - vi. Identify how frequently the data is received from the 3rd party and how frequently the new data will be updated in the model (e.g., yearly, weekly, real-time).
2. Data Quality
- a. Describe how the insurer vetted the 3rd party data for errors.
 - b. Explain any significant missing data (e.g., roof condition is not available for houses with significant tree overhang; vehicle use is not included for all people in the data).
3. Consumer Protection and Access
- a. Explain how applicants, customers, or claimants are made aware of 1) the data and information collected and its use; 2) how they can obtain their records and correct any errors; and 3) any time limitations or parameters imposed on the ability to correct the data and information.
4. Governance
- a. Submit a written data governance framework and controls for the insurer and 3rd party. The data governance framework and controls should include the following (at a minimum):
 - i. Explain security of the data. How is access to the data controlled? Can individuals access the data and tamper with the data? If so, explain.
 - ii. Identify the scope and process for validity testing. Describe procedures designed to reduce the risk of inaccurate or biased data.
 - iii. Identify how often the data will be updated and reevaluated for effectiveness, efficiency, and appropriate use and how updates of the data or initial rollouts of the data will be handled.
 - iv. Explain how the governance policy is implemented, monitored, and audited.

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

- v. Explain how the NAIC AI principles (see “Governance Framework and Controls” in the definitions section) are met in the data governance framework.
5. 3rd Party Contract:
 - a. Provide the purchase contract with the 3rd party owner of the data and highlight the contractual terms related to the data’s use.
 - b. Describe any contractual terms and limitations regarding the use of the data.
 - c. Describe the terms of purchase for the 3rd Party Data, specifically any contractual terms and limitations regarding the use of the data.

III. Definitions

The following definitions are provided but may be modified by the Innovation, Cybersecurity, and Technology (H) Committee in the future.

Artificial Intelligence (AI): The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings¹. It describes an automated process in which a system begins recognizing patterns without being specifically programmed to achieve a pre-determined result.

Bias: Bias is an effect that deprives a statistical result of representativeness by systematically distorting it, as distinct from a random error, which may distort on any one occasion but balances out on the average². Different kinds of bias include systemic, human, and statistical/computational. Please refer to “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence”³ for a description of different types of bias.

Governance Framework and Controls: A structure covering practices, guidance, and validation of models. It covers, but is not limited to, the following:

- **Fairness and Ethics Considerations:** Ensuring responsible adherence to fairness and ethical considerations. 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.

¹ <https://www.britannica.com/technology/artificial-intelligence>

² OECD, “Glossary of statistical terms,” OECD Online Resource, July 2007, <https://stats.oecd.org/glossary/detail.asp?ID=3605>

³ <https://arxiv.org/abs/1901.10002>

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

- **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:** 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.
- **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 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.

Please refer to the NAIC Principles on Artificial Intelligence (AI)⁴ for additional information. It is understood that governance models vary in terms of components and terms used to describe these risk areas. Where there may be concerns about overlap, the intention is for this additional information to clarify the unique intent of each.

⁴ https://content.naic.org/sites/default/files/inline-files/AI%20principles%20as%20Adopted%20by%20the%20TF_0807.pdf

12/2/2022

**Big Data and AI (H) Working Group
Workstream #2
Model and Data Regulatory Questions**

Machine Learning (ML): Machine Learning is a subset of Artificial Intelligence. It covers the discipline concerned with the implementation of computer software that can learn autonomously⁵.

Model: A simplified description of a real-world system using Machine Learning or Artificial Intelligence.

⁵ <https://www.britannica.com/technology/machine-learning>