

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>