

AI Risk Evaluation Supplement version 5.0 – Comments Received

Table of Contents

01 ACLI Comment Letter AIRES v5.0 - 9.29.26.pdf
02a AHIP Letter responding to AI Risk Evaluation Supplement V 5.0 Exposure Final.pdf
02b AHIP ai-risk-evaluation-supplement-v5.0-redline Final.pdf
03 APCIA Comments_AI Risk Evaluation Supplement version 5.0.pdf
04 BCBSA Recommendations_NAIC AI Risk Eval Supplement V5_FINAL.pdf
05 ByteStream.docx
06 CAI Comment Letter on AIRES (9.29.26)(55721210.1)-c.pdf
07a CFAIO_Comment_NAIC_AI_Risk_Evaluation_Supplement.pdf
07b CFAIO_Proposed_Additions_Exhibit_B.pdf
08 NAIC Consumer Reps Comments on AI Evaluation Supplement 5.0.docx.pdf
09 CT Comments.docx
10 Dorvelon_NAIC_Comment_AI_Risk_Evaluation_Supplement_v5.pdf
11 EMILIA_Protocol_AI_Risk_Supplement_v5_Comments.docx
12 Ali_Houshang_i_NAIC_AI_Risk_Evaluation_Supplement_v5_Comments_2026-09-08.pdf
13 IRI letter to NAIC re NAIC AI Risk Evaluation Supplement v5_FINAL (9-29-2026).pdf
14 Kaenan-Hertz-Comments-AI-Risk-Evaluation-Supplement-v5.0.pdf
15 Kennedy - NAIC_AI_Risk_Evaluation_Supplement_v5_Expanded_State_Law_Asymmetry.docx
16 Legion_Modular_IP_NAIC_SUBMISSION_PACKAGE_READY_2026-09-10.pdf
17 Yash_Lundia_NAIC_V5_Comments.pdf
18 Monitaur-NAIC Artificial Intelligence (AI) Risk Evaluation Supplement V5 Comments.pdf
19a 2026-0929_NAMIC Comments_NAIC BDAIWG ARES v.5.pdf
19b 2026-0929_NAMIC Redlines_NAIC BDAI WG ARES v.5.pdf
20 Pinnacle_Comments on the Application of the NAIC AI Risk Evaluation Supplement to GLMs Used for Pricing
21 Point32health comments.docx
22 Project OV — NAIC AI Risk Evaluation Supplement v5.0 Comments — READABLE.pdf
23 ReconAI_NAIC_AI_Risk_Evaluation_Supplement_v5_Comment.pdf
24 Sharer_Consulting_Comment_Letter_NAIC_AI_Risk_Evaluation_Supplement_v5.0_2026-09-14.pdf
25 Verdict-Systems-NAIC-AI-Risk-Evaluation-Supplement-v5-Comments.pdf
26 Brian Wroblewski.docx
27 NALC Comment Ltr Re Evaluation Supplement v5.0.092926.pdf
28 Demond Davis - Nisaba.docx
29 EasyEssence_NAIC_Comment_Letter_FILING_2026-09-08.pdf
30 Falsify-OU-comments-AI-Risk-Evaluation-Supplement-v5.0.pdf
31 Joseph Stephens - Spectralgraph LLC.docx
32 naic-ai-risk-evaluation-supplement-comment-letter-Crittenden-2026-09-25.pdf
33 Nicholas Templeman - CSOAI Ltd (Council of AI).docx



September 29, 2026

Commissioner Nathan Houdek, Chair
Commissioner Doug Ommen and Mary Block, Vice Chairs
NAIC Big Data and AI (H) Working Group
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1000
Kansas City, MO 64106-2197

Submitted Electronically

Re: ACLI Comments on AI Risk Evaluation Supplement Version 5.0

Dear Chair Houdek and Vice Chairs Ommen and Block:

The American Council of Life Insurers (ACLI) appreciates the continued work of the NAIC Big Data and AI (H) Working Group in developing the AI Risk Evaluation Supplement (Supplement). We commend the Working Group for its thoughtful consideration of stakeholder feedback and for the meaningful refinements reflected in Version 5.0. ACLI supports the Working Group's objective of providing regulators with a practical, risk-based resource that supplements existing market conduct, financial analysis, and financial examination processes rather than creating a separate and independent examination or regulatory regime.

ACLI is encouraged and grateful that Version 5.0 incorporates several recommendations from our July 21, 2026, comment letter. Among several updates, Version 5.0 adds a definition of AI Model, removes unused terminology, narrows portions of Exhibit A, recognizes the potential use of previously submitted information, and reinforces the importance of tailoring inquiries to risk and materiality. These changes improve the Supplement and demonstrate the value of the Working Group's collaborative process.

At the same time, ACLI member review of Version 5.0 identified several areas where further clarification or refinement would improve consistency, reduce unnecessary burden for both insurers and regulators, and help insurers provide responsive information. This includes further refinement to ensure that the terms AI System and AI Model are clearly distinguished and operate coherently throughout the Supplement. We offer the comments below in a constructive spirit and with recognition that the Working Group must balance regulatory objectives and practical implementation considerations.

General Comments

ACLI appreciates the revisions confirming that the Supplement supports existing market conduct, financial analysis, and financial examination processes, and that the governing handbooks remain the primary guidance. To reinforce that approach, ACLI respectfully requests additional direction distinguishing which portions of the Supplement are relevant to market conduct, financial analysis, and financial examination functions. Clearer direction would promote more consistent, focused, and

American Council of Life Insurers | 300 New Jersey Avenue, NW, 10th Floor | Washington, DC 20001

The American Council of Life Insurers (ACLI) is the leading trade association driving public policy and advocacy on behalf of the life insurance industry. 90 million American families rely on the life insurance industry for financial protection and retirement security. ACLI's member companies are dedicated to protecting consumers' financial wellbeing through life insurance, annuities, retirement plans, long-term care insurance, disability income insurance, reinsurance, and dental, vision and other supplemental benefits. ACLI's 275 member companies represent 94 percent of industry assets in the United States.

responsive use of the Supplement. ACLI respectfully recommends that the Working Group implement this guidance through an applicability matrix or comparable crosswalk identifying which questions are relevant to market conduct examinations, financial analysis, and financial examinations. We believe that such a resource would reduce duplicative requests and help regulators and insurers apply the Supplement consistently within the purpose and scope of each review.

ACLI further requests that the Supplement state expressly that it is an examination and analysis aid and does not independently create substantive legal duties, standards of conduct, admissions, or expanded regulatory authority. This clarification should apply to the entire Supplement, not solely to questions concerning an insurer's AIS Program. Questions addressing matters such as independent validation, explainability, governance structure, staffing, consumer notice, risk quantification, or vendor oversight should not be interpreted as establishing requirements that do not otherwise arise under applicable law or adopted regulatory guidance.

Sequential and Risk-Based Use of Exhibits

ACLI appreciates the framework under which a regulator may begin with Exhibit A and use the insurer response to determine whether further inquiry is warranted. We request that the Supplement state more expressly that Exhibits A, B, C, and D are intended to operate sequentially and as progressively escalating inquiries. Exhibit A should serve as the threshold screening tool. A regulator should proceed with Exhibit B only when further governance inquiry is warranted and Exhibits C and D should be used only when prior responses identify a specific material risk requiring model-level review.

When deciding whether to proceed to a more detailed exhibit, regulators should consider inherent risk, mitigating controls and residual risk. Relevant controls may include human review, deployment limits, testing and monitoring, exception handling, reversibility, and remediation. Considering those controls before escalating the inquiry would better align the inquiry with the remaining risk. ACLI further recommends that before proceeding to Exhibit C or D, the requesting regulator should identify the examination objective, the specific area of heightened risk, and the basis for concluding that information provided through the preceding exhibits is insufficient. This explanation would help ensure that escalation is connected to an identified regulatory need, remains proportionate to the risk under review, and would further assist in the insurer providing the appropriate and specific information to the regulator thus providing for greater overall compliance.

ACLI believes this overall clarification is important because insurers cannot be expected to assume sequential use absent express direction. In many examinations, Exhibit A, or Exhibits A and B, may provide sufficient information. If a regulator proceeds to Exhibit C or D, the request should be limited to identified high-risk AI Systems or AI Models with material consumer or financial impact, rather than applied broadly to an insurer's full inventory. Explicit sequencing and added clarity that the Supplement was drafted in this manner and intent would preserve regulatory flexibility while improving predictability and proportionality.

Materiality and Interstate Consistency

ACLI appreciates Version 5.0's greater emphasis on risk and materiality, including its recognition that a regulator may specify a materiality threshold or rely on a company's materiality determination. We respectfully request additional guidance to promote consistency in how those determinations are made and applied. If individual jurisdictions select materially different thresholds, reporting periods, use cases, scopes, or documentation expectations, insurers may face duplicative requests and may be required to prepare different, and at times conflicting, responses concerning the same systems. Such variation

could also create the appearance of inconsistency where the differences instead reflect jurisdiction-specific instructions rather than a coordinated and consistent approach.

Accordingly, ACLI encourages the Working Group to establish common parameters for materiality and customization, while retaining flexibility for examination-specific circumstances. The Supplement should also apply materiality consistently throughout Exhibit A. As drafted, requests for system counts and the model inventory in Exhibit A Part Two may capture low-risk or operational systems with limited relevance to consumer protection or financial soundness. Tying these requests to direct consumer impact or material financial impact would better align Exhibit A with the Supplement's risk-based purpose. The Supplement should also establish a presumption in favor of an insurer's documented materiality framework where that framework is reasonable, consistently applied, tied to consumer or financial impact, and grounded in existing examination guidance and authority. A regulator could request a different approach when warranted by identified examination-specific circumstances but should explain the basis for doing so.

Previously Submitted Information

ACLI appreciates the recognition that regulators may rely on information previously submitted to another state when that information remains current and applicable. We encourage the Working Group to reinforce this principle by establishing a reasonable look-back period during which an insurer may direct a regulator to a prior submission and by clarifying that information already provided in a rate filing or other regulatory review need not be reproduced absent a specific need for updated or supplemental information. Greater interstate coordination and broader acceptance of common submissions would reduce duplication for both regulators and insurers without limiting a regulator's ability to request information necessary for a particular examination. The Supplement should also expressly encourage reliance on reviews already conducted by a lead state, group supervisor, or coordinating jurisdiction. Follow-up requests by another jurisdiction should focus on material changes, information not previously reviewed, or issues genuinely specific to that jurisdiction or examination objective.

Coordinated and Phased Implementation

ACLI respectfully recommends a timely, coordinated, and phased implementation of the Supplement. Insurers will need reasonable time to assess the final framework and align internal systems, inventories, processes, and personnel with potential information requests. Coordination among jurisdictions during implementation will be especially important to avoid overlapping requests under differing expectations. A measured and coordinated rollout would support more complete responses, allow regulators and insurers an opportunity to identify practical issues not addressed during the pilot, and advance the goal of a consistent, risk-based examination resource.

Confidentiality and Legal Protections

ACLI appreciates the Supplement's recognition of applicable confidentiality protections. Because responses may include proprietary governance materials, sensitive model information, cybersecurity-related information, trade secrets, or material subject to privilege or protection, it would be beneficial for the requesting regulator to identify the applicable legal authority and confidentiality protections before production. Submission of information under the Supplement should not be construed as waiving any otherwise applicable privilege or protection.

Additionally, insurers should be permitted to provide appropriately framed narrative descriptions where producing underlying materials would disclose otherwise privileged or protected information. Information should be shared among jurisdictions only where equivalent confidentiality and legal protections apply.

Definitions

ACLI appreciates the Working Group's addition of an AI Model definition and its effort to distinguish AI Systems from AI Models. We also appreciate removal of the unused definition of Degree of Potential Harm to Consumers. These are helpful changes. Our review, however, indicates that several definitions remain difficult to apply together and could produce inconsistent results. ACLI therefore respectfully requests a targeted review of the definitions of AI, AI System, AI Model, Algorithm, Predictive Model, Automation, Consumer Impact, Direct Consumer Impact, Adverse Consumer Impact, and Validation Method so that each term describes the correct concept and the terms operate coherently as a whole across the Supplement.

AI Systems, AI Models, and Traditional Predictive Models

The current definitions may blur the distinction between the development process used to create a model and the resulting model that is deployed. The development of a GLM, GAM, or another predictive model may involve iterative calculations using data, but the resulting model is a fixed scoring formula or set of rules that does not learn, self-update, or change during use. A later model may be developed using more recent data, but that is a separate, intentional development process rather than autonomous learning by the deployed model.

ACLI therefore reiterates its request that traditional rule-based systems, deterministic algorithms, fixed mathematical formulas, and transparent predictive models such as GLMs and GAMs be excluded from the Supplement. These tools present materially different governance considerations from systems that learn, adapt, or operate with meaningful autonomy. In addition, insurance pricing models are already subject to state regulatory review through rate filing processes, and requiring their inclusion in the Supplement may duplicate information already available to regulators. At a minimum, the definitions should make clear that a fixed formula is not transformed into an AI Model merely because iterative techniques were used during its development.

Consumer Impact, Direct Consumer Impact, and Adverse Consumer Impact

ACLI appreciates the Working Group's effort to focus the Supplement on matters affecting consumers. Further refinement is needed, however, because "Consumer Impact" is currently framed broadly around a decision subject to insurance regulatory standards, while "Direct Consumer Impact" may encompass nearly any matter involving a customer. ACLI respectfully requests that these terms be narrowed and harmonized so they focus on decisions that directly determine or materially affect a consumer's eligibility, coverage, premium, rating, claim, benefit, or other substantive insurance outcome.

The Supplement should also define "material consumer impact" particularly because Exhibit C uses that concept to identify models that may warrant detailed review. Alternatively, the Supplement should remove the term "material consumer impact" if "Adverse Consumer Outcome" is intended to serve as the operative standard. Without a clear and consistently applied materiality component, Exhibit C could extend beyond the high-risk matters for which it appears intended. ACLI also requests that any reference to adverse consumer effects be tied to a defined and meaningful insurance outcome rather than to any decision subject to insurance regulation.

Material Financial Impact and Insurance Regulatory Standards

ACLI appreciates the additional structure around material financial impact. We respectfully request further clarification distinguishing material financial impact from ordinary operational consequences and explaining its relationship to inherent risk and traditional financial statement materiality. ACLI also

reiterates its request for greater specificity regarding “Insurance Regulatory Standards”. A clearer description of the applicable laws, regulations, and standards would reduce uncertainty and promote consistency across jurisdictions. We also ask the Working Group to review terminology that uses “adverse” in connection with insurance regulatory standards to ensure that it accurately describes the intended concept.

Exhibits A-D

Exhibit A

ACLI appreciates the narrowing of the “Other Insurance Practices” category to matters involving potential direct consumer impact or material financial impact, as well as the removal of Legal/Compliance and Producer Services as separately identified rows. These revisions appropriately focus Exhibit A on areas of greater regulatory relevance over the business of insurance.

We do, however, request other refinements to further improve upon Exhibit A. First, ACLI requests a single, uniform definition of “currently in use” and a consistent reporting period. We believe that the preceding twelve months provides a reasonable and administrable timeframe. Allowing different regulators to select different periods would increase burden and reduce comparability.

Second, as noted in our definitions comments, Exhibit A should more clearly distinguish AI Systems from AI Models to avoid confusion and double counting. ACLI recommends revisiting the columns requesting counts of AI Systems and focusing the exhibit on items that present direct consumer impact or material financial impact. We further recommend that the residual category for other AI use should be removed or expressly limited by materiality.

Third, ACLI requests that Exhibit A Part Two be reassessed. As an introductory inquiry, Exhibit A should provide only the high-level information needed to determine whether further review is warranted. A broad model inventory at this stage may expand the scope and burden and capture systems immaterial to the examination. If Part Two is retained, it should be limited to models within a specified business area that present direct consumer impact or material financial impact, and should focus on model name, use case, relevant operational area, inherent-risk classification, and potential consumer or financial impact rather than raw model counts separate from risk or materiality.

Fourth, ACLI recommends defining the term “implemented” to mean that an AI System or AI Model has entered full production use. Systems or models in development, testing, proof-of-concept, or limited pilot stages should not be counted as implemented as they may never be deployed operationally. The Supplement should also clarify whether the requested implementation count covers only an initial production deployment or also includes subsequent material modifications.

Finally, ACLI reiterates its request that the “Fraud/Waste & Abuse” category be clarified so that waste and abuse are used only in contexts where those terms have an established regulatory meaning, such as health-program integrity. ACLI recommends using “Fraud Detection” as the specific category which we believe is consistent with the underlying intent.

Exhibit B

ACLI respectfully requests clarification regarding the relationship between an insurer’s AI governance program and its Enterprise Risk Management (ERM) framework. For many insurers, an AI governance program may be aligned with and collaborate with ERM without being housed in the ERM organizational

structure. Exhibit B should recognize both arrangements and focus on whether governance, escalation, oversight, and risk-management functions are appropriately coordinated rather than where the program is organizationally located.

ACLI also recommends reconsidering the question concerning staffing reductions associated with AI implementation. The regulatory purpose of this inquiry is unclear in the context of market conduct or financial examinations, and a staffing change does not necessarily indicate reduced human oversight of AI systems. Any retained question in this area should be tied to the adequacy of oversight or controls relevant to the examination.

In addition, the narrative and checklist versions of Exhibit B appear to address many of the same governance, risk management, oversight, compliance, and vendor-management subjects. ACLI encourages the Working Group to consolidate, streamline, or allow an insurer to respond either through a narrative or checklist depending on which format most accurately reflects the insurer's governance structure and existing documentation. This would avoid duplication while preserving the information regulators need. This will benefit insurers compliance with the Exhibit and regulators in their review and understanding of the elements submitted. As such, the Exhibit B instructions should expressly memorialize that the insurer may select either the narrative or checklist format and need not complete both.

ACLI also recommends removing the consumer-notification question in Exhibit B, Checklist Item 3n, or expressly limiting it to disclosures required by applicable law or adopted regulatory guidance. The Supplement should not imply a general consumer-notice obligation where no such requirement otherwise applies.

Finally, for third-party systems and models, an insurer should be required to provide only information reasonably available through its contractual rights and ordinary vendor-oversight processes. The Supplement should not imply that an insurer must produce source code, training data, model weights, proprietary validation materials, or other information that it has no legal or contractual right to obtain.

Exhibit C

ACLI requests that Exhibit C be expressly limited to identified high-risk models with material consumer impact or material financial impact. Questions concerning effects on financial statements, risk assessment, or controls should not assume that every high-risk model affects each of those functions. Rather, insurers should respond only to the portions applicable to the specific model and examination context. As noted above, "material consumer impact" should be defined before it is used as a threshold for Exhibit C.

As a technical comment, Exhibit C is limited to models already identified as high risk. The separate request for a model-risk classification appears unnecessary and potentially confusing. ACLI recommends removing that field.

Exhibit D

ACLI continues to have significant concerns regarding the potential breadth and operational burden of Exhibit D. ACLI appreciates the revisions requiring identification of the AI Systems or AI Models associated with reported data and recognizing that requests may be tailored by line of business. Those changes provide helpful context, but further limitations are warranted.

If retained, Exhibit D should apply only after sequential review of Exhibits A through C identifies a specific high-risk AI System or AI Model with material consumer or financial impact. The request should be limited to the models relevant to the examination, exclude models already provided through a rate filing or other regulatory process, and not operate as an enterprise-wide data inventory. Exhibit D should also distinguish between data elements directly used by a model and data elements used to infer a characteristic, and the regulator should specify which category is relevant to the particular inquiry.

Because many pilot participants tested only Exhibit A, ACLI requests that Exhibits C and D not be used broadly until their operational burden, clarity, and consistency have been adequately tested through additional piloting or a phased implementation. These limitations would preserve the potential regulatory value of Exhibit D while making its application more targeted and manageable for both insurers and regulators alike.

Conclusion

ACLI appreciates the substantial progress reflected in Version 5.0 and the Working Group's continued willingness to consider stakeholder feedback. The revisions already made meaningfully improve the Supplement. As the Working Group moves toward final adoption, ACLI respectfully encourages particular attention to:

- Excluding traditional fixed and transparent predictive models;
- Clarifying and harmonizing core definitions;
- Expressly requiring sequential and risk-based use of the exhibits;
- Applying consistent materiality, jurisdictional, and reliance parameters; and
- Providing for a coordinated, phased implementation.

ACLI and its members remain committed to working collaboratively with regulators to identify practical compromises that support effective oversight, consistent application, and a manageable examination process. We greatly appreciate the opportunity to provide these comments and look forward to continued engagement as the Supplement advances.

Respectfully submitted,



Jared Kosky
Assistant Vice President & Counsel
American Council of Life Insurers
jaredkosky@acli.com

cc: Miguel Romero, NAIC
Scott Sobel, NAIC
Dorothy Andrews, NAIC

October 2, 2026

Commissioner Nathan Houdek, Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106-2197

By Email to: Miguel Romero at Miguel Romero at maromero@naic.org

Re: Version 5.0 of Artificial Intelligence Risk Evaluation Supplement

Dear Commissioner Houdek:

On behalf of AHIP¹, thank you for the opportunity to provide feedback on Version 5.0 of the NAIC Big Data and Artificial Intelligence (H) Working Group's Artificial Intelligence Risk Evaluation Supplement (Supplement). AHIP appreciates the substantial work reflected in the Supplement, including the opportunity for plans to participate in the Pilot program (Pilot) that informed Version 5.0. We appreciate that Pilot states have engaged participating plans and considered their feedback in Version 5.0. Additional visibility into how that feedback informed and was incorporated into Version 5.0 would be helpful. AHIP's letter focuses on the Supplement and its future use and is accompanied by a redline with additional suggested edits.

AHIP supports the Version 5.0 changes that clarify the Supplement's intent and scope, including alignment with "existing market conduct, financial analysis, and/or financial examination review procedures." We appreciate the effort to confirm that the Supplement should be used within existing financial and market regulatory frameworks and recommend a few additional clarifications:

- **Risk-Focused and Customization Guidance.** The NAIC's AI Model Bulletin (Bulletin) emphasizes governance, accountability, proportional controls, and flexibility based on risk, while the Supplement requires detailed inventories, documentation, and evidence regardless of risk level. This creates operational misalignment with the Bulletin. Each exhibit instructs regulators to customize the Supplement, but does not explain how that process should work in practice. AHIP recommends adding the specific examples to guide customization, align the Supplement's scope with the purpose of market conduct or financial examinations, and avoid overly expansive information requests. Certain low-risk

AI uses should also be left outside the scope, including internal company operations and low-risk customer service tools, such as chatbots that summarize or provide already approved information.

- **Consistent Application of Materiality Threshold.** AHIP appreciates that Version 5.0 begins to address materiality concerns raised by interested parties. However, the current definition could allow varying interpretations by states or examiners. This is especially important in the market conduct context because the threshold is tied to the definition of materiality in the Financial Condition Examiner’s Handbook, which may not align with questions of consumer impact. AHIP recommends additional guidance to support consistent application of the materiality threshold across states.
- **Consistency and Standardization Across States.** Consistent regulatory expectations across states are important for effective AI oversight. AHIP recommends a prescribed AI review approach incorporated into the Financial Condition and Market Conduct Examination Handbooks to support clear, consistent application by regulators and regulated entities. The Supplement should also align with the Model Bulletin’s definition of “AI Systems” rather than separately referencing “AI Systems” and “AI Models,” which could create unnecessary variation in interpretation and implementation.

Thank you for the opportunity to provide these comments. AHIP appreciates the continued efforts to clarify the Supplement’s intent and scope and looks forward to further engagement as this work advances. Please contact us with any questions.

Sincerely,
LaCosta Wix
AHIP Senior Regulatory Counsel, State Affairs and Policy
lwix@ahip.org
504-579-628

Cc: Miranda Motter, Senior Vice President, State Government Affairs and Policy

¹AHIP is the national association whose members provide health care coverage, services, and solutions to hundreds of millions of Americans every day. We are committed to market-based solutions and public-private partnerships that make health care better and coverage more affordable and accessible for everyone. Visit www.ahip.org to learn how working together, we are Guiding Greater Health.

Artificial Intelligence (AI) Risk Evaluation Supplement

Background, Intent, and Instructions

Background:

The rapid expansion of big data and the adoption of artificial intelligence and machine learning are significantly transforming insurance practices. These technologies can offer substantial benefits to both insurance companies and consumers by facilitating the development of innovative products, improving customer interface and enhancing service, simplifying and automating processes, and promoting efficiency and accuracy. However, without robust governance and effective controls, the use of *AI Systems* may lead to Adverse Consumer Outcomes or compromise the financial soundness of an insurance company. Insurers are responsible for managing the risks associated with the development and implementation of *AI Systems* and must demonstrate to regulators that appropriate risk-based oversight mechanisms are in place and are functioning effectively. Ultimately, insurers that make decisions or perform actions impacting consumers or financial soundness that are made or supported by advanced analytical and computational technologies, including from *AI Systems*, must comply with all applicable insurance laws and regulations.

Intent:

The NAIC's Innovation, Cybersecurity and Technology (H) Committee charged the Big Data and AI Working Group (BDAIWG) to create tool(s) that would assist regulators to identify and assess both financial and consumer risks from the use of *AI Systems* on an on-going basis.

The following exhibits are designed to supplement existing market conduct, financial analysis, and/or financial examination review procedures for reviewing *AI Systems*. As this supplements existing NAIC resources, regulators should continue to consider existing NAIC resources as authoritative but may consider drawing from this supplement to assist in understanding and assessing a company's use of *AI Systems*. Inquiries and information requests performed related to this supplement will be coordinated consistent with the guidance provided by the *Market Regulation Handbook*, *Financial Condition Examiners Handbook*, and the *Financial Analysis Handbook*.

With the intention of the document as a supplement acknowledged, there may be some important context that is not specified in this document, particularly related to deciding which companies are in scope of inquiry. This is to avoid the perception of conflicting guidance as many important decisions leading to the full use of this Supplement would be governed by the aforementioned Handbooks.

These optional exhibits allow regulators to determine the extent of *AI Systems* usage for an insurance company and whether additional analysis is needed focusing on financial and consumer risk.

This Supplement consists of the following exhibit questionnaires:

- **Exhibit A: Quantify Regulated Entity's Use of AI Systems**

Formatted

Commented [A1]: We appreciate the add of the intent section and this specifically

- **Exhibit B: AIS Program (Two Options: Narrative or Checklist)**
- **Exhibit C: High-Risk AI Models**
- **Exhibit D: AI Model Data**

Instructions:

Information obtained from the responses to the following Exhibits may supplement guidance and tools used during an existing market conduct, certification, financial analysis, and/or financial examination review, to enhance the regulator's understanding of the extent of *AI Systems* utilization and assessment of potential risk to consumers and the insurance company. Effective assessment requires regulators to maintain a fluent understanding of the applicable laws and regulations, including but not limited to those pertaining to unfair trade practices, unfair claims settlement practices, corporate governance annual disclosures, confidentiality, financial reporting, and rating.

Regulators using this Supplement may wish to first use Exhibit A to determine if further inquiry using Exhibits B-D is necessary. It is possible that while the company responding indicates usage of *AI Systems*, its use of AI is limited or low in inherent risk such that further inquiry as contemplated by subsequent exhibits is not warranted. For instance, if a regulator is using the Supplement to inquire about a company's claims practices and based on responses to Exhibit A, the regulator determines that the company's use of AI is out of scope of a claims-related targeted exam, the regulator may determine it is appropriate to thereafter rely on the Market Regulation Handbook to guide the remaining work.

If information requested through this Supplement has already been provided to this department or any other state department of insurance, the company's response should so state, in which case the regulators may accept prior submissions if the prior response is still current and applicable.

The responses to these exhibits will be considered by regulators when identifying the inherent risks of the insurer. They should also affect the planned examination or inquiry approach, as well as the nature, timing and extent of any further procedures performed.

General Guidance

Machine Learning (ML) Models

Machine learning models are widely used to estimate the relationship between a dependent variable and one or more independent variables. For example, an insurer may want to estimate how a driver's age relates to the probability of being involved in a car accident. There are many ML models that can be used for this purpose.

In particular, generalized linear models (GLMs) are one such technique. As with other machine learning techniques, GLMs are not without risk of causing unfair discrimination or other adverse consumer outcomes and require effective governance over data quality, model development, validation, implementation, ongoing performance monitoring, and compliance with applicable laws and regulations.

This supplement asks about a company's use of specific models in Exhibit A to give regulators flexibility in determining their scope of interest.

Materiality

This Supplement was designed to allow regulators the flexibility to either specify materiality or rely on the company's *materiality* calculations as companies respond to Supplement related inquiries.

The concept of *materiality* leverages the definition otherwise used to drive financial exam-related decisions and as defined in the FCEH. For purposes of completing responses related to this Supplement, regulators may opt to rely on a company's internal materiality thresholds provided the information supporting a company's *materiality*, including an explanation for why that threshold was considered appropriate.

However, nothing in the guidance provided within this Supplement should be construed to supersede FCEH guidance.

Risk Assessment

This Supplement also, at times, relies on a company's Risk Assessment. The concept of Risk Assessment centers around the discussion of *Inherent Risk* which leverages the definition reflected in the FCEH. For purposes of completing responses to this Supplement, regulators that Risk Assessment information provided to regulators include explicit indication of *Inherent Risk* Assessments (i.e., risk before the consideration of controls or mitigation measures).

Consistent with the discussion of *materiality*, nothing in the guidance provided within this Supplement should be construed to supersede FCEH guidance.

Confidentiality

The information requested via these Exhibits will be held confidential, consistent with the requesting state's relevant authority.

Regulators using any of these exhibits should cite examination or other authority as appropriate when requesting information from insurers. Regulators should cite all relevant confidentiality statutes or other specific protections related to documents, materials or other information in the possession or control of regulators that are obtained by or disclosed to the regulators or any other person in the course of a market conduct inquiry and all information reported or provided to the regulator pursuant to cited examination or other authority.

Commented [A2]: We are pleased to see that this version attempt to address the materiality issue

Exhibit A: Part One - Quantify Regulated Entity's Use of AI Systems

Purpose: To obtain information pertaining to the number of *AI Systems* and *AI Models* used by a company to help facilitate risk assessment. Based on the responses from the company, regulators may ask for additional information related to their *AI Program* (Exhibits B), high-risk models (Exhibit C), and the types of data (Exhibit D) where there is risk for *Adverse Consumer Outcomes* or material adverse financial impact.

Company Instructions:

- Provide the most current counts and use cases of the following, as requested.
- For purposes of responding to information requests related to this Exhibit, models that *augment* or *automate* decision-making related to consumers are considered to have direct consumer impact.
- “*AI System*”, “*Adverse Consumer Outcome*”, and “*Use Case*” are defined in the “Definitions and Appendix” section of this Supplement.
- Include **all lines of business**. If the governance differs by company, line of business, or state, work with your regulator to determine if multiple submissions are needed.
- If a company has an existing *AI Systems* inventory, it may suggest submission of the inventory in lieu of completing Exhibit A.
- If a company has questions about any of the fields, please submit them in writing to **<regulator contact email address>**.

Regulator Instructions:

- **Regulators should customize this Exhibit to limit information requested to more targeted inquiries for use in a limited scope exam.**
- **Regulators should either specify a materiality threshold to guide company responses to the Supplement or request that the company provide a threshold that it used in completing this Exhibit.**
- For column 3, regulators should specify the number of months.

Company Legal Name and Group Name: _____

NAIC CoCode and Group Code: _____

Company Contact Name: _____ Email: _____

Line(s) of Business for Which This Response Applies: _____

Materiality Threshold <regulator to provide or specify if the company will provide>: _____

Commented [A3]: This should be more specific to the scope of the review and not “all lines of business”

Commented [A4]: It's not quite clear how this would happen in practice. It would be helpful to maybe have some examples, otherwise the concern is that the supplement is not really customized and will be too broad in scope to truly align with the relevant market conduct or financial exam.

Commented [A5]: The discussion above about materiality seems to suggest that regulators look to the definition/guidance in the FCEH. The definition and guidance around materiality in the FCEH is in the context of financials. This seems to leave a gap in guidance for regulators to specify a materiality threshold here, that can translate to the AI questions in Exhibit A, particularly around consumer impact that is not financial in nature. More clarification is needed to help guide the regulator if they plan to specify a materiality threshold, which would in turn facilitate how the user/insurer response

Date Form Completed: _____

DRAFT

(1) Use of AI System in Operations or Program Area	(2) Number of AI Systems Currently in Use	(3) Number of AI Systems Implemented in the Past X Months	Number of AI Models with Direct Consumer Impact			Number of AI Models with Material Financial Impact			(10) Use Case(s)
			(4) Generalized Linear Models (GLMs)	(5) Generative/Agentic AI Models	(6) Other AI/ML Models	(7) Generalized Linear Models (GLMs)	(8) Generative/Agentic AI Models	(9) Other AI/ML Models	
Marketing									E.g., UC1: Identify potential consumers interested in product.
Premium Quotes & Discounts									
Underwriting/Eligibility									
Ratemaking/Rate Classification/Schedule Rating/Premium Audits									
Claims/Adjudication*									
Customer Service									
Utilization Management/Utilization Review/Prior									

Commented [A6]: Recommend clarifying to “material.” The term is a well-established concept in existing frameworks and supports consistent, risk-based application and appropriately focuses attention on AI systems that meaningfully affect consumer outcomes. “Direct” is not well-defined and not included in established frameworks, and could broadly sweep in low-risk tools that do not materially affect consumers.

(1) Use of AI System in Operations or Program Area	(2) Number of AI Systems Currently in Use	(3) Number of AI Systems Implemented in the Past X Months	Number of AI Models with Direct Consumer Impact			Number of AI Models with Material Financial Impact			(10) Use Case(s)
			(4) Generalized Linear Models (GLMs)	(5) Generative/ Agentic AI Models	(6) Other AI/ML Models	(7) Generalized Linear Models (GLMs)	(8) Generative/ Agentic AI Models	(9) Other AI/ML Models	
Authorization/ Level of Care Determination									
Fraud/Waste & Abuse									
Investment/ Capital Management									
Reserves/ Valuations									
Catastrophe Triage									
Reinsurance									
Other Insurance Practices** (if applicable)									
* Includes Salvage/Subrogation, adjuster assignments, etc.									
** Any additional operation or program area with the potential to result in either Direct Consumer Impact or Material Financial Impact.									

Commented [A6]: Recommend clarifying to “material.” The term is a well-established concept in existing frameworks and supports consistent, risk-based application and appropriately focuses attention on AI systems that meaningfully affect consumer outcomes. “Direct” is not well-defined and not included in established frameworks, and could broadly sweep in low-risk tools that do not materially affect consumers.

Exhibit A: Part Two – AI Systems Request List

In Support of the regulatory inquiries related to this Supplement, the regulators request information or documentation related to the following:

Regulator Instructions: Regulators should customize this Exhibit to limit information requested to more targeted inquiries for use in a limited scope exam.

1. Materiality – Explanation related to the company's *materiality* calculation (if applicable) used in responding to inquiries related to this Supplement.
2. Risk Assessment – Documentation explaining the company's risk assessment process for purposes of its use of *AI Systems*.
3. Model Inventory – A listing of the company's models, that specifies risk for the following program areas <regulator to specify program areas in scope of the request>. Additionally, regulators request the model inventory or accompanying documentation include:
 - a. *AI Model* Name
 - b. Description of how the *AI Model* is used
 - c. Description of the broader *Use Case* in which each *AI Model* is used
 - d. Operation or Program Area(s) the *AI Model* is Used Within
 - e. *Inherent Risk* Level Based Upon the Company's Risk Assessment
 - f. What consumer impact does this *AI Model* have?
 - g. What financial impact does this *AI Model* have?

Commented [A7]: Similarly, the guidance to regulator is to use the definition in the FCEH, which is financials focused. It's not clear what risk assessment they will be looking for as it relates to AI models/systems that are "consumer impact" focused. There need to be more specification/guidance for regulators and the insurers

Exhibit B: (Narrative) AIS Program

Purpose: To obtain the Company's *AIS Program*, including how the company handles governance, risk identification, mitigation, and management framework and internal controls, and third-party *AI Systems* and data for *AI Systems*; and the process for acquiring, using, or relying on third-party *AI Systems* and data. Market and financial regulators should coordinate to gain access to the relevant section of the policies governing the use of *AI Systems*.

Company Instructions: Below are both Checklist and Narrative sections of Exhibit B of the NAIC's AI Risk Evaluation Supplement. If the responses to the checklist indicate that an explanation of a policy, procedure or process can be found on a given page of a document, indicate the page number and the document title and include the document referenced in the submission. If other sections of the document are irrelevant, the relevant sections of the document may be submitted in lieu of the entire document. Please see definitions at the end of this document.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam.

The references to, and questions about, elements of a company's *AIS Program* are not intended to be interpreted as creating new requirements for *AIS Program*.

Group and Company Legal Name: _____

NAIC Group and Company CoCodes: _____

Company Contact Name: _____ Email: _____

Date Form Completed: _____

1. Provide your *AIS Program*. [Click or tap here to enter text.](#)
 - a. What role maintains the *AIS Program*? [Click or tap here to enter text.](#)
 - b. Discuss the governance structure, Board reporting and frequency. [Click or tap here to enter text.](#)
 - c. Discuss the process by which the *AIS Program* is integrated throughout the organization, assessed and remediated. [Click or tap here to enter text.](#)
 - d. Discuss the process by which the effectiveness of the *AIS Program* and individual models are assessed and modified. [Click or tap here to enter text.](#)

Commented [A8]: We appreciate this clarification. We would request additional clarification relative what this looks like in practice. The more a regulator can clearly scope the insurer's response the better given the breath of the ask in this section.

- e. Discuss how responsibility for governance within the organization is assigned and how the organization ensures consistency and alignment. [Click or tap here to enter text.](#)
 - f. Discuss the integration of the *AI Systems* in the Own Risk and Solvency Assessment (ORSA) and Enterprise Risk Management (ERM) assessments. [Click or tap here to enter text.](#)
 - g. How does the insurance company assess autonomy, reversibility, and reporting impact risk of *AI Systems*? [Click or tap here to enter text.](#)
2. Discuss the uses of *AI Systems* that:
- a. Generate a *material financial impact*. [Click or tap here to enter text.](#)
 - b. Generate a *direct consumer impact*. The company's response should include, but not be limited to, whether there is a human in the loop, whether there are high volume transactions, how any exceptions or errors are found and resolved, and whether there were any staffing reductions as part of the *AI System* implementation. [Click or tap here to enter text.](#)
 - c. Generate or impact material information reported in financial statements. [Click or tap here to enter text.](#)
 - d. Generate or impact risk or control assessment. [Click or tap here to enter text.](#)
 - e. Discuss the development, testing, and implementation of material *AI Systems* that the Company has implemented. If appropriate, include details regarding where any systems differ from established IT systems and data handling protocols. Discuss the basis for deviation from established practices. [Click or tap here to enter text.](#)
3. Provide the policy for, and discuss the use and oversight of, material *AI System* vendors, model design and testing:
- a. Discuss the validation and testing procedures performed on internally-developed *AI Systems*, including whether validation and testing are performed by someone that is independent from development. [Click or tap here to enter text.](#)
 - b. Discuss the validation and testing procedures performed on third-party vendor-supplied *AI Systems*. [Click or tap here to enter text.](#)
 - c. Discuss the testing and verification that has occurred including frequency, scope and methodology. [Click or tap here to enter text.](#)
 - d. Discuss how the company ensures explainability and transparency into any *AI models* recommending or making decisions. [Click or tap here to enter text.](#)
4. Provide the policy for, and discuss the use and oversight of, material *AI Systems* by third-party providers including actuarial, claim, MGA, audit, and/or other professional services. [Click or tap here to enter text.](#)
- a. Discuss the testing and verification that has occurred, frequency, scope, and methodology. [Click or tap here to enter text.](#)
5. Discuss additional aspects of the framework design and evaluation pertaining to *AI Systems*. [Click or tap here to enter text.](#)

- a. Discuss the unit(s) responsible for the *AIS Program*, assessment approach and frequency, and involvement with the program area to the extent it differs from that discussed previously. [Click or tap here to enter text.](#)

DRAFT

Exhibit B: (Checklist) AIS Program

Purpose: To obtain information about the Company's *AIS Program* including information about how the company handles governance, risk management and internal controls, and third-party *AI Systems* and data for *AI Systems*; and the process for acquiring, using, or relying on third-party *AI Systems* and data. The references to, and questions about, elements of a company's *AI Systems* Governance Risk Assessments are not intended to be interpreted as creating new requirements for *AI Systems* Governance Risk Assessments.

Company Instructions: Provide responses to the questions regarding how the governance of *AI Systems* fits within your company's system of supervision or Enterprise Risk Management program. Include all companies and lines of business. If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See [definitions](#) below.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam.

Group and Company Legal Name(s): _____

NAIC Group and Company Code(s): _____

Company Contact Name: _____ Email: _____

Date Form Completed: _____

Ref	AI Systems Use Questions for Company	Company Response
1	Has the company adopted a written <i>AIS Program</i> ? If yes, when was it adopted and what is the frequency of review for updating?	
2	Was the Board of Directors or management involved in the adoption of an <i>AIS Program</i> ?	
	2a. What is the role of the Board of Directors or management in the <i>AIS Program</i> ?	

Commented [A9]: This should be more specific to the scope of the review and not "all lines of business"

Ref	AI Systems Use Questions for Company	Company Response	
3	Reference the processes and procedures of the Company <i>AIS Program</i> that addresses the following:		
	How the Insurance Company...	Document Name and Page #	If not specified in the <i>AIS Program</i> , provide details below:
	3a. Assesses, mitigates, and evaluates <i>inherent risks</i> related to the use of <i>AI System</i> including the possibility of the company engaging in unfair trade practices		
	3b. Ensures <i>AI Systems</i> are compliant with applicable state and federal laws and regulations		
	3c. Evaluates the risk of <i>Adverse Consumer Outcomes</i>		
	3d. Considers data privacy and protection of consumer data used in <i>AI Systems</i>		
	3e. Evaluates whether <i>AI Systems</i> are suitable for their intended use and should continue to be used as designed		
	3f. Considers <i>AI System</i> risks within its Enterprise Risk Management (ERM)		
	3g. Considers <i>AI System</i> risks within the Own Risk and Solvency Assessment Report (ORSA), as applicable.		
	3h. Considers <i>AI System</i> risks within the software development lifecycle (SDLC) or within the broader <i>AIS Program</i>		
	3i. Considers <i>AI System</i> risk impact on financial reporting		
	3j. Trains employees about <i>AI System</i> use and defines prohibited practices (if any)		
	3k. Quantifies <i>AI System</i> risk levels		
	3l. Provides standards and guidance for procuring and engaging <i>AI System</i> vendors		
	3m. Considers consumer complaints related to consumer outcomes in material resulting from <i>AI</i>		

Commented [A10]: Health insurance already operates under robust, technology-neutral complaint tracking and resolution requirements, and existing systems are designed to capture consumer issues and outcomes rather than attribute complaints to specific tools such as AI

	Systems and whether they are identified, tracked, and addressed <u>through existing complaint and remediation frameworks</u>		
	3n. Promotes consumer awareness of the use of <i>AI Systems</i> through disclosures, policies, and procedures for consumer notification, as appropriate		
	3o. Determined “ <i>materiality</i> ” and what threshold was used for assessing <i>AI Systems</i> risk.		
	3p. Governs, monitors, tests and ensures transparency of <i>AI Systems</i> developed by vendors.		

Exhibit C: High-Risk AI Models

Purpose: To obtain detailed information on in-production high-risk *AI models*, such as models making automated decisions that could cause *Adverse Consumer Outcomes*, *material financial impact*, *material consumer impact*, or material financial reporting impact. *AI model risk criteria* are set by the insurance company. Note that *materiality* should be a consideration in determining the level of risk. To assist in identifying models for which this information is requested, regulators may request information on the company's risk assessment and a model inventory if such information has not otherwise already been provided.

Company Instructions: Fill in the details for each of the *AI model(s)* requested. Include *all companies and lines of business*. If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See definitions below. The template below refers to both *AI Systems* and *AI Models* depending on the information being requested. There may be some instances where a company feels information should be provided in relation to the *AI System* and not the *AI Model* or vice-versa. This should be discussed with regulators as part of the submission process to avoid misunderstanding.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam.

Group and Company Legal Name(s): _____

NAIC Group and Company Code(s): _____

Company Contact Name: _____ Email: _____

Date Form Completed: _____

Ref	AI Model Information Requests	Company Response
1	AI model name and version number.	
2	Use cases and purpose(s) of model	
3	Model type used in the AI System (GLM, generative/agent AI, or other AI/ML algorithm)	
4	Model Implementation Date	
5	Model development (internal, "owned" third party, or true third party, or a mix of both – include vendor name if applicable)	

Commented [A11]: This statement needs additional clarification. The term "AI model risk criteria" is not defined. If it's intended to clarify that insurance companies must identify those "high-risk AI models" based on their AI model risk criteria in order to respond to this section, then that should be made clear.

Commented [A12]: This should be more specific to the scope of the review and not "all lines of business"

6	Did the Company purchase the <i>AI model</i> , was the <i>AI model</i> developed specifically for the Company? Is the Company able to modify the <i>AI model</i> as necessary?	
7	Model risk classification (high, moderate, low, etc.)	
8	Model risk(s) (i.e., describe the potential for <i>adverse consumer impact</i>)	
9	Model limitations (i.e., limitations on <i>uses cases</i> or data inputs outside the domain of the training data, or other guardrails to manage model related risk)	
10	<i>AI model</i> type (automate, augment, support)	
11	Discuss testing model outputs (e.g. model drift, accuracy, unfair trade practices, unfair discrimination, performance degradation, etc.) and how the model was validated prior to being deployed as well as how its performance is monitored on an ongoing basis.	
12	Last date of model testing	
13	Discuss how the model affects the financial statements, risk assessment or controls.	
14	Discuss how the model is reviewed for compliance with applicable state and federal laws, including but not limited to the unfair trade practices act and unfair claims settlement laws.	
15	To the extent permitted by law, discuss if the company has had any actions taken against them for use of this model. Actions may include, but are not limited to, informal agreements, voluntary compliance plans, administrative complaints, ongoing third-party monitoring, cease and desist, remediation, restitution, fines, penalties, investigations, consent orders or other regulatory agency actions.	

Exhibit D: AI Model Data

Purpose: To obtain detailed information of the source(s) and type(s) of data used in an *AI model(s)* to identify risk of *adverse consumer impact*, unfair trade practices, material financial impact, or material financial reporting impact.

Company Instructions: Provide details below for the data used in *AI model(s)*. If any of the data elements listed are used in the training or test data as part of the development of an *AI model*, provide information on whether the data element is sourced internally or whether the data element is sourced from a third party, in which case provide the name of the third-party vendor. Leave blank if a data source is not used in the development of *AI model* for the insurance operation. **Include all companies and lines of business.** If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See [definitions](#) below.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam. Regulators may also wish to customize this supplement for the line of business as some data types may or may not be pertinent based on the responding company's lines of business. Depending on responses received, a regulator may wish to ask for a data dictionary to gain a more complete understanding of a data used for a given model.

Group and Company Legal Name(s): _____

NAIC Group and Company CoCode(s): _____

Company Contact Name: _____ Email: _____

List the *AI System* or *AI models* which rely on the data described in response to this submission:

Describe the Line of Business for Which This Response Applies (complete one for each line of business):

Date Form Completed: _____

Commented [A13]: This should be more specific to the scope of the review and not "all lines of business"

	(1)	(2)	(3)	(4)
Ref	Type of Data Element Used in AI Model	Describe how the AI Model uses this data including how the output is used within operations; or input data is used for training and testing	Internal Data Source	Third Party Data Source / Vendor Name
1	Aerial Imagery			
2	Age, Gender, Ethnicity/Race			
3	Consumer or Other Type of Insurance/Risk Score			
4	Crime Statistics			
5	Criminal Convictions (Exclude Auto-Related Convictions)			
6	Driving Behavior			
7	Education Level (Including school aptitude scores, etc.)			
8	Facial or Body Detection / Recognition / Analysis			
9	Geocoding (including address, city, county, state, ZIP code, lat/long, MSA/CSA, etc.)			
10	Geo-Demographics (including ZIP/county-based demographic characteristics)			
11	Household Composition			
12	Image/video Analysis			
13	Income			
14	Job History			
15	Loss Experience			
16	Medical, including Biometrics, genetic information, pre-existing conditions, diagnostic data, etc.			
17	Natural Catastrophe Hazard (Fire, Wind, Hail, Earthquake, Severe Convective Storms)			

18	Online social media, including characteristics for targeted advertising			
19	Personal Financial Information			
20	Reasonable Accommodations or Policy Modifications (granted or requested)			
21	Telematics/Usage-based insurance			
22	Vehicle-Specific Data including VIN characteristics			
23	Voice Analysis			
24	Weather			
25	Other: Non-Traditional Data Elements (Regulator to specify)			

DEFINITIONS AND APPENDIX

Where available, for the purposes of this evaluation terms are defined in accordance with the NAIC Model Bulletin on the Use of *AI Systems* by Insurers (https://content.naic.org/sites/default/files/2023-12-4%252520Model%252520Bulletin_Adopted_0.pdf):

“Adverse Consumer Outcome” refers to an *AI System* decision (output) by an insurance company that is subject to insurance regulatory standards enforced by the Department that adversely impacts the consumer in a manner that violates those standards.

“Agentic AI” refers to *AI Systems* that can pursue objectives across multiple steps without requiring human input at each stage, which could involve perceiving the environment, deciding what to do next, taking actions (often through tools or APIs), observing/analyzing the results, and/or adjusting the action accordingly. The difference is in agency – the system does not just respond to a single prompt or instruction, but actively executes a process.

“Algorithm” means a clearly specified mathematical process for computation; a set of rules that, if followed, will give a prescribed result.

“AI Model” refers to an integral component of an *AI System* that learns from data to perform tasks via computational, statistical and/or machine learning techniques. It combines algorithms, training data and learned parameters to produce outputs from a given set of inputs, such as making predictions or decisions on data it has never seen before.

“AI System” is a machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, content (such as text, images, videos, or sounds), or other output influencing decisions made in real or virtual environments. *AI Systems* are designed to operate with varying levels of autonomy.

“AIS Program” refers to the controls and processes that an insurer adopts and implements to ensure the responsible use of *AI Systems*. An *AIS program* may be comprised of general guidelines, governance framework, risk management and internal controls, and third-party oversight.

“Artificial Intelligence (AI)” refers to a branch of computer science that uses data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement, or the capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement. This definition considers machine learning to be a subset of artificial intelligence.

“Augmentation” refers an *AI System* that suggests an answer and/or advises a human who is making a decision.

“Automation” refers to an *AI System* that does not involve human intervention.

“Consumer Impact” refers to a decision by an Insurer that is subject to insurance regulatory standards enforced by the Department.

“Direct Consumer Impact” refers to an *AI System* that *augments* or *automates* human decision-making whose output has the potential to create final consumer outcomes including but not limited to underwriting including tier and coverage eligibility determinations, premium determinations, ratemaking, and claims denial and settlement-related decisions. This has ~~less applicability to~~ **does not include** *AI Systems or AI Models* used for internal analytics, model monitoring, and back-office tooling with no consumer-visible output or other internal operational or administrative functions that do not determine or materially affect an outcome for a specific consumer.

Commented [A14]: Recommend this change to clarify these are out of scope.

“Externally Trained Models” refers to transferred learnings from pre-trained models developed by a third party on external reference datasets.

“Generative Artificial Intelligence (Generative AI)” refers to a class of *AI Systems* that generate content in the form of data, text, images, sounds, or video, that is similar to, but not a direct copy of, pre-existing data or content.

“Generalized Linear Model (GLM)” refers to a statistical modeling technique that extends linear regression to by relating predictive variables to an outcome variable using an assumed additive or multiplicative relationship through a link function and an appropriate probability distribution.

“Inherent Risk” refers to the risk of economic loss or inaccurate financial reporting before **considering risk-mitigation strategies or** internal controls.

Commented [A15]: Recommend this addition.

“Internally Trained Models” refers to models developed from data internally obtained by the company.

“Machine Learning (ML)” refers to a field within artificial intelligence that focuses on the ability of computers to learn from provided data without being explicitly programmed.

“Material Financial Impact” should be determined based on either the company’s materiality threshold which should be disclosed when responding to Supplement related inquiries or based on the threshold to be provided by regulators.

“Materiality” refers to the dollar amount above which the examiner’s perspective of an insurer’s financial position will be influenced.

“Model Drift” refers to the decay of a model’s performance over time arising from underlying changes such as the definitions, distributions, and/or statistical properties between the data used to train the model and the data on which it is deployed.

“Neural Network Model” refers to a machine learning model that mimic the complex functions of the human brain. These types of models consist of interconnected nodes or neurons that process data, learn patterns and enable tasks such as pattern recognition and decision-making, including but not limited to: Single/multi-layer perceptrons/fully connected networks (MLPs/FCs), Deep Learning (DL), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory Neural Networks (LSTMs), Sequence Models, Large Language Models (LLMs), and Reinforcement Learning Models (RLs).

“Predictive Model” refers to the mining of historic data using algorithms and/or machine learning to identify patterns and predict outcomes that can be used to make or support the making of decisions.

“Residual Risk” refers to an assessment of risk after considering risk-mitigation strategies or internal controls.

Commented [A16]: Recommend this addition.

“Support” refers to an *AI System* that provides information but does not suggest a decision or action to a human.

“Third Party” for purposes of this Supplement means an organization other than the insurance company that provides services, data, or other resources related to AI.

“Validation Method” refers to the source of the reference data used for validation, whether Internal, External, or Both.

“Use Case” refers to a description of a specific function in which a product or service is used.

Operations

Claims - Examples: Claim assignment, triage/fast-tracking, individual/bulk claim reserving including loss estimation, imaging/video analysis, fraud detection, litigation, estimation of closure rates, salvage/subrogation, examination/report ordering.

Customer Service - Examples: Agent/broker/internet/customer service interaction (chatbots), online/smart phone apps, loss prevention/risk mitigation advice, payment plans, complaints.

Marketing - Examples: market research, target advertising, market/coverage expansion, customer segment target marketing, demand modeling, agent/broker incentive plans, up/cross-selling.

Other: Cyber Security, Strategic Operations, Reserving, Investments, Capital Management, Financial Reporting, Reinsurance, Legal, Legal Exposure, Reputation Risk.

Ratemaking/Pricing - Examples: Development of overall/base rates, expense/loss loadings, estimation of trends and loss development, development of manual rating factors, tiering criteria, insurance credit scoring, territory boundary definitions, numeric/categorical level groupings and interactions, individual risk rating, telematics/UBI, price optimization, schedule rating factors.

Underwriting - Examples: Policy/coverage acceptance or eligibility, company placement/tiering, schedule rating, decisions based on telematics/UBI, report ordering, retention modeling, inspections, anomaly detection.

DRAFT

September 29, 2026

NAIC Big Data and Artificial Intelligence (H) Working Group

1100 Walnut Street

Suite 1500

Kansas City, MO 64106

Attn: Scott Sobel (ssobel@naic.org) and Miguel Romero (maromero@naic.org)

RE: *APCIA Comments on the AI Risk Evaluation Supplement version 5.0*

Chair Houdek and Members of the Big Data and Artificial Intelligence (H) Working Group:

The American Property Casualty Insurance Association (APCIA) appreciates the opportunity to comment on Version 5.0 of the AI Risk Evaluation Supplement. We recognize the significant effort that regulators and pilot participants have invested in this process.

APCIA supports the Working Group's objective of developing a practical supervisory tool to help regulators identify and assess meaningful consumer and solvency risks associated with AI systems. The revisions reflected in Version 5.0 demonstrate meaningful progress, particularly the clearer articulation that the Supplement is intended to complement existing examination authorities rather than serve as a standalone regulatory framework, the increased emphasis on a risk-based approach, and refinements to the materiality and inherent risk concepts. We also appreciate the Working Group's continued willingness to solicit feedback and make substantive revisions throughout the pilot process. The changes made to date demonstrate the value of that collaborative approach and have improved both the clarity and usability of the Supplement. We remain committed to working constructively with regulators as the Supplement continues to evolve.

While Version 5.0 represents a substantial improvement over prior drafts, APCIA believes additional refinement remains warranted before the Supplement is finalized and broadly implemented. In particular, further clarification regarding the scope of covered models, the treatment of generalized linear models (GLMs), implementation expectations, and certain reporting and governance provisions would help ensure the Supplement remains appropriately risk-based, operationally workable, and focused on technologies that present the most significant consumer and regulatory concerns. Because we view the treatment of GLMs as a particularly significant issue with implications extending beyond this Supplement, we are contemporaneously submitting separate comments from Pinnacle Actuarial Resources addressing the application of the Supplement to generalized linear models used for pricing.¹ We have also included selected drafting suggestions in the Appendix.

¹ See Pinnacle Actuarial Resources, Inc., *Comments on the Application of the NAIC Artificial Intelligence Risk Evaluation Supplement to Generalized Linear Models Used for Pricing* (Sept. 29, 2026).

Generalized Linear Models (GLMs)

APCIA remains concerned that the current draft continues to blur the distinction between traditional statistical models and artificial intelligence systems. GLMs have been used, reviewed, and regulated for decades. Extending AI-specific oversight frameworks to encompass these models risks creating uncertainty regarding the intended scope of the Supplement while diverting regulatory and industry resources away from technologies that present the novel risks the Working Group is seeking to evaluate.

We support a risk-based framework that focuses enhanced regulatory attention on technologies that present novel governance, transparency, consumer protection, or operational risks not already addressed through existing regulatory frameworks. Conventional GLMs are generally transparent, explainable, and subject to well-established governance, actuarial review, and regulatory oversight processes. Similar concerns have been raised internationally. In April 2026, EIOPA observed that generalized linear models and related actuarial techniques are subject to established governance frameworks and do not present many of the "black-box" characteristics that motivate enhanced AI-specific oversight requirements.² We believe these considerations are equally relevant to the Working Group's evaluation of the treatment of conventional GLMs under the Supplement.

More broadly, APCIA encourages the Working Group to focus on whether a model presents novel risks warranting enhanced AI-specific review, rather than on the specific methodology used to develop the model. Additional guidance regarding the meaning of "explainability" and how that concept should be evaluated in practice would help promote more consistent application of the Supplement. Applying AI-specific reporting and review requirements to models that are already transparent, explainable, and subject to established governance frameworks could substantially expand the scope of the Supplement without a commensurate regulatory benefit, create uncertainty regarding the treatment of longstanding actuarial practices, and dilute attention from technologies that present the most significant emerging risks.

Given the importance of this issue, we have submitted separate comments discussing the distinction between generalized linear models used for pricing and AI systems.³ We believe further consideration should be given to whether conventional GLMs should be distinguished from AI systems for purposes of the Supplement.

² See [Letter to EU institutions on AI Act and EU Insurance legislation proposal for clarifying application of the AI Act \(1\)](#), European Insurance and Occupational Pensions Authority (EIOPA) (Apr. 13, 2026), recommending that generalized linear models, including linear and logistic regression, be excluded from high-risk AI requirements because they are transparent, interpretable, subject to established actuarial governance frameworks, and do not exhibit the "black-box" characteristics that drive many AI governance concerns.

³ See Pinnacle Actuarial Resources, Inc., *Comments on the Application of the NAIC Artificial Intelligence Risk Evaluation Supplement to Generalized Linear Models Used for Pricing* (Sept. 29, 2026).

Scope, AI Model Reporting, and Use Cases

APCIA appreciates the Working Group's efforts to distinguish among AI systems, AI models, and use cases. However, additional guidance would improve consistency in how insurers identify and report these concepts.

In practice, a single business process may involve multiple systems, models, and use cases. We recommend that the Working Group include illustrative examples demonstrating how insurers should inventory and report AI systems, AI models, and use cases where a single platform contains multiple models or supports multiple business functions. We also recommend expressly confirming that insurers may leverage existing inventories, model documentation, and governance artifacts when responding to Exhibit A. Additional clarification regarding the practical relationship between AI Systems and AI Models would help insurers structure inventories more consistently and reduce the risk of inconsistent reporting across companies and jurisdictions.

The Supplement should remain focused on AI systems used within regulated insurance activities and the consumer and financial risks arising from those activities. Consistent with that objective, we recommend reconsidering the inclusion of broad categories such as "Other AI/ML Models" in Exhibit A. As currently drafted, this category could function as an open-ended catchall, expanding reporting obligations and creating uncertainty regarding the intended scope of the Supplement. We also recommend reconsidering how marketing-related use cases are reflected in Exhibit A, as those activities may present different risk profiles than core insurance functions such as underwriting, ratemaking, claims, or fraud detection.

Additional guidance regarding the intended scope of machine learning models and traditional predictive models would also be helpful, particularly regarding how materiality, business impact, and existing governance programs should inform whether additional review is warranted. APCIA supports language recognizing that not all AI use cases warrant additional inquiry and that regulators may appropriately limit or conclude their review following Exhibit A. We generally supports a sequential review approach in which Exhibit A serves as a threshold assessment to help determine whether additional exhibits are necessary. This approach reinforces the Supplement's risk-based design by focusing more detailed inquiries on systems presenting higher levels of risk, complexity, or materiality.

APCIA also encourages clarification regarding instructions that regulators may "customize" portions of the Supplement. The Working Group should clarify that customization is intended to allow regulators to narrow or tailor existing information requests for a particular examination, rather than introduce new requests beyond those reflected in the Supplement. Additional clarity would promote more consistent implementation across jurisdictions.

The pilot process has demonstrated that responding to the Supplement can require significant time and resources. We therefore appreciate the Working Group's recognition that the Supplement is intended to be used in connection with examinations rather than as a standalone reporting exercise. To further reinforce this principle, we recommend removing "on an ongoing basis" from the first sentence of the Intent section

and adding language emphasizing the importance of coordinating requests through the examination process.

Finally, we question the value of requiring reporting based on a model's implementation date when insurers are already reporting systems currently in use. Because implementation milestones may be defined differently across organizations and projects, reliance on implementation dates may reduce consistency without materially improving regulatory understanding.

Definitions and Terminology

APCIA appreciates the Working Group's efforts to further refine key definitions and improve consistency throughout the Supplement. However, several definitions would benefit from additional clarification. In particular, we would welcome further explanation regarding the purpose and intended application of the newly added "AI Model" definition and its relationship to existing concepts such as AI Systems, predictive models, and GLMs. While we appreciate the effort to distinguish among these concepts, we are concerned that the introduction of a new AI Model definition may add complexity and create uncertainty regarding reporting expectations. To the extent possible, we encourage consistency with terminology and definitions already reflected in existing NAIC guidance, including the AI Model Bulletin.

APCIA continues to hear significant questions regarding the meaning of "consumer" for purposes of the Supplement. While Version 5.0 adds definitions for Consumer Impact and Direct Consumer Impact, additional clarification regarding who constitutes a consumer, when an outcome is considered consumer-facing, and how those concepts relate to other terms used throughout the Supplement, such as Adverse Consumer Outcome and adverse consumer impact, would promote more consistent interpretation and reporting across insurers and jurisdictions. Additional clarification regarding the meaning of regulated insurance activities would also be beneficial, including how that concept relates to Direct Consumer Impact, material financial impact, and the overall scope of the Supplement.

Finally, we recommend clarifying that AI systems used exclusively for internal analytics, model monitoring, back-office functions, or other internal operational purposes with no consumer-facing output or effect are generally outside the intended scope of Direct Consumer Impact. The current statement that such uses have "less applicability" leaves uncertainty regarding whether those systems are intended to fall within scope and could benefit from more definitive guidance.

Model Inventories and Existing Documentation

APCIA supports the flexibility allowing insurers to provide existing AI or model inventories. To reduce duplicative reporting and better align with the Supplement's risk-based objectives, the Supplement should expressly confirm that insurers may rely on existing model inventories, governance documentation, model risk management documentation, and other comparable internal artifacts containing substantially similar information, rather than creating new reports solely for purposes of responding to the Supplement.

Materiality and Risk Assessments

APCIA supports the Working Group's efforts to align materiality and risk concepts with existing NAIC examination guidance and appreciates the emphasis on inherent risk. We also appreciate the flexibility afforded by the revised materiality provisions.

We question whether further prescriptive definition of materiality within the Supplement is necessary. Given the inherently context-specific nature of materiality assessments and the flexibility reflected elsewhere in the Supplement, overly prescriptive definitions may create confusion rather than promote consistency. Any materiality framework should remain grounded in existing examination guidance, company governance frameworks, and the facts and circumstances of the particular use case under review.

To further support implementation, additional examples illustrating how materiality thresholds may be applied in practice would be helpful. To promote consistency and leverage existing governance frameworks, we believe that where an insurer has established and documented materiality thresholds through its existing risk management, model governance, or control processes, those thresholds generally should be utilized unless regulators identify a specific reason that additional review is warranted.

Third-Party AI Systems

Version 5.0 appropriately recognizes the role of third-party providers within the AI ecosystem. However, insurers' ability to obtain information about externally developed models may vary significantly depending on the product, vendor relationship, and proprietary protections associated with the underlying technology.

We recommend clarifying that oversight, validation, transparency, and documentation expectations for third-party systems should be risk-based and appropriately calibrated to the information reasonably available to insurers. The Supplement should recognize that insurers may not have visibility into all aspects of a vendor's model development, training data, or proprietary methodologies. In such circumstances, regulatory expectations should be calibrated to information available through contractual rights, vendor oversight processes, and other commercially reasonable means.

Additional guidance regarding the minimum information insurers are expected to obtain from third-party vendors would help companies better understand how to satisfy governance, validation, transparency, and oversight expectations for externally developed systems. We also recommend clarifying that explainability and transparency expectations should be commensurate with both the technology involved and the information reasonably available to the insurer, particularly for externally developed models and tools.

Exhibit D and Data Sources

APCIA appreciates the efforts to refine Exhibit D and make data-related inquiries more targeted. However, additional clarification would be helpful regarding whether regulators are seeking information about data

elements directly used by a model, data elements used to infer a characteristic, or both. The distinction is important. For example, there is a meaningful difference between using information directly provided by an applicant and using other data sources or modeling techniques to infer a similar characteristic. Clarifying the intended focus of Exhibit D would promote more consistent responses across insurers and help ensure the information collected aligns with the Supplement's risk-based objectives.

We also recommend clarifying that reporting should focus on data sources and data elements used in production models, rather than data evaluated during development, testing, or experimentation but not ultimately incorporated into the final model. Additional guidance regarding whether Exhibit D is intended to be completed once for all relevant models or separately for each model under review would further improve consistency.

APCIA supports continued flexibility in how Exhibit D is applied. Exhibit D should remain focused on models that are high-risk, material, or otherwise warrant enhanced regulatory review. As currently drafted, the information requests could become extensive when applied across multiple models and lines of business. APCIA therefore supports the Working Group's efforts to tailor requests based on the insurer's business, risk profile, and examination objectives.

We also recommend providing additional guidance regarding the scope of the "other non-traditional data elements" category, which may otherwise be interpreted broadly and lead to inconsistent reporting across insurers.

Finally, given the ongoing pilot process and the operational complexity associated with Exhibits C and D, we encourage the Working Group to continue evaluating whether certain information requests provide incremental supervisory value beyond information already available through existing examinations and governance documentation.

Agentic AI

APCIA appreciates the addition of a definition for agentic AI and agrees that regulators should remain informed about emerging technologies and evolving use cases. Given the rapidly evolving nature of agentic AI technologies, we support continued monitoring and discussion of these systems as use cases mature.

At the same time, we would welcome additional clarity regarding the purpose of this provision. Specifically, the Working Group should clarify whether the inclusion of agentic AI is intended primarily to facilitate regulator awareness and inventorying of emerging technologies or whether distinct governance, documentation, or oversight expectations are anticipated for such systems. Illustrative examples would also promote more consistent interpretation by helping distinguish between process automation, code generation and decision-support tools, and systems that independently initiate actions, make decisions, or provide recommendations subject to human review.

Pilot Program Findings and Implementation Timeline

In addition to the substantive revisions discussed above, APCIA encourages the Working Group to continue providing transparency regarding implementation of the Supplement through the ongoing pilot process.

We appreciate the significant effort invested by participating regulators and insurers throughout the pilot. As the process moves forward, additional insight regarding how the Supplement has been applied in practice, lessons learned, and the information regulators found most useful would promote broader industry understanding of regulatory expectations and assist insurers in preparing for potential implementation.

We also encourage continued communication regarding the anticipated timeline for finalization and use of the Supplement. Clarity regarding next steps, implementation expectations, and the relationship between the pilot process and eventual adoption of the final Supplement would help insurers allocate resources appropriately and continue engaging constructively throughout the process.

Conclusion

APCIA appreciates the substantial improvements reflected in Version 5.0 and the Working Group's continued engagement with industry throughout the pilot process. The revisions demonstrate meaningful progress toward a more focused, risk-based, and operationally workable framework.

At the same time, additional refinement remains warranted before the Supplement is finalized and implemented. In particular, we encourage further consideration of the treatment of generalized linear models, the distinction among AI systems, AI models, and use cases, the application of materiality concepts, key definitions and terminology, expectations for third-party systems, and the scope and implementation of Exhibits C and D.

We look forward to continuing to work with regulators and other stakeholders as the Supplement evolves and welcomes further dialogue regarding these comments.

Thank You,

Kristin Abbott

Sr. Director and Counsel, Cyber & Privacy

Dave Snyder

Vice President, International & Counsel

Appendix: Selected Drafting Suggestions

APCIA has included selected drafting suggestions for the Working Group's consideration on several issues discussed in these comments. These examples are illustrative and are not intended to constitute a comprehensive redline of the Supplement.

Redline Convention: Text shown in ~~red strikethrough~~ indicates proposed deletions. Text shown in green underline indicates proposed additions.

Exhibit A - Existing Inventories

~~If a company has an existing AI Systems inventory, it may suggest submission of the inventory in lieu of completing Exhibit A.~~

If a company maintains an existing AI Systems inventory, model inventory, or equivalent model risk management documentation, regulators may accept such documentation in lieu of requiring a separate inventory format.

Exhibit B.3p- Third-Party AI Systems

~~Governs, monitors, tests and ensures transparency of AI Systems developed by vendors.~~

Governs, monitors, and performs risk-based oversight of AI Systems developed by vendors, based on information reasonably available to the insurer.

Exhibit D - Data Sources

~~To obtain detailed information of the source(s) and type(s) of data used in an AI model(s)...~~

To obtain detailed information regarding data elements directly used by, or relied upon to infer characteristics within, AI model(s), as applicable to the regulator's inquiry.

Only data sources used in production models should be reported. Data elements evaluated during development, testing, experimentation, or validation but not incorporated into the final model need not be reported.

Definitions- Agentic AI

Suggested addition to the definition: For purposes of this Supplement, the inclusion of Agentic AI is intended to facilitate awareness and identification of emerging technologies and should not be interpreted as creating distinct governance or reporting requirements absent further regulatory guidance.

Definitions- Direct Consumer Impact

~~This has less applicability to AI Systems used for internal analytics, model monitoring, and back-office tooling with no consumer-visible output...~~

AI Systems used exclusively for internal analytics, model monitoring, back-office tooling, or other administrative functions with no consumer-visible output and that do not determine or materially affect an outcome for a specific consumer are generally outside the scope of Direct Consumer Impact.

Definitions- AI Model

The distinction between AI Models, AI Systems, predictive models, and statistical models should be interpreted consistently with existing NAIC guidance, and regulators may wish to provide examples illustrating how these concepts should be applied for purposes of reporting under this Supplement.

Intent

Regulators may use this Supplement ~~on an ongoing basis~~ to evaluate AI Systems used by insurers...

The Supplement is intended to be used in connection with existing examination processes and should be applied in a coordinated and risk-based manner to avoid unnecessary duplication of regulatory requests.

The Supplement is intended to focus on AI Systems used in regulated insurance activities and on systems that directly affect consumers or present material financial, operational, or compliance risks.

Instructions

~~Regulators may wish to first use Exhibit A to determine if further inquiry using Exhibits B-D is necessary.~~

Regulators should consider using Exhibit A to determine whether additional inquiry using Exhibits B-D is necessary and should tailor any subsequent requests based on the nature, complexity, materiality, and risk profile of the AI systems identified.

Materiality

Regulators may establish a materiality threshold or rely on one provided by the insurer. Where an insurer has established and documented materiality thresholds as part of its existing governance, risk management, or model risk management framework, those thresholds should generally be used absent a specific supervisory reason to apply a different threshold.

Regulator Instructions

Regulators may customize requests. Customization should be limited to narrowing, tailoring, or selecting relevant portions of the Supplement based on examination objectives and should not be interpreted as authorizing the creation of new information requests outside the Supplement.

September 29, 2026

Commissioner Nathan Houdek, Chair
Commissioner Doug Ommen, Co-Vice Chair
Deputy Commissioner Mary Block, Co-Vice Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1000
Kansas City, MO 64106-2197

Submitted via email to Scott Sobel (ssobel@naic.org) and Miguel Romero (maromero@naic.org)

RE: Artificial Intelligence Risk Evaluation Supplement Version 5.0

Dear Chair Houdek and Vice Chairs Ommen and Block:

The Blue Cross Blue Shield Association (BCBSA) appreciates the opportunity to provide comments on the National Association of Insurance Commissioners (NAIC)'s Artificial Intelligence (AI) Risk Evaluation Supplement version 5.0.

BCBSA is a national federation of independent, community-based and locally operated Blue Cross and Blue Shield companies (Plans) that collectively provide health care coverage for one in three Americans. BCBS Plans contract with 97% of hospitals and 83% of doctors across the country and serve those who are covered through Medicare, Medicaid, an employer, or purchase coverage on their own.

BCBSA appreciates the Working Group's continued engagement with stakeholders and the progress reflected in Version 5.0 of the AI Risk Evaluation Supplement. We are particularly supportive of the Supplement's clearer grounding in the existing Market Regulation Handbook, Financial Condition Examiners Handbook and Financial Analysis Handbook, which helps position the Supplement as an extension of established regulatory processes rather than a parallel framework. We also appreciate the clearer treatment of the distinction between AI systems and AI models, and the Supplement's greater emphasis on materiality and inherent risk, which lay important groundwork for focusing oversight on the AI systems that pose the greatest risk to consumers and to financial soundness.

Drawing on Blue Plans' experience with the Supplement during the pilot, we have identified areas where further refinement would strengthen the Supplement and better serve its purpose of assisting regulators in identifying and assessing the financial and consumer risks arising from the use of AI systems. Together, these recommendations seek to focus regulatory attention on the AI systems that pose the greatest risk, to promote consistent application across states and to avoid unnecessary burden that could divert resources from the responsible AI innovations

that improve care and service for consumers. To advance these aims, we offer the following key recommendations:

- **Strengthen the Supplement's risk-based approach, consistent with the NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers (Model Bulletin).** The Supplement already contemplates using Exhibit A to determine whether the more detailed inquiries in Exhibits B through D are warranted, but it offers limited guidance on when these additional exhibits are necessary. As a result, companies may be asked to document their full range of AI use, including low-risk internal and administrative tools, rather than the systems that meaningfully affect consumers or financial soundness. Directing regulators to reserve the substantive exhibits for higher risk uses and focusing the Supplement on AI applications that materially affect consumer outcomes such as coverage eligibility and benefit determinations would concentrate oversight where it matters most, reduce unnecessary burden and produce more useful information for regulators. In particular, the definition of "Direct Consumer Impact" already draws on a materiality concept, treating functions that do not materially affect an outcome for a specific consumer as having less applicability. Adopting the label "Material Consumer Impact," and carrying that concept into the operative standard would better signal the intended scope to the companies and examiners who rely on it. This graduated, materiality-based approach reflects the Model Bulletin's core principle that oversight should be commensurate with the degree of potential harm to consumers.
- **Promote a consistent, baseline application of the Supplement across states, including by setting common expectations for the information requested, the evidence expected and response timeframes.** BCBSA recognizes that individual states need the flexibility to tailor the Supplement to meet their specific regulatory needs. A shared baseline lowers the burden of implementing the Supplement for states, while leaving each state free to adapt it to its specific needs. A common baseline for states to build from also provides consistency for companies responding, which can drive more meaningful reporting that states can act on. A reasonable and consistent response timeframe, of at least 30 business days, would further support complete and accurate submissions.
- **Refer the Supplement to the Market Regulation and Consumer Affairs (D) Committee and the Financial Condition (E) Committee, for incorporation or appropriate cross-reference within existing regulatory Handbooks.** BCBSA supports the Working Group's clarification that the Supplement is intended to complement, not operate separately from existing NAIC market conduct and financial regulatory processes. The Market Regulation Handbook, Financial Condition Examiners Handbook, and Financial Analysis Handbook are the resources examiners and analysts rely on for related inquiries and information requests. Because the Supplement may be used in market conduct examinations, financial condition examinations, and financial analysis, the Handbooks should provide guidance on when its use is appropriate, how inquiries should be scoped based on risk and materiality, and how its use should be coordinated across regulatory functions and jurisdictions. This approach is consistent with the responsibilities of the relevant D and E Committee groups and the NAIC's prior commitment to cross-committee coordination on the Supplement. Further, it also follows

NAIC precedents. For example, during implementation of the Own Risk and Solvency Assessment (ORSA) framework, the NAIC referred the new assessment tool to the groups responsible for the Financial Analysis and Financial Condition Examiners Handbooks to integrate it into existing risk-focused supervisory processes. We recommend a similar referral here to promote consistent, proportionate and appropriately scoped use of the Supplement within the NAIC's existing regulatory framework.

- **Remove the separate quantification of generalized linear models (GLMs) from Exhibit A or, at a minimum, require GLM reporting only if it is limited to models that materially affect consumers or financial soundness.** Version 5.0 asks companies to separately count GLMs for the first time. As a well-understood technique with a long-established use in insurance and whose outputs can be easily traced to their inputs, GLMs do not present the similar risks that characterize the generative and agentic applications the Supplement targets. Separately counting GLMs shifts focus away from the higher risk uses that are more likely to materially affect consumer outcomes and warrant regulatory attention. Removing the separate GLM count, or, at a minimum, scoping it to models with material consumer or financial impact, would keep the Supplement focused on the AI that presents genuine risk.

In what follows, we expand on these key recommendations and offer additional recommendations, organized to follow the structure of the Supplement, beginning with instructions and general guidance, followed by exhibit-specific recommendations and concluding with definitions. Thank you for the opportunity to provide feedback on the AI Risk Evaluation Supplement. If you have any questions, please contact Randi Chapman at Randi.Chapman@bcbsa.com.

Sincerely,



Clay S. McClure
Senior Director, State Affairs
Blue Cross Blue Shield Association

BCBSA'S DETAILED COMMENTS ON THE NAIC ARTIFICIAL INTELLIGENCE RISK EVALUATION SUPPLEMENT VERSION 5.0

Instructions and General Guidance Recommendations

Recommendation #1:

Establish a reasonable and consistent response timeframe of at least 30 business days for inquiries made under the Supplement, with extensions available upon request.

Rationale:

The Supplement asks companies for detailed information that often requires coordination across technical, actuarial, compliance and governance functions. During the pilot, the time allowed to respond varied considerably, with some states requiring responses in fewer than two weeks. Timeframes that do not account for the cross-functional nature of the information requested can compromise the completeness and accuracy of the responses, which ultimately affects the quality of the information available to regulators. A consistent response timeframe of at least 30 business days, with extensions available upon request, would set a workable and predictable expectation, support more complete and accurate submissions and reduce the follow-up and rework that incomplete responses create for regulators and companies alike.

Recommendation #2:

Specify within the Confidentiality section how submitted information will be maintained, whether and how it may be shared across financial and market conduct examiners and across states, and confirm that information will not be published or used adversely against a company.

Rationale:

The Supplement requires companies to provide detailed and often sensitive information about governance processes, models and data. Version 5.0's confidentiality section is an important addition, but it remains general as to where information is held, how and with whom it may be shared, and whether it may be published. Clear protections are particularly important for the most sensitive categories of information the Supplement draws in, such as model documentation and training data descriptions, model and AI system inventories, third-party vendor identities and governance artifacts including risk classifications and incident records. Because the Supplement contemplates coordination among financial and market conduct examiners and across states, the absence of clear information-sharing parameters creates a real risk that sensitive and proprietary information could be disseminated beyond its intended purpose. Where information is shared across states without a common protection standard, the protections effectively default to those of the least protective jurisdiction. Clear confidentiality parameters would encourage companies to document and share sensitive assessments candidly, improving the quality and completeness of the information available to regulators. Clear protections and defined limits on how this information may be used and shared are essential to maintaining trust between regulators and regulated entities.

Recommendation #3:

Include earlier and more prominently in the Supplement's general instructions, the option for companies to satisfy inquiries using existing materials — including an existing AI systems inventory and existing governance documentation — in lieu of completing the exhibits.

Rationale:

Version 5.0 allows a company to suggest submitting an existing AI systems inventory in lieu of completing Exhibit A, which we support. However, that option currently appears only within Exhibit A, and companies would benefit from seeing it, and similar options, in the Supplement's general instructions. Confirming that existing documentation can satisfy relevant requests, and presenting that option up front, would allow companies to meet regulatory objectives using materials they already maintain and rely on, rather than creating new, examination-specific documentation. This accelerates regulator review, promotes more efficient implementation across states and reduces compliance burden.

Recommendation #4:

Clarify in the Instructions that the Supplement is intended to help regulators evaluate the governance and risk management of AI systems, and not to serve as a mechanism for the routine collection or cross-company comparison of enterprise AI inventories.

Rationale:

The Instructions state that the Supplement is intended to enhance a regulator's understanding of the extent of AI systems use and to assess potential risk to consumers and the company, and that responses will be considered in identifying the insurer's inherent risks. Making explicit that this is intended for risk and governance evaluation, rather than inventory collection, the Supplement's purpose would help ensure the tool is applied consistently with that intent. It would also reflect the practical reality that AI inventories are dynamic and depend heavily on how each organization defines, groups and tracks its AI assets, which makes them appropriate for a risk assessment but of limited value for direct comparison across companies. This clarification would reinforce the risk-based approach the Supplement already contemplates and keep the focus on how companies govern and manage AI risk.

Exhibit A: Part One - Quantify Regulated Entity's Use of AI Systems

Recommendation #1:

Replace "Direct Consumer Impact" with "Material Consumer Impact" throughout the Supplement, including the Exhibit A column heading over columns 4 through 6, the Exhibit A Company Instructions and the corresponding definition.

Rationale:

Materiality is an established concept in existing regulatory frameworks. For example, the NAIC Financial Condition Examiners Handbook uses materiality to focus examiner attention, and the Supplement draws its own materiality definition from that Handbook. Its use here is also consistent with the risk-proportionate approach in the Model Bulletin, as well as nationally

recognized frameworks such as the National Institute of Standards and Technology (NIST) AI Risk Management Framework. The term "material consumer impact" would capture AI outputs that materially affect a consumer outcome, such as coverage eligibility, benefit determinations and claims decisions, and it would exclude routine, internal or purely informational tools, such as customer service FAQ chatbots that retrieve and summarize pre-approved, static information.

The Supplement's own definition of "Direct Consumer Impact" already focuses on whether an AI output affects a consumer outcome. It applies to AI outputs that have the potential to create a final consumer outcome, and it treats internal analytics, model monitoring and back-office tools that do not determine or materially affect an outcome for a specific consumer as having less applicability. However, the "direct" label does not convey a focus on whether an AI output materially affects a consumer outcome, and because companies and examiners look first to the column heading when scoping their responses, a heading that matches the definition would support more accurate and consistent reporting. Adopting "material" also parallels the existing "Material Financial Impact" heading over columns 7 through 9, improving internal consistency. By contrast, "direct" is not a concept used in the Model Bulletin, and does not track the materiality-based standard the definition applies. Version 5.0's more detailed definition of "Direct Consumer Impact" is a meaningful step toward appropriate scoping, and adopting "material" achieves that scoping more reliably and in closer alignment with the Model Bulletin.

Recommendation #2:

If NAIC retains the term "Direct Consumer Impact," BCBSA recommends incorporating an explicit materiality standard into the definition, building on the materiality concept it already references, so that the category captures only AI whose outputs materially affect consumer outcomes.

Rationale:

While BCBSA recommends replacing "Direct Consumer Impact" with "Material Consumer Impact" for the reasons described above, the current definition references materiality, treating tools that do not materially affect an outcome for a specific consumer as having "less applicability." Making that concept explicit and operative in the definition, rather than expressing it only as a matter of "less applicability," would achieve much of the same benefit under the current term. This standard would be qualitative, focusing on whether an AI system's output materially affects a consumer outcome, and is therefore distinct from the dollar-based materiality threshold used for Material Financial Impact. It could be incorporated without disturbing the Supplement's existing structure. Even as a targeted change, it would sharpen the Supplement's focus on the AI of greatest consequence to consumers and support more consistent scoping under the current term.

Recommendation #3:

Remove the separate quantification of generalized linear models (GLMs) from Exhibit A.

Rationale:

Version 5.0 introduces GLMs into Exhibit A for the first time and asks companies to separately quantify them in columns 4 and 7. GLMs are among the most transparent and interpretable modeling techniques used in insurance. Because the relationship between each input and the resulting output is directly observable, GLMs do not present the similar risks that distinguish the generative and agentic applications the Supplement is designed to surface. Requiring GLMs to be counted separately directs attention toward a lower-risk technique rather than toward the higher-risk uses that warrant regulatory focus, diffusing regulators' capacity to substantively review the impactful application. It also risks overstating the scope of AI-related risk by folding a long-established statistical method, used across a range of insurance functions, into an inventory intended to identify AI risk. Removing the separate count would keep Exhibit A focused on the AI that presents genuine risk.

Recommendation #4:

Clarify in Exhibit A how companies should count AI systems and AI models, and add disclosure language acknowledging that a single AI system may appear under more than one program area.

Rationale:

The table asks companies to report counts by "Use of AI System in Operations or Program Area," and to separately quantify AI systems in columns 2 and 3 and AI models in columns 4 through 9. It is not clear whether a single AI model used across multiple systems should be counted once or multiple times, how models should be counted when one AI system comprises several models, or how to treat multi-purpose and enterprise-wide systems that span program areas, such as fraud detection tools that operate across underwriting and claims. Absent guidance, companies will categorize and count inconsistently, and regulators may misinterpret the totals, particularly where a single system is counted under several program areas. Clarifying the counting methodology, and stating expressly that AI systems may be listed under more than one program area and that totals across program areas should not be assumed to represent unique systems, would improve the accuracy and comparability of responses and prevent misleading aggregation.

Exhibit B: (Narrative) AI Systems Program

Recommendation #1:

Remove the reference to staffing reductions from Exhibit B, Narrative Question 2(b).

Rationale:

Question 2(b) addresses AI systems that generate a direct consumer impact, and asks companies to include, among other things, whether there were staffing reductions as part of the AI system implementation. While staffing decisions may accompany an AI implementation, they relate to workforce and operational planning rather than to whether an AI system's output affects a specific consumer. As a result, this element may be a less direct indicator of consumer impact than the other factors identified in Question 2(b), such as whether there is a human in

the loop or how exceptions and errors are resolved. Focusing Question 2(b) on those factors would keep it closely aligned with the consumer impact it is intended to capture.

Recommendation #2:

Clarify Narrative Question 3(b), that validation and testing of third-party vendor-supplied AI systems may be satisfied by either the vendor or the insurer, including for AI systems supplied by fourth parties.

Rationale:

Many AI systems are supplied by third-party vendors that retain proprietary rights over their models, and insurers frequently do not have access to the technical detail necessary to independently test every element. In these circumstances, it should be acceptable for a company to rely on vendor testing and assurance reports rather than duplicating work that the insurer cannot reasonably perform. This is especially true where AI is supplied by a fourth party, such as a subcontractor or AI operating under the vendor's own agreements, because the insurer has no direct relationship with that party and even less ability to access or test the underlying model. Confirming that either the vendor or the insurer may satisfy these requirements would align the question with how third-party AI oversight works in practice and would avoid setting an expectation that insurers cannot meet for models they did not build and cannot fully access.

Exhibit B: (Checklist) AI Systems Program

Recommendation #1:

Revise Exhibit B, Checklist item (3m) so that it addresses consumer complaints related to consumer outcomes involving material AI systems and confirms that such complaints are captured through existing complaint and remediation frameworks.

Rationale:

Health insurers already operate under robust, technology-neutral requirements for tracking and resolving consumer complaints, and existing complaint systems are generally organized by the nature of the issue, rather than by the underlying technology involved. Asking companies to identify complaints "resulting from" AI systems would require them to determine and tag whether AI was involved in each complaint. This would produce data of questionable accuracy and limited regulatory value, because AI is often one of several factors in a process and its role may be indirect or not apparent, so companies would apply the "resulting from" standard inconsistently, and a complaint typically reflects the outcome a consumer experienced, such as a denial or an access issue, rather than the tools involved in reaching it. Consumers are already protected through established complaint, grievance and appeal pathways regardless of whether AI or a traditional process was involved. Scoping the item to consumer outcomes involving material AI systems and confirming that such complaints are addressed through existing frameworks, would keep the item workable and consistent with how consumer protection operates in practice. Consumers are best served by these existing pathways, which provide reliable access to review, appeal and remediation focused on correcting the outcome they

experienced rather than on the technology behind it. Consistent with this, BCBSA suggests the following edit (additions in bold, deletions in strikethrough):

*3m. Considers consumer complaints **related to consumer outcomes in material** ~~resulting from~~ AI Systems and whether they are identified, tracked, and addressed **through existing complaint and remediation frameworks.***

Exhibit C: High-Risk AI Models

Recommendation #1:

Revise the Exhibit C Company Instructions to direct companies to complete Exhibit C only for the high-risk AI models identified.

Rationale:

Exhibit C is titled "High-Risk AI Models," and its stated purpose is to obtain detailed information on in-production high-risk models. The Company Instructions direct companies to "fill in the details for each of the AI model(s) requested" and to include all companies and lines of business, which could be read to extend the exhibit's detailed, model-by-model reporting to models beyond those that are high-risk. During the pilot, some member Plans were asked to complete Exhibit C for all AI use cases identified in Exhibit A, including models the company had already assessed and determined to be low risk. Because Exhibit C calls for substantial detail on each model, extending it to lower-risk models in this way creates compliance and resource demands that are not matched by a corresponding oversight benefit. Revising the Company Instructions to refer to "each high-risk AI model identified" would align the instructions with the exhibit's stated purpose and help ensure this detailed reporting remains focused on the models that present the greatest risk.

Recommendation #2:

Clarify the Exhibit C instruction addressing when information should be provided in relation to the AI System versus the AI Model.

Rationale:

The Company Instructions note that there may be instances where a company feels information should be provided in relation to the AI System and not the AI Model, or the reverse. As written, this instruction is difficult to apply consistently, and companies would benefit from additional guidance and examples describing when a response should address the system rather than the model. Clearer guidance would reduce the likelihood of misaligned responses and the follow-up inquiries that such misalignment tends to prompt.

Exhibit D: AI Model Data

Recommendation:

Make disclosure of specific third-party data sources and vendor names in Exhibit D optional or accommodate contractual confidentiality restrictions.

Rationale:

Requiring companies to disclose the identity of third-party data sources and vendor names may conflict with existing confidentiality agreements and nondisclosure obligations, as many vendor contracts restrict disclosure of the vendor's identity or solutions in regulatory filings outside of privileged examination contexts. Companies cannot readily resolve this by renegotiating their agreements. Confidentiality of a vendor's identity and solution is often a core commercial term, and companies have no unilateral right to change it; vendors with significant market power may simply decline and subsequently limit innovation that is in the best interest of members and patients. A mandatory disclosure requirement could therefore narrow the pool of vendors willing to be named. Making vendor identification optional, or providing an express accommodation for contractual restrictions, would avoid placing companies in the position of choosing between completing the exhibit and honoring their contractual obligations, while still allowing regulators to obtain the information through appropriate channels where needed.

Definitions

Recommendation:

Align the definitions of "Inherent Risk" and "Residual Risk" so that both refer to risk-mitigation strategies and internal controls in parallel terms.

Rationale:

As drafted, "Inherent Risk" refers to risk before considering internal controls, while "Residual Risk" refers to risk after considering risk-mitigation strategies or controls. The two definitions are meant to be counterparts, and conforming them so they mirror one another would improve internal consistency and prevent unintended distinctions in how the two concepts are applied. BCBSA suggests the following edits (additions in **bold**):

*"Inherent Risk" refers to the risk of economic loss or inaccurate financial reporting before considering **risk-mitigation strategies or** internal controls.*

*"Residual Risk" refers to an assessment of risk after considering risk-mitigation strategies or **internal** controls.*

From ByteStream Strategies Inc.

29 September 2026

Scott Sobel, AI Policy Advisor
Miguel Romero, Director of Innovation, Cybersecurity and Technology
National Association of Insurance Commissioners

Re: Comment on the AI Risk Evaluation Supplement, version 5.0

Scott and Miguel:

ByteStream Strategies Inc. advises organizations on AI governance programs, with particular attention to whether governance evidence supports the conclusions drawn from it. We have read version 5.0 and offer one comment, with two shorter notes attached to it.

The comment concerns an ambiguity in how far the materiality threshold reaches across Exhibit A.

Exhibit A, Part One asks for the "Number of AI Models with Direct Consumer Impact" and, in the column beside it, the "Number of AI Models with Material Financial Impact". The Regulator Instructions to that Exhibit provide that regulators "should either specify a materiality threshold to guide company responses to the Supplement or request that the company provide a threshold that it used in completing this Exhibit".

The two columns are built differently. "Materiality" is defined as "the dollar amount above which the examiner's perspective of an insurer's financial position will be influenced", and "Material Financial Impact" is expressly determined against that threshold. "Direct Consumer Impact" is defined by what the model decides: underwriting, tier and coverage eligibility, premium determinations, ratemaking, and claims denial and settlement. Several of those involve money, and sometimes a great deal of it. What the definition does not do is make inclusion conditional on financial materiality to the insurer. The test is what the decision is, not what it costs the carrier.

Both readings are available on the text as drafted. A company could take "responses to the Supplement" at its word and apply the financial threshold to both columns, or read the two definitions as keeping the consumer column outside its reach. The definitions support the second reading and we think it is the intended one, but the instruction does not say so.

The first reading omits models the definition includes. A claims denial model at a smaller carrier can touch every claimant it sees and never reach a dollar amount that influences an examiner's view of the financial position. A count built that way is defensible against the instruction and short against the definition, and a reviewer comparing the two columns cannot see which reading produced them.

We recommend one sentence in the Exhibit A Regulator Instructions: that the financial materiality threshold determines inclusion in the Material Financial Impact column, and does not exclude otherwise qualifying models from the Direct Consumer Impact column. We would confine it to Exhibit A. Exhibit C already provides that materiality is a consideration in determining the level of risk, alongside Adverse Consumer Outcomes and material consumer

impact, and a clarification written for Exhibit A should not be read as making financial materiality control the consumer-risk assessment there.

Two shorter notes.

First, Exhibit A provides that "If a company has an existing AI Systems inventory, it may suggest submission of the inventory in lieu of completing Exhibit A". A company inventory is built to that company's own scope rules, which need not match the Exhibit's. Where a substitution is accepted, we suggest the company be asked to state the scope rule its inventory uses, so that a regulator can see whether the substitute covers the same ground.

Second, the Supplement captures the materiality threshold in three places: the header field in Exhibit A, the disclosure obligation within the definition of Material Financial Impact, and checklist item 3o. We take no issue with any of them. We would add one question on subsequent responses: whether the threshold has changed since the last response and, holding the model population constant, how many models entered or left scope as a result. That isolates the effect of the threshold from model retirements, new deployments and shifts in business use, which would otherwise sit in the same number.

None of this asks the Working Group to set a numeric standard. Version 5.0 leaves that with the regulator and the company, which we think is right. It asks only that the financial materiality threshold used under the Supplement be pointed at the column it can actually measure.

We are glad to expand on any of this.



Chair Nathan Houdek (WI)
Co-Vice Chair Mary Block (VT)
Co-Vice Chair Doug Ommen (IA)
2025 NAIC Big Data and AI (H) Working Group NAIC Central Office
1100 Walnut Street
Suite 1500
Kansas City, Missouri 64106

Sent via email to: ssobel@naic.org

RE: Comments on the AI Risk Evaluation Supplement (Version 5.0)

Dear Chair Houdek and Co-Vice Chairs Ommen and Block:

We are writing on behalf of the Committee of Annuity Insurers¹ ("CAI" or "Committee") to provide comments on Version 5 of the Artificial Intelligence (AI) Risk Evaluation Supplement (the "Supplement") that was exposed for comment by the Big Data and Artificial Intelligence (H) Working Group ("BDAIWG" or "Working Group") on August 29, 2026. We appreciate the Working Group's dedication and commitment to developing the Supplement, and its consistent efforts to engage key stakeholders throughout the drafting and pilot process.

The CAI's comments address both procedural and substantive topics that we hope the Working Group will take into consideration as it works to finalize the Supplement and the NAIC moves towards its adoption. Our overarching request is that the final Supplement be revised so that it is flexible enough that it will enable regulators to focus on those AI developments and deployments that pose the greatest risks for consumers and the industry as powerful new AI Systems ("AIS") evolve with unique capabilities and risks.

¹ The Committee of Annuity Insurers is a coalition of life insurance companies that issue annuities. It was formed in 1981 to address legislative and regulatory issues relevant to the annuity industry and to participate in the development of public policy with respect to securities, state regulatory and tax issues affecting annuities. The CAI's current 33 member companies represent approximately 80% of the annuity business in the United States. More information is available at <https://www.annuity-insurers.org/>.

COMMENTS

1. Clarify whether, and if so how, the Supplement will be integrated with the Market Regulation, Financial Condition Examiners and Financial Analysis Handbooks (“Handbooks”)

The instructions contained in the Supplement state that the Supplement “will be coordinated consistent with the guidance provided by the *Market Regulation Handbook*, *Financial Condition Examiners Handbook*, and the *Financial Analysis Handbook*” and go on to explain that “many important decisions leading to the full use of this Supplement would be governed by the aforementioned Handbooks.” The BDAIWG has mentioned in previous meetings that efforts are underway to coordinate with other relevant NAIC committees regarding the use and implementation of the Supplement. However, while the BDAIWG has taken the lead in drafting the Supplement, it is unclear the extent to which other NAIC committees that have mandates over the regulatory exam processes will adopt the Supplement into the Handbooks. The CAI requests that the BDAIWG clarify what specific plans, if any, are underway at the NAIC to integrate the Supplement into the Handbooks prior to, or after, adoption of the Supplement by the NAIC.

2. Explain how regulators will evaluate and use the responses provided to the Supplement in the future.

The pilot initiated by the BDAIWG has provided regulators participating in the pilot with valuable opportunities to coordinate and communicate about the experiences they have had administering the Supplement and the responses they have received from insurers. This has resulted in a process that we understand has promoted more uniformity in interpretations of various aspects of the Supplement among the states and efficiency in the exam process. This coordinated approach among the states is particularly important now when the capabilities and risks of AIS and AI models are evolving so rapidly. The CAI requests that the BDAIWG consider further bolstering regulatory consistency by providing additional written guidance on lessons learned from how the Supplement has been interpreted and used in the pilot program. For instance, a report on the Pilot program that includes anonymized information on the results of testing of the various Exhibits would be valuable for both regulators and the industry. That report could also detail the extent to which each Exhibit was tested during the pilot and provide recommendations as to how regulators can best tailor each Exhibit, as well as how and when each Exhibit should be used during an exam.

3. Reconsider the quantitative emphasis of Exhibit A.

As the NAIC Model AI Bulletin states, insurers are expected to readily produce an inventory of their material AIS upon regulatory request. Currently, Exhibit A in the Supplement requests an inventory of AIS and AI models in a complex numerical fashion. We believe that Exhibit A – Part Two, with its narrative focus, is a less burdensome and more effective way for regulators to obtain a complete picture of those AIS that are used by an insurer. The quantitative approach currently presented in Exhibit A – Part One will be administratively burdensome, will be difficult for insurers to produce accurately, and will only provide regulators with a rudimentary “heat map” of where insurers are

concentrating their AI deployments. The Committee requests that Exhibit A Parts One and Two be streamlined into a single principles-based request for insurers to disclose material AIS, that any quantitative elements be simplified, and that insurers be allowed to explain in a clear narrative the scope of their AIS deployments.

4. Ensure conformity between the different formats of Exhibit B

Exhibit B is intended to provide regulators with flexibility as to how they choose to request information about an insurer's current risk management posture and procedures (what the Model Bulletin refers to as an "AIS Program"). However, the narrative and checklist formats diverge not just in format, but also in the actual substance of the information being requested. The CAI requests that the content of the two Exhibit B formats be revised to be consistent, so as to encourage uniformity and efficiency in the responses regulators obtain to Exhibit B.

The following examples illustrate instances in which the Narrative and Checklist formats request substantively different information:

- **Assessment of Autonomy and Reversibility.** The Narrative format (Question 1g) asks how the company assesses autonomy, reversibility, and reporting impact risk — a critical inquiry given the rise of agentic AI. The Checklist format contains no corresponding question.
- **Data Privacy and Consumer Data Protection.** The Checklist format (Question 3d) asks whether the AIS Program considers data privacy and protection of consumer data; the Narrative format does not include a corresponding inquiry.
- **Employee Training and Prohibited AI Practices.** The Checklist format (Question 3j) asks whether the company trains employees about AI System use and defines prohibited practices; the Narrative format includes no analogous question.

These discrepancies could produce materially different regulatory assessments of the same company's AI governance depending solely on which format the regulator selects, undermining the uniformity and efficiency the Supplement is designed to promote.

5. Enhance the Supplement's focus on forward-looking AI risks, such as Agentic AI

The CAI is strongly opposed to the new emphasis placed on including Generalized Linear Models ("GLMs") within scope of the Supplement. While GLMs can present risks and impact consumer outcomes, this technology is tried, trusted and transparent. The governance and risk management necessary to oversee GLMs has been long established, is dramatically different from that required for Generative AIS, and is fundamentally misaligned with the governance and risk management tools required for Agentic AI with its focus on permissions, perimeter boundaries and agent identities, as well as third party uses. The Supplement, like the Model Bulletin, will remain most relevant the more it follows a risk- and principles-based approach. By doubling down on the discussion and promotion of regulating GLMs under the Supplement, the Supplement risks failing to provide the instruction and guidance that is necessary for industry and regulators to effectively and efficiently oversee the significant risks presented by both generative and

agentic AIS and other frontier AI deployments. The Committee requests that the emphasis on GLMs be removed from the Supplement and that the Working Group consider specific revisions to the Supplement to ensure that it systematically requests information needed for regulators to assess how insurers are governing and mitigating the risks arising from their own deployments or from third party deployments of Agentic AI.

Thank you for the opportunity to comment on the exposed Version 5 of the AI Systems Risk Evaluation Supplement. We hope that you find these comments helpful at this stage in your strategic thinking. Please do not hesitate to contact us if you have any questions and/or would like to discuss our comments further.

Sincerely,

For The Committee of Annuity Insurers

Eversheds Sutherland (US) LLP

By: Mary Jane Wilson-Bilik
Mary Jane Wilson-Bilik
Partner

Cc: Stephen E. Roth, Eversheds Sutherland (US) LLP
Jeremy Bloomstone, Eversheds Sutherland (US) LLP

Comment on the AI Risk Evaluation Supplement, Version 5.0

Exhibit B Is the Oversight Exhibit

September 26, 2026

Via email to ssobel@naic.org and maromero@naic.org

Commissioner Nathan Houdek, Chair Commissioner Doug Ommen, Co-Vice Chair Mary Block, Co-Vice Chair Big Data and Artificial Intelligence (H) Working Group National Association of Insurance Commissioners

Re: AI Risk Evaluation Supplement, Version 5.0 (exposed August 31, 2026)

Chair Houdek, Co-Vice Chairs Ommen and Block:

The Center for AI Oversight submits these comments on version 5.0 of the AI Risk Evaluation Supplement. A one-page summary follows, then the comment itself, and proposed additions to Exhibit B are attached separately in the exhibit's own format with an editable version for staff use. We would welcome a few minutes on the October 8 call to present the recommendations, and are glad to speak with staff beforehand.

Respectfully submitted,

Brian J. Allen, Esq. Executive Director, Center for AI Oversight info@cfaio.org

Summary

The finding. The Supplement is already built in three layers. Exhibits A, C and D ask what the insurer built, bought and tested. Exhibit B asks who chartered the program and who is accountable for it. Exhibit B is the oversight exhibit. It is distinct in kind from the other three rather than a higher-altitude version of them, and it is the only part of the instrument that will still be true at the next examination, because oversight decisions attach to the company's decision process rather than to the models, use cases and risks that change continuously beneath them.

The recommendation. Complete Exhibit B. Five recommendations, expressed as one amendment to item 1 and seven new Checklist entries, plus one optional item. All are answerable from a document, none requires an examiner to evaluate a model, and none prescribes governance content.

#	Addition	Exhibit	What it fixes
1	What the program governs, its granularity, who approved that as appropriate, and the	B, item 1	A program scoped by decision and one scoped by default answer identically

#	Addition	Exhibit	What it fixes
	standards elected		
2	Accountability at two levels, declared exclusions, and competency	B, 2 series	"Board of Directors or management" can be answered accurately and tell the examiner nothing
3	Ratified AI risk tolerance, with body and dates	B, 3 series	A tolerance that was calculated but never ratified is a working assumption
4	Authority to operate, dated, scoped and expiring	B, 3 series	Automating a decision is a grant of authority, not a design outcome
5	The escalation and disclosure line, with a destination and a record	B, 3 series	Nothing establishes what happens when a system runs outside tolerance

Plus four fields in Exhibit C recording who approved the company's risk criteria, and one change separating what a vendor's officer can attest to from what requires assurance.

The Supplement is already built in three layers

OVERSIGHT GOVERNANCE	What is governed, at what depth, and who answers. Chartered once, then operated.	Still true at the next examination
Exhibit B <i>Five additions proposed</i>		
MANAGEMENT	What the company does: model inventory, use cases, risk criteria, vendor sourcing.	Accurate on the day collected
Exhibits A and C		
TECHNICAL CONTROLS	How the systems are built and tested: testing methodology, data element categories.	Accurate on the day collected
Exhibits C and D		

Exhibits A, C and D ask what the insurer built, bought and tested. Exhibit B asks who chartered the program and who answers for it.

Exhibit B is the oversight exhibit, and it is the only part of the instrument that does not decay between examinations.

Why it matters. This instrument is used inside examinations that run on statutory cycles of up to five years, and in Florida market conduct examinations run "as often as it deems necessary." Exhibit A's model inventory is accurate on the day it is collected and materially stale by the next cycle. The decisions Exhibit B records are stable by nature. When they do change, the program carries the record of the change.

The program is the examinable object. A committee is a governing body. What it governs is a program, and the program does not govern the technology. It governs the company's capacity to make informed decisions about its use of AI. An examiner cannot examine a committee's diligence and should not be asked to examine a model. An examiner can examine a program, with the questions examiners have asked of compliance and anti-fraud programs for decades.

None of it is a new disclosure. Where AI is mission critical, the duty to maintain a board-level monitoring and reporting system already attaches under the Caremark line, applied to mission-critical risk in *Marchand v. Barnhill*. A company that has met it can answer these fields from documents it already keeps.

The mechanism exists, and extending it means naming the object. Insurance regulation names accountable officers twice: the CGAD signature under Model Act #305 Section 3.B and the ORSA Summary Report signature under Part C of the Guidance Manual's Introduction. Neither reaches AI, and neither states what the officer is accountable for with enough precision to examine. Extending them is not a matter of adding a third signature. It is a matter of naming the object: the AI systems program, and what that program governs.

The precedent is recent, and it is cyber. The 2014 Cybersecurity Framework had five functions and no GOVERN. The oversight layer went unbuilt for a decade, was then closed from outside the standards community, and is still not fully closed. This Working Group can put the layer in its own instrument now, on its own terms.

If only two additions are taken, take 2 and 5. With any system that automates a decision, add 4.

What we do not comment on. Exhibit A, Exhibit C's testing methodology fields, and Exhibit D. Those concern how an insurer builds, tests and sources its systems. They belong to the operator and the examiner.

1. Scope

The Center for AI Oversight is an independent nonprofit research and education organization. It defines oversight governance for AI in regulated industries: what must be governed and who is accountable for it. It does not prescribe how AI systems are built, tested, validated or run.

This comment observes that boundary. It addresses Exhibit B, two provisions in Exhibit C, and the attestation in the companion *Regulatory Framework for Third-Party Data and Model Vendors*. The Center takes no positions on matters involving specific institutions.

2. Exhibit B is the oversight exhibit

Three layers, and why the distinction is not academic

Technical controls govern how the system is built, constrained and watched. **Management** governs how the organization pursues its AI objectives and manages the risk of doing so; ISO/IEC 42001 is written for this layer, with the NIST AI Risk Management Framework supplying the risk practice inside it. **Oversight governance** sits above both. It decides what is governed, at what depth, and who answers, and it never decides how.

Management here names an activity, not a rank. Setting objectives, prioritizing, resourcing and treating risk are management wherever performed, including at the top of the company. What makes a decision oversight is not the seniority of the person making it, but that it settles what is governed and who answers for it.

The Supplement already reflects this

Exhibits A, C and D sit at the management and technical layers. Exhibit B sits at the oversight layer, in both of the options version 5.0 offers. The Narrative asks the company to provide its AIS Program and to discuss the governance structure, Board reporting and frequency, how responsibility for governance is assigned, and the integration of AI Systems into the ORSA and ERM assessments. The Checklist asks whether a written AIS Program was adopted, whether the Board of Directors or management was involved in adopting it and what their role in it is, and then runs the 3 series across ERM at 3f, ORSA at 3g, vendor procurement at 3l, and vendor-developed systems at the new 3p. Every one of those is an oversight question.

Version 5.0 moved further in the same direction. The Checklist now carries a Document Name and Page # column, and the instructions direct that "if the responses to the checklist indicate that an explanation of a policy, procedure or process can be found on a given page of a document, indicate the page number and the document title and include the document referenced in the submission." That is the working principle behind every field proposed here: an oversight question should be answerable from a document the company either holds or does not, and the absence of the document is itself the finding.

This comment does not ask the Working Group to adopt an outside framework. It asks the Working Group to finish the exhibit it already wrote.

What the exhibit is examining is a program

Most companies respond to AI by forming a committee. That is not wrong, but it is not governance. A committee is a governing body: it is the WHO. What it governs is a program: the WHAT. And the program does not govern the technology. It governs the company's capacity to make informed decisions about the balance of risk and reward in its use of AI.

That is what makes the subject supervisable. An examiner cannot meaningfully examine a committee's diligence, and should not be asked to examine the technology. An examiner can examine a program. Does it exist. Was it chartered. Is it scoped. Is it granular in proportion to the environment. Did the accountable body approve it. Those are the questions market conduct and financial examiners have asked of compliance programs, anti-fraud programs and risk management programs for decades, and they require no new authority to ask.

A committee that meets without a program beneath it is evidence of attention, not of accountability.

One consequence for the drafting

Version 5.0 defines an AIS Program as "the controls and processes that an insurer adopts and implements to ensure the responsible use of AI Systems," and says such a program "may be comprised of general guidelines, governance framework, risk management and internal controls, and third-party oversight."

That is a management definition. Controls and processes are what an organization does. The definition places governance inside the program as one component among several rather than above it as the layer that charters it, and it is the reason Exhibit B cannot presently distinguish a company that has built an oversight program from one that has built controls and called them governance.

The Center does not propose that the Working Group redefine the term. The narrower fix is to ask the company. A program whose stated object is the company's AI systems is a management program, and Exhibits A, C and D already reach that ground. A program whose stated object is the company's capacity to

decide about those systems is an oversight program, and it is the only one Exhibit B can test without duplicating the other exhibits. Recommendation 6.1 records it in the opening item, in one clause.

The Working Group's own data supports the emphasis

In the NAIC health insurance AI/ML survey released in May 2025, 93 companies responded and 84 percent said they apply AI or machine learning to health insurance operations, which is 78 companies. Among those 78, 55 percent develop internally using third-party components and 15 percent rely entirely on third-party vendors, with a further 13 percent using a combination. On the two clearest categories alone, seven in ten AI-using health insurers run systems built wholly or partly by someone else. For those companies the model is not theirs to explain. The decision to deploy it, under whose authority and with what withdrawal criteria, is.

3. The answers already exist in a well-governed company

The duty attaches before the exhibit asks

Nothing proposed here is a new disclosure. Where AI is mission critical to underwriting, pricing or claims handling, the obligation to maintain a board-level system of monitoring and reporting already attaches. The Delaware courts developed that duty from *In re Caremark Int'l Inc. Derivative Litig.*, 698 A.2d 959 (Del. Ch. 1996), and applied it to mission-critical risk in *Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019), where the claim survived dismissal because the board had no committee charged with the company's central safety risk, no protocol requiring management to report on it, and no record of having discussed it.

A company that has met that duty can answer every field proposed in this comment from documents it already keeps. A company that cannot has that gap whether or not this exhibit asks.

The record is protective as well as exposing, and this sector holds the exposure twice

In *Firemen's Ret. Sys. of St. Louis v. Sorenson*, C.A. No. 2019-0965-LWW (Del. Ch. Oct. 5, 2021), a documented and monitored program earned dismissal of the oversight claim after a breach of roughly 500 million guest records. The artifacts that answer an examiner's question are the same artifacts that defend the board.

Insurers in scope of this Supplement hold that exposure from both sides. They deploy AI in their own underwriting, pricing and claims operations. Many of them also write directors and officers liability, where the oversight failures of their insureds arrive as claims on the same theory that has already reached cybersecurity.

That second exposure is not theoretical. Cornerstone Research counted sixteen AI-related securities class actions in 2025 and fifteen in the first half of 2026 alone, on pace to nearly double the 2025 total. Those fifteen were thirteen percent of core filings and seventy-three percent of the Disclosure Dollar Loss index, and five of the seventeen mega filings in the period were AI-related. A smaller and growing set of derivative suits pleads board oversight failure directly.

No court has yet decided an AI oversight claim on the merits, and the Delaware pleading bar remains high. That distinction matters less to a D&O writer than it might appear. Defense costs are incurred from the complaint forward whatever the outcome, and settlements in this line have been reached on governance reforms without any merits ruling at all. The loss experience is already running.

This Supplement is written for financial analysis and financial examination as well as market conduct, and that is where the two sides meet. Exhibit B asks the insurer about its own AIS Program. The D&O book is an underwriting, pricing and reserving question this instrument does not reach. But the document set is the same one on both sides, and the examiner reading Exhibit B is reading the record the insurer would need to produce as a defendant and would want to see from a D&O applicant.

This comment does not ask the Working Group to adopt, interpret or enforce any corporate law standard. Model Act #305 Section 1.B reflects the same limit in the corporate governance context, providing that nothing in that act shall be construed to prescribe corporate governance standards beyond applicable state corporate law. The point is practical: these fields impose a collection cost, not a governance cost.

4. Oversight is the stable half, and cyber shows the cost of leaving it out

The lower layers move at the speed of the technology

A model deployed this quarter behaves differently next quarter. Use cases are added and retired, and risks are reclassified as the environment changes. The oversight layer does not move at that speed, because the obligation attaches to the company's decision process rather than to any generation of the technology. That matters here because of the cycle this instrument runs on.

Among the pilot states, Virginia, Pennsylvania and California each examine licensed insurers at least once every five years (Va. Code Section 38.2-1317(B); 40 P.S. Section 323.3(b); Cal. Ins. Code Section 730(b)). Florida conducts financial examinations on a risk-tiered basis of three to five years (Fla. Stat. Section 624.316) and market conduct examinations "as often as it deems necessary" (Fla. Stat. Section 624.3161).

Set that against Exhibit A, which asks for the number of AI Systems currently in use, the number implemented in a regulator-specified lookback window, and counts of AI Models with direct consumer impact and with material financial impact. Those figures are accurate on the day they are collected. Three years later a substantial share of the inventory will have been retrained, replaced, retired or reclassified. An examiner reading the prior examination's Exhibit A is reading a photograph of a moving object.

Exhibit B does not decay that way. A charter, a ratified tolerance, a named accountable officer and an escalation line change only when the company decides to change them, and the program carries the record of that change.

Two regulators have already acted on this

Connecticut, a pilot state, requires its domestic insurers to complete an annual Artificial Intelligence Certification under Insurance Department Bulletin No. MC-25 of February 26, 2024, due September 1, 2024 and annually thereafter. The bulletin rests entirely on pre-existing authority, principally the Unfair Insurance Practices Act (Conn. Gen. Stat. Sections 38a-815 to 38a-819), market conduct examination authority (Sections 38a-15 and 38a-16), the corporate governance annual disclosure statute (Section 38a-142a) and the rating laws. A recurring AI oversight filing, with no AI statute behind it. NYDFS Insurance Circular Letter No. 7 (2024), at paragraph 26, states that an insurer's board or other governing body, its committees, or senior

management under delegated authority "should review and approve" the insurer's AI-related policies and procedures at least annually. It is guidance rather than a rule, and it places the annual approval at the governing level.

Why existing practice does not supply the layer

Governance, risk and compliance practice produces control inventories, risk registers, policy sets and testing programs, each with an owner. It rarely states which layer a responsibility belongs to, or who is answerable at the top of each. The layers collapse into one undifferentiated function, and a company can then describe in detail how its AI risk is managed without being able to say who is accountable for a given deployment. An examiner reading that record sees activity everywhere and answerability nowhere.

The management instruments do not close it, and they are right not to try. They leave scope, granularity and reporting depth to the organization: testing proportionate to consequence, review calibrated to criticality and risk tolerance. Each is a calibration they are correct to leave to the operator. None says who holds it, or on what record. A calibration nobody ratified is a working assumption, and an examiner cannot tell the two apart.

That is the shape of the gap. It is also the shape cybersecurity left open for a decade.

The last time this layer went unbuilt, the gap took a decade to close

This is not the first control discipline to mature from the bottom up. The 2014 Cybersecurity Framework organized its core around five functions, Identify, Protect, Detect, Respond and Recover, and contained no GOVERN function. The control layers matured first and the oversight layer went largely unbuilt for a decade.

It was filled eventually, and mostly from outside the framework. The Securities and Exchange Commission adopted cybersecurity disclosure rules on July 26, 2023, requiring registrants to describe "the board of directors' oversight of risks from cybersecurity threats" under Item 106 of Regulation S-K. The courts applied the oversight duty to cybersecurity on their own schedule. NIST then added GOVERN as a sixth function, including an Oversight category, in Cybersecurity Framework 2.0 on February 26, 2024.

The sequence is the point. For ten years the artifact everyone reached for had nothing to say about who answers, and the answer arrived from a disclosure rule and a courtroom before the standard caught up.

The point for this Working Group is narrow. The oversight layer arrives eventually, either inside the examination instrument on terms the regulator sets, or afterward through a disclosure rule or a courtroom on terms set elsewhere. The Working Group holds an advantage NIST did not: it is an examination body rather than a standards body, and it can simply ask the question. This exposure is the moment to ask it.

5. The naming mechanism already exists, and extending it means naming the object

Insurance regulation already requires a named individual to attest, annually, on a record filed with the department. It does so twice, and neither instrument reaches AI.

Corporate Governance Annual Disclosure

Model Act #305 Section 3.B requires "a signature of the insurer or insurance group's chief executive officer or corporate secretary attesting to the best of that individual's belief and knowledge that the insurer has implemented the corporate governance practices."

The signer is named. The practices are not defined, and they are not broken out by layer. An insurer can satisfy that attestation by describing management processes and calling the result governance, and an examiner reading a CGAD filing cannot tell whether AI oversight decisions are covered by it, covered in part, or not addressed at all.

That is not a defect this Working Group is charged with repairing, and the Center does not propose that it try. It is the reason Exhibit B should ask the question directly rather than rely on a company's corporate governance filing to have answered it. One note for whichever NAIC group holds that charge. Defining the attested practices and breaking them out by layer expectation, so that an attestation distinguishes the decisions made at the oversight layer from the processes operated beneath it, would make both instruments more examinable.

Own Risk and Solvency Assessment

The Introduction to the ORSA Guidance Manual, at Part C (General Guidance), requires "a signature of the insurer's chief risk officer or other executive having responsibility for the oversight of the insurer's ERM process attesting to the best of his/her belief and knowledge that the insurer applies the ERM process described in the ORSA Summary Report and that a copy of the ORSA Summary Report has been provided to the insurer's board of directors or the appropriate committee."

The attestation stops at confirming that the report was *provided* to the board. Provision is receipt. It does not establish what the board did with it, and that is where AI oversight most often fails. Recommendation 6.5 closes that half step by asking for the recorded response.

What extending the mechanism actually requires

The Working Group does not need to invent an accountability mechanism for AI. The sector has one, in two forms, filed by every domestic insurer. But extending it is not a matter of adding a third signature to a third form, because an attestation is only as examinable as its object. CGAD leaves the practices undefined. ORSA defines the object as the ERM process, which is a management process. Neither states that anyone is accountable for an AI oversight program, or what such a program governs.

Extending the mechanism therefore means three things, and Exhibit B can do all three: name the committee that holds the program by charter and the officer who answers to it; state what that officer is accountable for, which is the AIS Program; and state what the program governs, which is the company's capacity to make informed decisions about its use of AI, as distinct from the AI systems it operates. A signature with no stated object attests to nothing in particular.

6. Five additions to Exhibit B

Each is a question about the program. None requires the examiner to evaluate a model. None imposes a substantive governance standard.

6.1 Ask what the program governs, at what granularity, who approved that as appropriate, and which standards the company elected

Checklist item 1 asks whether a written AIS Program has been adopted, when, and how often it is reviewed for updating. Narrative item 1 asks the company to provide the program. Neither asks what the program governs, at what granularity, whether anyone decided that scope and granularity were right for this company, which external standards were elected, at what depth, or who decided.

Two gaps sit inside that, and the first is the more consequential.

Scope and granularity are decisions, and the exhibit treats them as givens. A rural domestic carrier and a multinational group perform the same oversight acts at very different depth. That is correct, and it is why a national instrument should not set one depth for both. But a program that covers the right ground because someone decided it should and a program that covers it because nobody looked further produce identical answers to item 1 as written. Reviewing a document for updating is not the same as testing whether its scope is still appropriate to a company whose use of AI has changed since the last examination. This is the field that distinguishes a chartered program from an inherited one.

An elected standard is a commitment. Adopting a voluntary framework at a chosen depth creates an obligation the company set for itself, and it is the commitment, not the framework, that assurance later tests. The exhibit does not record what was elected or how far.

Proposed: state what the AIS Program governs and the granularity at which it operates; identify the individual or body that approved that scope and granularity as appropriate to the company's size, complexity and use of AI Systems, with the date the scope and granularity are next reviewed for continued appropriateness; and identify the external frameworks or standards elected, the depth of application elected, and the individual or body that approved the election, with date.

A program that states it governs the company's AI systems has described a management program. One that states it governs the company's capacity to decide about those systems has described an oversight program. The exhibit cannot currently tell which it received.

6.2 Ask who is accountable for the program, at two levels

Four questions come close and none establishes accountability for the program. Narrative item 1a asks what role maintains the AIS Program. Narrative item 1e asks how responsibility for governance is assigned and how the organization ensures consistency. Checklist item 2 asks whether the Board of Directors or management was involved in adopting the program, and item 2a asks what the role of the Board of Directors or management in it is.

A company can answer all four accurately and still not have said who is accountable for the AIS Program itself. Maintaining a program is an administrative function. Assigning responsibility describes an organizational chart. Involvement in adoption is a historical fact. The disjunctive running through items 2 and 2a permits a response that is true and uninformative.

The object is what matters. Accountability for AI in general is a portfolio description. Accountability for the AIS Program is a position that can be examined, because the program has a stated scope, a stated granularity and required outputs, and someone either answers for them or does not.

That accountability sits at two levels, and the two are not interchangeable. A company that names its Audit Committee as holding oversight of the program by charter has answered the first level, and for many insurers that is the right body. Adding the committee chair's name does not answer the second, because the chair sits inside that body rather than beneath it. The question that settles it is short: when the committee asks whether the program did what it was chartered to do, who answers? If the answer is the committee, the exhibit has recorded a self-report. If a named officer answers to a named committee, the examiner can test both sides.

Nothing in this requires a new role or a new committee. The officer is whoever already owns it, and insurance regulation already names accountable officers twice, as section 5 sets out.

Proposed: identify the committee holding oversight of the AIS Program by charter; identify by name and role the officer accountable for the program, with any delegation and its date; identify any AI systems declared out of scope, the named individual who accepted each exclusion, and the review date; and describe how the company establishes that the accountable officer and the committee hold the competency to interrogate what the program reports to them.

The out-of-scope field is the one most likely to produce a finding. Systems fall outside a program because they are low risk, legacy or acquired with a book of business, and sometimes because no one decided. The first three are decisions, the fourth is not, and the exhibit cannot currently distinguish them.

6.3 Ask who ratified the risk tolerance and when it is next reviewed

Checklist item 3k asks how the company "Quantifies AI System risk levels," and the new item 3o asks how it "Determined 'materiality' and what threshold was used for assessing AI Systems risk." Both ask how the assessment was made. Neither asks who ratified it. Assessment is an act performed by a modeling or risk function. Ratification is an act performed by a body that can be held to it, and a tolerance calculated but never ratified is a working assumption the company is not committed to.

Proposed: identify the stated AI risk tolerance, the body that ratified it, the date of ratification, and the date of next scheduled review.

6.4 Ask whether the program treats removing the human as a grant of authority

Narrative item 1g asks how the company "assess[es] autonomy, reversibility, and reporting impact risk of AI Systems." Assessing the risk of autonomy is not the same as authorizing it, and a company that has assessed the risk well may still be unable to say who granted the system permission to act.

The oversight question is not what any one system was permitted to do. It is whether the program treats the removal of a person from a decision as a grant of authority at all, or as something that happens when a model is deployed. Those produce different records and different examinations.

Proposed, as an Exhibit B item: identify how the company grants, bounds and periodically re-ratifies authority to operate for AI Systems recorded as automating a decision, recording the authorizing individual or body, the date of authorization, its scope, and its expiration or re-ratification date.

Section 7.2 proposes the per-model complement in Exhibit C, where the automate, augment and support classification is already recorded. That field is the record the program produces. This one is the program.

6.5 Ask where an out-of-tolerance system escalates to, and what was done about it

Narrative item 1b comes closest, asking the company to discuss "the governance structure, Board reporting and frequency." That establishes that reports move and how often. Checklist items 3m and 3n are consumer-facing, addressing complaints and consumer awareness.

None of them establishes what the report carries, what triggers one outside the cadence, who can act on it, or what was done. That gap matters most for the system operating within specification and outside tolerance, which is the harder case and the one that produces supervisory surprise.

Proposed, as two items rather than one, because eight elements in a single checklist cell will not be answered well.

Escalation: the cadence on which the program reports to the accountable committee independent of events; whether that reporting carries adverse information, incidents, findings, inquiries and deviations from tolerance and not only progress; the escalation triggers and materiality thresholds fixed in advance; the party with authority to continue, constrain or withdraw a system; whether at least one reporting pathway to the governing body does not depend on management's discretion; and whether the response to an escalated matter is recorded.

Disclosure: the company's disclosure position, including the criteria for disclosure to the department, to consumers and to the public, and any clock that statute attaches to it.

Two elements carry most of the weight. A routing label without a party who can withdraw the system does not resolve to an action, and a red flag received without a recorded response is, on the record, a red flag ignored. This is the field an examiner can test directly from minutes.

Optional sixth: ask whether anyone independent has tested the program

Narrative item 1d asks how "the effectiveness of the AIS Program and individual models are assessed and modified," and Narrative item 3a asks whether validation and testing of internally developed systems are "performed by someone that is independent from development." The Working Group has therefore already applied the independence principle, but to model validation rather than to the program.

Proposed: identify the provider of any independent assurance over the AIS Program, the scope of the review and the date of the most recent one. The useful object of such a review is whether the company applied what it said it would apply, at the depth it recorded in the opening item, rather than conformance to a standard in the abstract. ISO/IEC 42006:2025, published July 7, 2025, gives this market a published accreditation standard. A blank answer is information, not a deficiency.

Proportionality

Recommendations 6.2 and 6.5 carry most of the value for a small or resource-constrained insurer, with 6.4 added wherever a system automates a decision. That minimum viable form is why an oversight instrument scales from a small domestic carrier to a global group without a separate small-entity version.

7. Two provisions in Exhibit C

7.1 Keep company-set risk criteria, and record who approved them

Exhibit C states that "AI model risk criteria are set by the insurance company." The Center supports that provision and asks the Working Group not to retreat from it. A national instrument cannot set one threshold right for both a small domestic carrier and a multinational group. Allowing the company to set its own criteria is what lets the Supplement scale, and it is consistent with the principle that the obligation is not the right decision but an informed one, reached through a process defensible to the scale of the risk.

The gap is not that the company sets the criteria. It is that setting them is itself a decision, and the exhibit does not treat it as one.

Proposed, four fields alongside the existing statement: the individual or body that approved the risk criteria; the level at which approval was given; the date of approval; and the date the criteria are scheduled for review.

7.2 Treat automate, augment and support as a grant of authority, not a property of the system

Exhibit C item 10 records AI model type as automate, augment or support. The Supplement defines Automation as "an AI System that does not involve human intervention" and Augmentation as one that "suggests an answer and/or advises a human who is making a decision." Those definitions are precise and the Center does not propose changing them.

The field built on them is the most consequential in the exhibit, and it is recorded as a property of the model. It is better understood as a grant of authority. Removing the human from a decision is a decision, and the difference between the two definitions above is the difference between a recommendation and an action taken in the company's name.

Proposed: for any model recorded as automating a decision, identify the individual or body that authorized operation at that level, the date, the scope of the authorization, and its expiration or next re-ratification date.

Authority to operate that does not expire is not authority. It is inertia.

8. The attestation in the third-party vendor framework

*A note on routing. The companion * Regulatory Framework for Third-Party Data and Model Vendors * is before the Third-Party Data and Models (H) Working Group, and its comment period closed August 5, 2026. This point is raised here because the two documents interlock and because Exhibit B addresses vendor procurement at item 3l and, as of version 5.0, vendor-developed AI Systems at the new item 3p. The Center offers the material to that Working Group directly if it is useful.*

The framework provides for an annual attestation by a senior officer that the vendor's governance program is implemented and effective for the data and models in scope. The Center has worked from the version exposed to that Working Group and does not quote it verbatim here; the point below turns on the pairing of implemented with effective rather than on the precise wording.

Most of that is sound. Requiring the attesting person to hold the approval authority rather than merely sign is a deliberate and correct choice, and *implemented* is an existence and operation claim that leader is positioned to make.

Effective is different in kind. Effectiveness is an assurance conclusion, and the framework asks the vendor to assert it about the vendor's own program with no independent party testing the assertion. Sarbanes-Oxley pairs the two deliberately: Section 302 is management's certification, Section 404(a) is management's assessment of internal control, and Section 404(b) is the independent auditor's attestation on it. The framework has written the 302 and not the 404(b).

The Center is not asking the Working Group to mandate an audit. Two narrower changes close the gap.

8.1 Separate what the officer attests to from what requires assurance

Ask the attesting leader to attest to implementation and scope, which that leader knows. Ask separately whether independent assurance has been obtained, from whom, on what scope and as of what date, treating no answer as information rather than as a deficiency.

8.2 Name the object of assurance as the commitment, not the framework

Where assurance has been obtained, the question is whether the vendor applied what it said it would apply, at the depth it recorded, rather than whether the vendor's program conforms to an external standard in the abstract.

This carries weight beyond insurance. OCC Bulletin 2026-13 of April 17, 2026, which replaced the 2011 model risk management guidance, states that generative and agentic AI models "are not within the scope of this guidance." The instrument financial supervisors have relied on for model risk does not reach the systems this Supplement is about.

9. What this comment does not ask for

It does not ask for a new requirement on any insurer. Every field proposed is a question, answerable from a document the company either holds or does not, and the absence of the document is itself the finding. Version 5.0's own addition of a Document Name and Page # column is the same principle, and the Supplement's own instruction that "the references to, and questions about, elements of a company's AIS Program are not intended to be interpreted as creating new requirements for AIS Program" applies to these items exactly as it applies to the existing ones.

It does not ask the Working Group to prescribe governance content. Asking who ratified a decision is not prescribing what the decision should be, which is the distinction Model Act #305 Section 1.B draws in the corporate governance context, as discussed in section 3. Nor does it ask the Working Group to abandon the optional and supplemental design, or the provision that each state may modify or add questions.

And it does not ask for anything the model bulletin has not established. The NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers, adopted December 4, 2023, expects a written program, vests responsibility with senior management accountable to the board, and outlines the information regulators may request. The Supplement is the operational expression of that bulletin, and these recommendations complete the expression rather than extending its reach.

10. About the Center

The Center for AI Oversight is an independent nonprofit research and education organization. Its tagline is *Defining oversight governance for AI in regulated industries*. It works with corporate boards and senior executives in regulated industries, with the professional bodies serving governance practitioners, and with regulators at the federal and state level. Center materials are educational, and the Center takes no positions on matters involving specific institutions.

Its advisory board includes Jim Cunha, former Executive Vice President, Federal Reserve Bank of Boston; Chuck Senatore, former Regional Director, U.S. Securities and Exchange Commission; Rear Admiral David Simpson, U.S. Navy (retired), former Chief of the Public Safety and Homeland Security Bureau, Federal Communications Commission; and Brian Peretti, former Chief Technology Officer, U.S. Department of the Treasury. Its Senior Fellows include Ankur Singhal, former Head of Technology Risk at PayPal and at The Hartford.

The Center's Executive Director, Brian J. Allen, Esq., was appointed by the Chairman of the Federal Communications Commission to represent the communications sector during development of the Cybersecurity Framework, and represented the banking sector through the Bank Policy Institute during development of the NIST AI Risk Management Framework.

This comment applies the Center's general position on the oversight layer, *Distinguishing the Oversight Layer of AI Governance*, available at cfaio.org.

Contact: info@cfaio.org

Proposed Additions to Exhibit B

AI Risk Evaluation Supplement, Version 5.0

Exhibit B: AIS Program (Checklist)

Submitted by the Center for AI Oversight, September 26, 2026

Companion to the Center's comment, "Exhibit B Is the Oversight Exhibit"

Summary

Five recommendations, expressed as one amendment to item 1 and seven new entries, plus one optional item. Each is answerable from a document the company either holds or does not. None asks an examiner to evaluate a model, a data source or a testing method. None prescribes governance content.

Proposed ref	Item	Completes
Item 1, amended	What the program governs, its granularity, who approved that as appropriate, the standards elected and the depth	What the program covers, and whether anyone decided it was the right depth for this company
2b, new	Committee and officer accountability, with delegation	Who answers, at two levels
2c, new	Competency at the accountable level	Whether the recipient can interrogate the report
2d, new	AI Systems declared out of scope, and who accepted them	The boundary of the program
3q, new	Ratified AI risk tolerance	Whether the tolerance is a decision
3r, new	Authority to operate, dated and expiring	Who authorized automation
3s, new	The escalation line	What happens when a system runs outside tolerance
3t, new	The disclosure position	What the company will report, and when
3u, new, optional	Independent assurance over the program	Whether the commitment was tested

On reference letters. These are drafted against the version 5.0 Checklist. The 2 series contains items 2 and 2a, so the accountability items are proposed as 2b, 2c and 2d. The 3 series runs continuously from 3a to 3p, with 3o and 3p new at this exposure, so the proposed 3 series items continue at 3q. The letters are a convenience rather than a recommendation, and the Center has no view on placement. Note that the Narrative option uses its own numbering, in which 2a to 2e already exist and address a different subject.

On the two options. Version 5.0 offers Exhibit B as a Narrative or a Checklist. These items are drafted for the Checklist. Part 3 notes where the Narrative already reaches nearby ground and what it leaves open, so the Working Group can decide whether to mirror the additions there.

Items in the 3 series complete the existing stem, *"Reference the processes and procedures of the Company AIS Program that addresses the following: How the Insurance Company..."*

Part 1. Clean text for insertion

Amendment to item 1

Existing (v5.0, Checklist item 1): "Has the company adopted a written AIS Program? If yes, when was it adopted and what is the frequency of review for updating?"

As amended:

Has the company adopted a written AIS Program? If yes, when was it adopted and what is the frequency of review for updating? State what the AIS Program governs and the granularity at which it operates, identify the individual or body that approved that scope and granularity as appropriate to the company's size, complexity and use of AI Systems, and state the date the scope and granularity are next reviewed for continued appropriateness. Identify the external frameworks or standards the company elected to apply to the AIS Program, the depth of application elected, and the individual or body that approved the election, with the date of approval.

Note on placement. As amended this item carries two distinct questions, one about the Program's own scope and granularity and one about the external instruments elected. The Working Group may prefer to split it into item 1 and a new item 1a. The Center has no view on which.

New items, 2 series

Ref	AI Systems Use Questions for Company	Company Response
2b	Identify the committee or governing body that holds oversight of the AIS Program by charter. Identify by name and role the officer accountable to it for the AIS Program, together with any delegation of that accountability and the date the delegation was recorded.	
2c	Describe how the company establishes that the accountable officer and the committee hold the competency required to interrogate what the AIS Program reports to them.	
2d	Identify any AI Systems the company has declared outside the scope of the AIS Program, the named individual who accepted each exclusion, and the date each exclusion is scheduled for review.	

New items, 3 series

Completing the existing stem, *"Reference the processes and procedures of the Company AIS Program that addresses the following: How the Insurance Company..."*

Ref	AI Systems Use Questions for Company	Company Response
3q	Ratifies AI System risk tolerance at the accountable level, identifying the body that ratified it, the date of ratification, and the date of next scheduled review	
3r	Grants, bounds and periodically re-ratifies authority to operate for AI Systems recorded as automating a decision, recording the authorizing individual or body, the date of authorization, its scope, and its expiration or re-ratification date	
3s	Escalates an AI System operating outside stated tolerance, identifying (i) the standing cadence on which the AIS Program reports to the accountable committee independent of events; (ii) whether that reporting carries adverse information, incidents, findings, regulatory inquiries and deviations from tolerance and not only progress; (iii) the escalation triggers and materiality thresholds fixed in advance; (iv) the party with authority to continue, constrain or withdraw the System; (v) whether at least one reporting pathway to the governing body does not depend on management's discretion; and (vi) whether the response to an escalated matter is recorded	
3t	Fixes the company's disclosure position for AI System matters, identifying the criteria for disclosure to the department, to consumers and to the public, and any clock that statute or regulation attaches to it	

Optional item

Ref	AI Systems Use Questions for Company	Company Response
3u	Obtains independent assurance over the AIS Program, identifying the provider, the scope of the review and the date of the most recent review. Where assurance has been obtained, the object of review is whether the company applied the frameworks and depth of application it recorded in response to item 1.	

A blank response to the optional item is information rather than a deficiency. Independent assurance over an AI systems program is not yet common practice, and the value of the question in the near term is that it establishes a baseline.

Part 2. Basis for each item

Not proposed for inclusion. Recorded for the Working Group's consideration.

The amendment to item 1. The existing item establishes that a program exists and when it was adopted. It does not establish what the program governs, or what the company undertook to do.

The object clause is four words in the exhibit and it separates two different documents. Version 5.0 defines an AIS Program as "the controls and processes that an insurer adopts and implements to ensure the

responsible use of AI Systems," which "may be comprised of general guidelines, governance framework, risk management and internal controls, and third-party oversight." That is a definition of what a company does, and Exhibits A, C and D already reach that ground thoroughly. A program whose stated object is instead the company's capacity to make informed decisions about those systems is an oversight program, which is what this exhibit is positioned to test. Without the clause the examiner cannot tell which one was filed, and a company that has built only the first will answer the rest of the exhibit in terms of its model inventory.

"Depth of application elected" is deliberate: a company that adopts a framework in part, deferring specified elements, has made a legitimate decision that should be recorded rather than obscured. The approver is asked for because adoption with no identified approver is an administrative act, not a governance decision.

2b, committee and officer. Four existing questions come close. Checklist item 2 asks whether the Board of Directors or management was "involved in the adoption," and item 2a asks "what is the role of the Board of Directors or management in the AIS Program." Narrative item 1a asks "what role maintains the AIS Program." Narrative item 1e asks how "responsibility for governance within the organization is assigned."

A company can answer all four accurately and leave the examiner unable to name anyone. Maintaining a program is an administrative function, assigning responsibility describes an organizational chart, involvement in adoption is a historical fact, and a role is a description rather than a name. None of the four resolves to a party who answers.

Insurance regulation already has the naming mechanism in two places, quoted in the accompanying comment: the CGAD signature under Model Act #305 Section 3.B, and the ORSA Summary Report signature under Part C of the Guidance Manual's Introduction. This item asks the AIS Program equivalent of a question the sector already asks on adjacent subjects. It asks for a role as well as a name, because the role survives personnel changes and can be tested against the company's own delegation records at the next examination.

2c, competency. The ORSA attestation confirms that a copy of the report was "provided to the insurer's board of directors or the appropriate committee." Provision is receipt. What makes receipt into oversight is the capacity to interrogate what was received. The item does not ask the accountable officer to validate a model. It asks whether the company can show the officer is equipped to question a report about one.

2d, exclusions. The item most likely to produce a finding, and the one with no existing analogue in either option. Every AI systems program has a boundary. Systems fall outside it because they are low risk, because they are legacy, because they were acquired with a book of business, or because no one has decided. The first three are decisions. The fourth is not. A company answering Exhibits A and C in good faith may simply not report systems it considers out of scope, leaving the examiner unaware that a scoping judgment was made at all. A company that cannot name an acceptor has not made a decision about the system. It has omitted it.

3q, ratification. Existing item 3k asks how the company "Quantifies AI System risk levels," and new item 3o asks how it "Determined 'materiality' and what threshold was used for assessing AI Systems risk." Both ask for the assessment and how it was derived. Assessment is a technical act performed by a modeling or risk function, and it may be qualitative as readily as quantitative. Ratification is a governance act performed by a body that can be held to it, and a tolerance assessed but never ratified is a working assumption the company is not committed to.

This item is also what makes Exhibit C's provision that "AI model risk criteria are set by the insurance company" supportable. The Center supports that provision, because allowing the company to set its own

criteria is what lets the Supplement scale from a small domestic carrier to a global group. But criteria set by the company are a decision by the company, and this is where that decision is recorded.

3r, authority to operate. Existing item 3e asks how the company "Evaluates whether AI Systems are suitable for their intended use and should continue to be used as designed," and Narrative item 1g asks how it "assess[es] autonomy, reversibility, and reporting impact risk." Both are the right questions asked in the passive register of evaluation rather than the active register of authorization. Assessing the risk of autonomy is not the same as authorizing it.

Exhibit C item 10 already records AI model type as automate, augment or support, and the Supplement defines Automation as "an AI System that does not involve human intervention." Removing the human is a decision, not a design outcome. The four attributes requested, authorizing party, date, scope and expiration, are the minimum that makes an authorization testable. An authority that does not expire is not an authority.

3s and 3t, escalation and disclosure. Narrative item 1b asks the company to discuss "the governance structure, Board reporting and frequency," which establishes that reports move and how often. Checklist items 3m and 3n are consumer-facing. None of them establishes what the report carries, what triggers one outside the cadence, who can act on it, or what was done.

Escalation is not incident response. Incident response addresses events already recognized as incidents. Escalation addresses the system behaving as designed and outside the boundary the company set for it, which is the harder case and the one that produces supervisory surprise.

Each element of 3s is testable from the company's own records. A standing cadence establishes that reporting does not wait on something going wrong. Adverse information establishes that the report is decision-useful rather than a progress update. Triggers fixed in advance establish that the threshold was not set after the event. A destination with authority to withdraw establishes that escalation resolves to an action. A pathway independent of management's discretion establishes that the governing body can learn what management would prefer it not learn. And a recorded response establishes what was done, because a red flag received without one is, on the record, a red flag ignored.

Disclosure is proposed as a separate item at 3t rather than a seventh element of 3s, because it is answered by a different document and often by a different function. A company that has not fixed its criteria in advance will fix them under pressure, and the department is the party most affected by that.

3u, independent assurance. Narrative item 1d asks how "the effectiveness of the AIS Program and individual models are assessed and modified," and Narrative item 3a asks whether validation and testing are "performed by someone that is independent from development." The Working Group has therefore already applied the independence principle, but to model validation rather than to the program.

The object of assurance stated in the item is deliberate. The useful question is not whether a program conforms to an external standard in the abstract. It is whether the company applied what it said it would apply, at the depth it recorded in item 1. That is testable, it is within the competence of an internal audit function or an external provider, and it does not require the assurance provider to evaluate a model. ISO/IEC 42006:2025, published July 7, 2025, gives this market a published accreditation standard, which allows the Working Group to reference the availability of independent assurance without mandating its use.

Part 3. Three observations on the existing exhibit

No new items required.

Items 2 and 2a both read "Board of Directors or management." The disjunctive is understandable in an instrument accommodating a range of company sizes and structures, and item 2a already asks the right question in asking for the role. It also permits a response that is accurate and uninformative, because a role can be described without anyone being named. If the Working Group prefers not to add a separate accountability item, the narrower alternative is to amend 2a to ask which body holds the role, under what instrument it holds it, and who answers for the program by name.

The 3 series stem is "How the Insurance Company..." Every item in the series is therefore a process question, which is appropriate for most of the series. Items 3q through 3t are decision questions rather than process questions. The Working Group may prefer to group them under a separate stem, for example *"Identify the following decisions and the party that made them,"* rather than extending the existing series.

The Narrative option asks companies to discuss; the Checklist asks them to point at a document. That difference is the Working Group's own design and the Center does not propose changing it. It is worth noting only that a discussion is not a record. Where the Working Group elects to mirror these additions in the Narrative, the Center would suggest preserving the identifying verbs, so that the company identifies a committee, an officer, a body, a date and a party rather than discussing how accountability is approached.

Proportionality

For a small or resource-constrained insurer, 2b and 3s carry most of the value, with 3r added wherever a system automates a decision. That minimum viable form is why an oversight instrument scales from a small domestic carrier to a global group without a separate small-entity version.

Date: September 29, 2026

From: NAIC Consumer Representatives

**To: Big Data and Artificial Intelligence (H) Working Group Chair Commissioner Nathan Houdek,
Vice Chairs Commissioners Mary Block and Doug Ommen**

CC: Miguel Romero and Scott Sobel

Re: Comments to AI Risk Evaluation Supplement

We, the undersigned Consumer Representatives to the National Association of Insurance Commissioners (NAIC), appreciate the important work of the Big Data and Artificial Intelligence Working Group to assist regulators to identify ways to monitor and oversee the growing use of AI across insurance operations and claims adjudication. Tools like the AI Risk Evaluation Supplement are critical to assisting regulators to operationalize the broad principles of consumer protection and regulatory oversight over AI articulated in the AI Model Bulletin. Below are our comments on the most recent draft of the AI Risk Evaluation Supplement.

Background, Intent, and Instructions

The Background discussion seems unnecessarily timid when it states, that “However, without robust governance and effective controls, the use of AI Systems may lead to Adverse Consumer Outcomes” We know that it likely will, if not almost certainly, lead to adverse consumer outcomes, and the Supplement can and should state this more explicitly. In the last two weeks alone, there have been multiple stories about AI systems going rogue, with the possible harm far greater than negative consumer outcomes. We think tone is important, especially in the first paragraph, and the possible risks are why this Supplement is being drafted. Substituting “may lead” with “likely to lead” or “carries a significant risk” that could lead to “adverse consumer outcomes” may be more accurate ways to describe the stakes of this conversation.

Instructions

We appreciate the updates to the instructions and think this section gives better guidance to both regulators and regulated entities. However, we think that it is important to identify state non-discrimination laws more explicitly at the outset, which may operate differently than unfair trade practices statutes. Identifying the risk of “unfair discrimination and other adverse consumer outcomes” that is referenced at the bottom of p. 2 with regard to machine learning and generalized linear models requires the common understanding, terminology, and precedents that exist in state and federal non-discrimination laws. We also think that it is important to recognize privacy protection as conceptually distinct from confidentiality in light of AI’s capacity to mine and use personal information from sources without an individual consumer’s knowledge or express consent. We recommend the following language:

- Effective assessment requires regulators to maintain a fluent understanding of the applicable laws and regulations, including but not limited to those pertaining to unfair

trade practices, unfair claims settlement practices, non-discrimination protections, corporate governance annual disclosures, confidentiality, privacy protection, financial reporting, and rating.

Confidentiality

We think the statement regarding confidentiality should be amended to specifically reference state freedom of information laws, which may lead to different confidentiality standards depending on the jurisdiction. The term “consistent with the requesting state’s relevant authority” may be accurate, but seems unnecessarily ambiguous.

Exhibit B: (Checklist) AIS Program

While we appreciate the inclusion of question 3(a) on how the company assesses, mitigates, and evaluates residual AI system risks of unfair trade practices, we think this question should be expanded to reference that some states have broader definitions of “unfairly discriminatory” that go beyond simply actuarial fairness, including the risks of bias and discrimination against protected classes. In many if not most states, there are specific prohibitions in the state’s insurance code or elsewhere in state law, that prohibits discrimination on the basis of race and ethnicity, and often additional characteristics. Limiting unfair discrimination to actuarial fairness is no longer the default model. Similarly for question 3(b), given the specific and acute risk to particular communities, we think the reference to “state and federal laws and regulations” broadly is not sufficient. We recommend adding potential relevant statutory frameworks, including non-discrimination protections.

In Section 2b, The “human in the loop” question is important, but it may not provide regulators sufficient information. To make a human in the loop relevant, the checklist should also include the amount of time the human is expected to take to review model outcomes, and also the process the human must go through to supersede or amend the model’s outcome. There are reports that human review of model outcomes may be less than 10 seconds, which provides little assurance that the company utilizing AIS is providing meaningful human review prior to finalizing the outcome. A review process that encourages the human to approve model outcomes and discourages overriding them, may also jeopardize meaningful oversight. Additional language in this section could include “whether there is a qualified and objectively authorized human in the loop to provide meaningful human review.”

Exhibit C: High-Risk Model Details

We continue to be concerned that AI model risk criteria are set by the insurance company. We believe that this poses an inherent conflict of interest and that consumers would benefit from a uniform definition of high-risk models. Moreover, agentic AI developments of the recent past would make every such system a potential source of high-risk models since AI could independently develop such models as the best way to efficiently achieve corporate goals.

Exhibit D: AI Model Data

We are concerned that the data details included may not be sufficiently representative of data sources used in the health insurance line. At row 16, the reference to “Medical, including Biometrics, genetic information, pre-existing conditions, diagnostic data, etc” fails to capture the reality that many sources of health information that would be commonly considered “medical” in nature are not collected/analyzed/shared by entities that are recognized as healthcare entities. For example, AI models could scrape data collected by fitness and retail apps to identify purchases of health-adjacent over-the-counter items such as pain killers and incontinence supplies, irregular exercise patterns, and dietary/allergy information. The increased prevalence of lesser regulated insurance products means that the potential consumer profiles that can be built from these small nuggets of data can become relevant to AI decision-making in a health insurance context, irrespective of their accuracy or overall relevance.

Row 20’s reference to reasonable accommodations is an attempt to capture one such source of information that could be unfairly used in a health context as well as other insurance contexts such as life or auto insurance, but there are others as well that should be more explicitly captured.

Definitions and Appendix

While we appreciate the Working Group’s attention to the definition section, we are concerned that some of the new definitions no longer track the language in the rest of the supplement. For example, there is now a separate definition for “direct consumer impact.” However, in Exhibit C of the supplement the term “material consumer impact” was used and it’s not clear what the difference between these terms is.

We are also concerned that separating out definitions for “consumer impact” and “direct consumer impact” could cause confusion and inconsistent reporting of information. The definition of “direct consumer impact” could be interpreted as a subset of “consumer impact” but even within that subset, there seems to be an implicit recognition that “AI Systems used for internal analytics, model monitoring, and back-office tooling with no consumer-visible output or other internal operational or administrative functions” are risk free. However, these are systems that can deeply affect how consumers interact with an insurer and how they can try to get recognition of and recourse for erroneous outcomes. For instance, a “back-office tooling” system may determine when and how consumers can get past a chatbot, how complaints are received and categorized, what happens to requests for data updates/corrections, and how customer service and satisfaction queries are distributed and the answers mined for feedback.

We recommend combining these terms and noting within the definition that consumer impact will ultimately vary depending on the specific insurer action. This will allow regulators, and not regulated entities, to decide what rises to the level of direct impact that warrants further scrutiny.

We look forward to continuing to partner with you to provide the consumer perspective in this process.

Consumer Representatives,

Lucy Culp
Adam Fox
Amy Killelea
Peter Kochenburger
Silvia Yee

1. Goals for Financial Analysis in the AI Risk Evaluation Supplement

With the intention of the document as a supplement that would assist regulators to identify and assess both financial and consumer risks from the use of AI Systems on an on-going basis. If the focus is on financial exams for financial risks and market conduct exams for consumer risks, we should clarify the goals of including Financial Analysis Handbook and financial analysis in this supplement.

2. Consistency and Principal Based AI Risk Evaluation

- a. On page 1, we should add ("FCEH") after Financial Condition Examiners Handbook as it was used in the document but never defined.
- b. Under intent on page 1, we should replace tools with supplements for consistency.
- c. Company Contacts is listed and we should keep the spirit of Principal Based AI Risk Evaluation and not ask Company Contact for Follow Up inquiry by functions and SME, etc. The regulators need to discuss the risks with the company accordingly.

3. Exhibit C:

Exhibit C, question 11: Discuss the testing and validation of model outputs, including, as applicable, testing for model drift, accuracy, performance degradation, unfair trade practices, unfair discrimination, and other potential Adverse Consumer Outcomes. Describe the testing methodologies and metrics used, the populations or comparison groups evaluated, the frequency of testing, and how results are evaluated and documented. Discuss validation performed prior to deployment and ongoing monitoring after deployment, including testing following material changes to the model, data, or use of the model. Where testing identifies material differences or potential Adverse Consumer Outcomes, discuss the company's process for evaluating the results, identifying contributing factors, and determining whether changes to the model, data, methodology, controls, or other reasonably available alternatives are appropriate.

ADD 11.a: 11a. Where applicable, discuss whether the company evaluates input variables, derived variables, or model outputs for potential proxy relationships or effects that may contribute to Adverse Consumer Outcomes or unfair discrimination, including the methodology used and any resulting action taken.

Big Data and Artificial Intelligence (H) Working Group

Re: Comment — AI Risk Evaluation Supplement (version 5.0)

Dorvelon welcomes the opportunity to comment on the AI Risk Evaluation Supplement (version 5.0). We support positioning the instrument as a supplement to existing examination procedures rather than as a freestanding AI examination regime, and we support leaving risk criteria with the company that owns the system.

Our comment addresses one point. Version 5.0 introduces a definition of agentic AI and counts generative and agentic models separately, but the risk classification in Exhibit C is not built to carry what that definition implies. For AI systems whose exposure arises from what they can reach and what they can do, rather than from how well they predict, the classification will not report the risk the examiner is looking for — and the company that produces it will not have done anything wrong. This comment draws on direct experience setting firm-wide model definition and risk-tiering standards, and auditing enterprise model inventories, under continuous federal examination.

WHAT VERSION 5.0 RESOLVED, AND WHAT REMAINS OPEN

Several changes in this draft are material improvements, and we note them because they narrow the gap we are describing. Exhibit A now counts generative and agentic models separately under both direct consumer impact and material financial impact. Exhibit C Rows 5 and 6 distinguish internal development from an “owned” third party and a true third party, and ask whether the company can modify the model. Exhibit B adds item 3k on quantifying risk levels and item 3l on standards for procuring and engaging AI System vendors, and the evidence column now requires a document name and page number rather than a page number alone.

What remains open is the risk taxonomy itself. Exhibit C states that “AI model risk criteria are set by the insurance company,” and adds that “materiality should be a consideration in determining the level of risk.” Row 7 then requests a “Model risk classification (high, moderate, low, etc.).” Row 8 defines what is being classified: “Model risk(s) (i.e., describe the potential for adverse consumer impact).” Row 11, which governs testing, enumerates model drift, accuracy, unfair trade practices, unfair discrimination, and performance degradation. Each of these is a property of how the model performs.

Information security is not treated as a risk dimension anywhere in the Supplement. It appears twice: in the Committee’s own name, and in the Exhibit A “Other” business-function list — a place where AI is applied, not a risk AI creates. Privacy appears once, at Exhibit B item 3d, which asks how the company “considers data

privacy and protection of consumer data used in AI Systems.” That is an input-side control. It does not reach the model as a disclosure vector: memorization of training data, extraction through queries, or inference from outputs.

THE DEFINITION OF AGENTIC AI IS NOT REFLECTED IN THE CLASSIFICATION

The Supplement defines agentic AI as systems that “can pursue objectives across multiple steps without requiring human input at each stage,” involving “taking actions (often through tools or APIs),” and observes that “the difference is in agency.” We agree, and regard this as the most consequential addition in version 5.0.

Nothing downstream asks about agency. Row 3 records that a model is generative or agentic. No row asks which tools or APIs it can invoke, what data it can reach, under whose identity and permissions it operates, whether any action executes without human review, or how an action outside its authority would be detected. A model is identified as agentic and then classified on properties that have nothing to do with its agency.

The result is a classification that is internally correct and reports the wrong thing. Consider a generative assistant supporting claims handling. It neither prices nor underwrites; assessed on drift, accuracy, and disparate outcomes it classifies low or moderate, and that assessment is sound. But it may hold read access to claim files containing health and financial information, and emit free text directly to a consumer. Its capacity to produce an Adverse Consumer Outcome derives from its data access and its output channel, neither of which any row records.

Consider also a true third-party model. Rows 5 and 6 now establish provenance and whether the company can modify it — a real improvement — but not whether the company can test or inspect it. A company with no contractual right of inspection has no evidence on which to rate that model high, and absence of evidence resolves downward. The models a company understands least remain the ones most likely to be classified low.

This matters because the classification gates the examination. Regulators are directed to “first use Exhibit A to determine if further inquiry using Exhibits B-D is necessary,” and Exhibit C collects detail on models the company has designated high risk. A model classified low does not surface.

TWO ASSESSMENTS ALREADY EXIST; THE SUPPLEMENT DOES NOT CONNECT THEM

In most insurers, model risk and information security are assessed by different functions, against different inventories, on different cadences, using different rating scales. An AI system appears in both. Nothing in the Supplement tells an examiner how the two assessments relate.

The Working Group has already solved this problem for other regimes. Exhibit B asks how the company considers AI System risks within Enterprise Risk Management (item 3f), within the Own Risk and Solvency Assessment (item 3g), and within the software development lifecycle or the broader AIS Program (item 3h). The pattern stops at the software lifecycle and does not reach information security.

The NAIC holds the other half. The Insurance Data Security Model Law (#668) establishes the information security program, board oversight, third-party service provider oversight, and incident response; the information technology review in the Financial Condition Examiners Handbook is the corresponding examination instrument, and the Supplement already coordinates with that Handbook. We are not proposing a new assessment — only that the Supplement state how the one it creates relates to those already in force.

We recommend stating a dependency rather than prescribing a sequence. The information security assessment establishes the system boundary: what data the model can reach, what it can write or transmit, the identity and permissions under which it operates, and which components a third party controls. Those facts are inputs to the risk classification, not a parallel track. A stated dependency preserves the examiner's discretion over sequencing and scope, which the regulator instructions rightly protect.

RECOMMENDED AMENDMENTS

Row and item references are to version 5.0 as exposed; the Working Group may place the substance wherever it best fits.

- Exhibit C, Row 7 — require the company to state which risk dimensions the classification incorporates, or to identify the dimensions it excluded.
- Exhibit C, new row — Data access scope: the classes of data the AI model can read or retrieve, and the identity and permissions under which it operates.
- Exhibit C, new row — Action authority and output channel: which tools or APIs the model can invoke, whether any action executes without human review, and whether outputs reach consumers directly or internal users only. This gives the agentic definition somewhere to land.
- Exhibit C, new row — Information security assessment: the date and reference of the most recent information security assessment of the AI System, and whether it preceded the risk classification.
- Exhibit C, new row — Adversarial and disclosure testing: whether the model has been tested for prompt injection, data exfiltration, model extraction, membership inference, or memorization of training data; by whom; and when. Row 11 reaches none of these.
- Exhibit C, Rows 5 and 6 — extend beyond provenance and modifiability to whether the company holds contractual rights to test, inspect, or obtain documentation for a true third-party model, and what it was unable to obtain.
-

Exhibit B, item 3 series — add an item asking how the company coordinates its AI System risk assessment with the information security program required under the Insurance Data Security Model Law (#668), where adopted. Item 3l addresses vendor procurement, but not this coordination.

- Exhibit A — the gate should capture data access and output channel alongside the generative and agentic counts, so that a model that is low-materiality by business function but high-exposure by access still reaches Exhibit C.

CLOSING RECOMMENDATION

We recommend one clarification: a risk classification is interpretable only when the dimensions it incorporates are declared. This preserves the Supplement's structure — criteria set by the company, scope controlled by the regulator — while ensuring that a classification of “low” reports a judgment the examiner can read, rather than the boundary of the taxonomy that produced it. We would welcome the opportunity to contribute further as the Supplement advances toward adoption.

Respectfully submitted,

Jucheng (Jim) Zhao, PhD, CFA, AIGP

Principal AI Governance Architect
Dorvelon

From Iman Schrock (EMILIA Protocol, Inc.)

Dear Mr. Sobel and Mr. Romero,

Version 5.0 asks insurers to assess autonomy, reversibility, human involvement, and how exceptions or errors are resolved. For systems that can take actions through tools or APIs, four additional questions would make those controls easier to examine: can the company prove that the action executed was the action authorized?

This is becoming a live responsibility question. Visa's April 2026 rules (section 4.1.24.3) require agentic payment providers to obtain cardholder acknowledgement that the cardholder is responsible for actions the provider takes. American Express's published Agent Purchase Protection terms identify eligible charges involving purchases that deviate from authenticated intent, although full terms are forthcoming. Both developments make the evidence behind an agent-initiated action relevant to insurance risk.

We suggest adding the following under the direct consumer impact and material financial impact questions:

1. For AI Systems that take actions with financial or consumer consequence, describe how the authorization for an action is bound to the specific action executed (counterparty, amount, resource, and effect), rather than to a session, a role, or a general permission.
2. Describe whether the authorization is verified at the point where the action becomes irreversible, and by which party. If verification occurs only upstream, describe how the company detects an action that was altered, widened, or substituted between approval and execution.
3. Describe whether an authorization can be used more than once, and how the company prevents a retry, a duplicate submission, or a replay from producing a second consequential effect under the same approval.
4. Describe how the company records and resolves actions whose outcome is unknown at the time of failure (for example,
5. a timeout after a payment instruction has left the company's control), including who owns the unresolved case and whether the original authorization can be reused during resolution.
6. Describe whether a party other than the company, such as an examiner or auditor, can confirm from the evidence itself who authorized a specific action and what exactly was authorized, without relying on records or conclusions of the party that operates the AI system.

These questions do not prescribe a protocol or evidence format. They ask whether the company can produce evidence that lets an examiner distinguish an authorized execution, a duplicate or substituted execution, and an unresolved outcome. That directly extends the draft's existing request for how exceptions or errors are found and resolved.

Ali Houshang, MSc, CISA, CISSP

Huntington Beach, California | ali.houshang@gmail.com
September 8, 2026

Scott Sobel, AI Policy Advisor

Miguel Romero, Director of Innovation, Cybersecurity and Technology

National Association of Insurance Commissioners

Re: Comments on AI Risk Evaluation Supplement version 5.0 - action authority, control boundaries, and review evidence for agentic AI systems

Individual capacity. These comments are submitted solely in my individual capacity. They do not represent any employer, client, professional association, or other organization.

Dear Mr. Sobel and Mr. Romero:

Thank you for the opportunity to comment on the National Association of Insurance Commissioners (NAIC) AI Risk Evaluation Supplement version 5.0 (v5.0). I have more than fifteen years of experience evaluating technology risk, governance, internal controls, and evidence in regulated environments. My request is narrow: add two yes/no triage questions in Exhibit A, two conditional governance questions in Exhibit B, and four items in Exhibit C consisting of one use-case trigger and three conditional fields. The purpose is to make the Supplement's existing treatment of agentic AI and automation follow through to the actions an AI System is authorized to execute, the controls that constrain those actions, and the records used to review what occurred. [1]

I am not asking the Working Group to create a separate agentic AI risk category, to treat agentic AI as inherently high risk, or to prescribe an architecture. v5.0 already contains the concepts needed to make a modest, technology-neutral refinement. The proposed wording below is intended to fit within the Supplement's existing risk-based and examination-oriented structure. [1][4]

1. v5.0 already recognizes execution and human involvement

1.1 Agentic AI is defined at the AI System level and by execution

v5.0 newly defines Agentic AI as an AI System that can pursue objectives across multiple steps without human input at each stage and may take actions through tools or APIs. The definition closes by distinguishing agency in practical terms: the system "actively executes a process." v5.0 separately defines an AI System as operating with varying levels of autonomy. These are system-level concepts, not merely model labels. [1][2]

That distinction matters because a structured inquiry that recognizes execution should be able to ask, when the use case is sufficiently important, what the system is allowed to execute. This is not a definitional request. The Supplement already has the definition. It is a follow-through request.

1.2 Exhibit C field 10 already makes the human-involvement distinction

Exhibit C is the high-risk deep dive. Its purpose expressly includes in-production high-risk models making automated decisions. Field 10 then asks whether the AI model type is "automate, augment, support." The definitions make the distinction concrete: Augmentation advises a human decision-maker, while Automation is an AI System that does not involve human intervention. Exhibit A similarly treats models that augment or automate consumer decision-making as having direct consumer impact. [1]

This is the strongest reason for the proposed change. The structured inquiry already identifies the no-human-involved case. The remaining question is whether the fields that follow allow a regulator to understand the authorized action, the relevant control boundary, and the evidence available to review a particular action. Today, they do not ask those questions expressly.

2. The gap is follow-through into action authority, controls, and evidence

The existing fields address important model questions. Field 9 asks about model limitations and guardrails to manage model-related risk. Field 11 asks about model-output testing and ongoing performance monitoring. Field 13 asks how the model affects financial statements, risk assessment, or controls. Fields 14 and 15 address legal compliance review and prior regulatory action. Those are important inquiries, but none expressly asks three action-level questions: what the AI System may execute, what pre-execution control applies to that action, and

what documentation or records can be used afterward to determine what occurred and whether the control operated. [1]

The evidence dimension is particularly easy to miss because it is not simply another model-performance question. The Supplement's structured requests contain no use of "reconstruct" or "reconstruction," and no use of "log," "logs," "logging," or "audit trail." The sole occurrence of "retention" is "retention modeling" in an underwriting example, not records retention. These textual observations do not prove that regulators lack authority or that companies lack evidence. They show only that this dimension is not expressed in the Supplement's structured requests. See Appendix B.

Illustrative claims example

Consider a high-risk claims intake use case within Exhibit A's existing Claims/Adjudication row. Assume an AI System meeting the Supplement's Agentic AI definition can perform three material actions: open a claim record; adjust a reserve; and initiate a payment step. Field 10 can record that the model is used to automate, but it does not identify which of those actions the AI System may execute. Field 9 can describe model limitations or guardrails, but it is framed around model-related risk rather than the system's action authority. Field 11 tests what the model produced, not what the system subsequently did. [1]

- Which of those three actions may the AI System execute without a person approving that action?
- What limit or pre-execution control applies to each material action?
- For a given reserve adjustment, what record allows a reviewer to determine the action taken, the outcome, and whether the applicable control operated?

Those questions can be answered without a new taxonomy and without dictating technology. They are the questions Section 3 is designed to add.

The point is also consistent with the NAIC's existing broader framework. The Model Bulletin calls for identification of constraints and controls on automation and design, includes data and record retention within risk management and internal controls, and describes information and documentation that may be requested during regulatory oversight. At the same time, the Bulletin states that its goal is not to prescribe specific practices or documentation requirements. The proposed wording below follows that same balance. [4]

For context only, draft minutes included in the Working Group's August 31, 2026 materials record an AM Best presentation that distinguished action-oriented workflows and discussed system access, action records, intervention, and reversibility. The presenter expressly stated that the discussion did not change AM Best criteria or create a formal AI governance assessment framework. I cite this only as contemporaneous discussion material in the Working Group record, not as an NAIC position and not as evidence of what insurers generally do. [5]

3. Proposed drafting

The numbering below is illustrative. Where lists are still evolving, I recommend functional placement, for example following the final lettered item under Exhibit B item 3 or following the final numbered request in Exhibit C, so the drafting remains stable if items are inserted elsewhere.

3.1 Exhibit A - system-level triage

First, rename columns 5 and 8 from "Generative/Agentic AI Models" to "Generative AI Models." Agentic AI is defined as a property of an AI System, while those columns count models. Removing "agentic" does eliminate v5.0's only explicit agentic count, but a model-type column cannot reliably count a system-level behavior. The two system-level yes/no questions below more directly answer the triage question the combined label appears to be reaching for. [1]

Placement / item	Proposed wording
Exhibit A triage 1	For each reported use case, does the use case include an AI System that meets the Supplement's definition of Agentic AI? Yes / No.
Exhibit A triage 2	If yes, is that AI System authorized to execute one or more actions that can directly create or materially alter a consumer, financial, or financial-reporting outcome without human approval before the action? Yes / No.

3.2 Exhibit B - conditional governance questions

The two concepts should appear in both versions of Exhibit B. In the Narrative version, they can be combined in a single new item following the final sub-item under item 2. In the Checklist version, they should appear as two separate items following the final lettered item under item 3, illustratively 3q and 3r.

Placement / item	Proposed wording
Narrative item 2f	For an AI System that meets the Supplement's definition of Agentic AI and is authorized to execute actions without human approval before each action, discuss (i) how the company establishes and maintains the boundaries of the system's action authority and the controls applicable to material categories of action, and (ii) how the company determines what documentation or records are maintained to permit review of actions taken and whether applicable controls operated, consistent with the AIS Program, applicable law, and applicable data and record retention practices.
Checklist item 3q	For an AI System that meets the Supplement's definition of Agentic AI and is authorized to execute actions without human approval before each action, establishes and maintains boundaries on the categories of actions the AI System is authorized to execute and the conditions or limits applicable to that authority.
Checklist item 3r	For such an AI System, maintains documentation or records sufficient to review actions taken and the operation of applicable controls, consistent with the AIS Program, applicable law, and applicable data and record retention practices.

3.3 Exhibit C - one trigger and three conditional fields

Placement / item	Proposed wording
After the current model-type request (illustratively 3A)	Does the use case include an AI System that meets the Supplement's definition of Agentic AI? If yes, briefly identify the AI System and the objectives it pursues across steps, and indicate whether it is authorized to execute one or more actions without human approval before each action. If no, the conditional action fields below need not be completed.
Following the final numbered request (illustratively 16)	For the AI System identified above, describe the material categories of actions it is authorized to execute in this use case and the material limits or conditions on that authority, including thresholds, prohibited actions, escalation conditions, or human approval requirements where applicable.
Following the final numbered request (illustratively 17)	For each material category of action identified above, describe the pre-execution control that applies, such as human approval, a rule-based gate, an escalation threshold, or another control; the conditions that trigger the control; and what occurs if the control condition is not satisfied.

Placement / item	Proposed wording
Following the final numbered request (illustratively 18)	Describe the documentation or records relied upon to review or reconstruct a specific action, determine the action taken and resulting outcome, and determine whether the applicable controls described above operated. Identify material limitations, including limitations arising from third-party systems. Describe how the relevant documentation and records are addressed under the company's AIS Program, applicable law, and applicable data and record retention practices.

4. Scope, burden, and confidentiality

The proposal is intentionally conditional. Exhibit A performs lightweight triage. Exhibit B asks about program-level governance only where the identified system has action authority. Exhibit C asks the detailed questions only in the high-risk deep dive and only where the use case includes the relevant agentic, action-taking system. This structure preserves v5.0's existing risk-based sequencing. [1]

v5.0 already permits a company with an existing AI Systems inventory to propose submitting that inventory instead of completing Exhibit A. The two added triage questions can therefore be handled in the same inventory-based scoping process rather than through a new standalone exhibit. I do not assume that every existing inventory already contains those two fields; a company may need to supplement its inventory response. [1]

The proposal also fits within the Supplement's existing confidentiality framework. v5.0 states that information requested through the Exhibits will be held confidential consistent with the requesting state's authority and directs regulators to cite relevant confidentiality statutes or protections. The proposed fields would sit inside that existing protection rather than create a separate disclosure channel. [1]

The requested information is deliberately limited. It does not seek credential values, endpoint details, API inventories, system prompts, proprietary source code, or a prescribed technical record format. Where third-party limitations affect the ability to review an action, field 18 asks that the limitation be identified rather than assuming perfect observability.

5. What this proposal does not do

- It does not treat agentic AI as inherently high risk. Risk still depends on the use case, impact, materiality, and the company's risk assessment.
- It does not prescribe identity, gateway, tracing, logging, or other technical architecture. The question is what authority and evidence exist, not how a company must implement them.
- It does not require human approval for every action. Human approval is one possible control; the proposed fields also permit rule-based gates, escalation thresholds, or other controls appropriate to the action and risk.

6. Qualifications and limitations

I am not an actuary and do not offer an actuarial view on model-risk classification, reserving methodology, or materiality thresholds. I did not participate in the NAIC pilot and have not relied on nonpublic carrier submissions, confidential examination information, or other nonpublic insurer data. These comments are based on the public text of v5.0, related public NAIC materials, and my experience evaluating control design and evidence.

Thank you for considering these comments. I respectfully request that the Working Group consider the Section 3 drafting for inclusion in the next exposure draft. I would be glad to appear on a Working Group call or provide further detail if that would be useful.

Respectfully submitted,
Ali Houshang, MSc, CISA, CISSP
 Huntington Beach, California
 ali.houshang@gmail.com

Appendix A - Public sources

[1] [NAIC, AI Risk Evaluation Supplement version 5.0 \(clean exposure draft\)](#) Accessed September 8, 2026.

- [2] [NAIC, AI Risk Evaluation Supplement - Summary of Changes from version 4.0](#) Public exposure materials, 2026.
- [3] [NAIC, AI System Evaluation Tool 4.0 - Clean](#) Prior public version, used only for change context and structural checks.
- [4] [NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers](#) Adopted by Executive (EX) Committee and Plenary December 4, 2023.
- [5] [NAIC Big Data and Artificial Intelligence \(H\) Working Group, Materials for August 31, 2026 public meeting](#) Includes draft August 13, 2026 minutes and the draft Supplement materials. The AM Best discussion referenced in Section 2 is characterized as discussion material, not an NAIC position.

Appendix B - Textual and structural checks

Method. The term counts below were independently checked against the clean v5.0 document across all paragraphs and all cells of all six tables using case-insensitive whole-word or exact-phrase matching. The terms shown are a selection relevant to the action and evidence dimension, not an exhaustive list. [1]

Term or phrase	Occurrences in v5.0
agentic	4
tools / tool	2 / 1
APIs / API	1 / 0
reconstruct / reconstruction	0 / 0
log / logs / logging / audit trail	0 / 0 / 0 / 0

Structural checks:

- Exhibit C contains 15 numbered information requests. Field 10 asks "AI model type (automate, augment, support)." [1]
- Exhibit B Checklist item 3 runs from 3a through 3p. Narrative item 1g asks how the company assesses autonomy, reversibility, and reporting impact risk, while Narrative item 2b asks about human involvement for direct consumer impact. [1]
- Exhibit A columns 5 and 8 are headed "Generative/Agentic AI Models" and sit within the model counts for direct consumer impact and material financial impact. Exhibit A includes a Claims/Adjudication row. [1]
- Exhibit C field 3 asks for the model type used in the AI System and includes "generative/agentic AI" as an example. Exhibit C Company Instructions permit discussion with regulators where information is more appropriately provided at the AI System rather than AI Model level, or vice versa. [1]
- Exhibit A permits a company with an existing AI Systems inventory to suggest submitting that inventory in lieu of completing Exhibit A. [1]
- v5.0 contains a standalone Confidentiality section stating that requested information will be held confidential consistent with the requesting state's authority and directing regulators to cite relevant confidentiality protections. [1]
- The only occurrence of the word "retention" in v5.0 is "retention modeling" in an underwriting example; it is unrelated to records retention. The Model Bulletin, separately, includes "Data and record retention" within risk management and internal controls. [1][4]
- This comment does not rely on the change in Exhibit D naming between versions 4.0 and 5.0 as evidence and proposes no amendment to Exhibit D. [1][3]



Electronically Submitted to ssobel@naic.org and maromero@naic.org

September 29, 2026

TO: The NAIC Big Data and Artificial Intelligence (H) Working Group (the “Working Group”)

Re: AI Risk Evaluation Supplement (Version 5.0)

Dear Members of the Working Group:

On behalf of our members, the Insured Retirement Institute (IRI)¹ appreciates the opportunity to provide comments to the Working Group on version 5.0 of the AI Risk Evaluation Supplement (the “Supplement”). As we have noted in prior comments, our members do not oppose a tool that provides structure to a state regulator’s evaluation of an insurer’s use of artificial intelligence, provided that any such tool is appropriately tailored to address potential consumer harm, focused on key components from a risk perspective, and aligned with existing laws, regulations, and the Model Bulletin on the Use of Artificial Intelligence Systems by Insurers (the “Model Bulletin”). Our primary concern with this exposure is a significant scoping issue, discussed below. We also offer brief comments on materiality, Exhibit A, and Exhibits C and D.

The scope of the exhibits should be narrowed to exclude predictive models with transparent and explainable structures, including Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs).

The Supplement should focus on AI systems that involve more complex decision-making processes, or where it is more difficult to connect inputs to outputs. Traditional predictive models with transparent methodologies are well understood, have been in use across the industry for decades, and are already subject to well-established governance, actuarial, and regulatory oversight practices. Including these models within the scope of the exhibits would substantially expand reporting obligations without providing commensurate regulatory value.

GLMs, in particular, have been used in the industry for many years, and regulators have long reviewed them through existing market conduct and financial examination processes. Those

¹ The Insured Retirement Institute (IRI) is the leading association for the entire supply chain of insured retirement strategies, including life insurers, asset managers, and distributors such as broker-dealers, banks and marketing organizations. IRI members account for more than 95 percent of annuity assets in the U.S., include the top 10 distributors of annuities ranked by assets under management, and are represented by financial professionals serving millions of Americans. IRI champions retirement security for all through leadership in advocacy, awareness, research, and the advancement of digital solutions within a collaborative industry community.

models do not present characteristics such as adaptivity, or difficulty in tracing inputs to outputs, that motivate a new and AI-specific method of review. Applying a new AI-specific evaluation framework to models that regulators already examine effectively would generate significant reporting burden for insurers, and volume for regulators, without improving consumer protection.

This approach is also consistent with recent analysis by the European Insurance and Occupational Pensions Authority (EIOPA), who concluded that the risks associated with these models arise primarily from governance and business practices, rather than from the model architecture itself, and that those residual risks are already addressed through existing sectoral regulation and supervision. The same reasoning applies here: where a model's structure is transparent, the appropriate regulatory focus is the insurer's governance and business practices, which existing examination guidelines already reach.

Accordingly, IRI recommends that the Working Group narrow the scope of the exhibits to exclude predictive models with transparent and explainable structures. At minimum, we recommend that GLMs and GAMs be excluded from the scope of the Supplement.

Additional Comments:

- With respect to materiality, IRI members have concerns about creating regulator-specific materiality thresholds. Any discussion of materiality should instead be grounded in company-based thresholds tied to existing examination guidance and authority. We also recommend deleting the definition of materiality, as defining a term that the rest of the Supplement treats as flexible creates confusion.
- IRI supports removing "Marketing" from Exhibit A. Marketing should not be grouped with much higher-impact functions such as ratemaking and underwriting.
- IRI recommends that Exhibits C and D be removed or deferred until further piloting can be conducted. These exhibits are duplicative in some cases and are not relevant to financial or solvency risk.

We appreciate the Working Group's continued engagement with stakeholders on these issues, and we would welcome the opportunity to discuss these comments further. Thank you again for the opportunity to provide input.

Sincerely,

Sarah E. Wood

Sarah Wood
Director, State Policy & Regulatory Affairs
Insured Retirement Institute
swood@irionline.org

Recommendations to Strengthen the AI Risk Evaluation Supplement

Comments on version 5.0 from the regional-carrier practitioner perspective

Insurtech Advisors | September 2026

TO: NAIC Big Data and Artificial Intelligence (H) Working Group

FROM: Dr. Kaenan Hertz, Insurtech Advisors

RE: AI Risk Evaluation Supplement version 5.0

DATE: September 2026

I build AI governance programs for regional carriers, where small compliance teams often rely heavily on vendor-supplied AI. I worked through the pilot tool as a regional carrier would, which gives these comments a practitioner basis. The recommendations below address version 5.0 of the AI Risk Evaluation Supplement and focus only on gaps that remain before adoption.

The Supplement gets the core structure right: risk-based scoping, regulator customization, company-specific materiality, and closer attention to third-party models. The remaining issue is answerability. A policy is not a control if a carrier cannot support its response with evidence it can realistically obtain.

What version 5.0 already improves

- It reframes the document as a supplement to the Market Regulation Handbook, Financial Condition Examiners Handbook, and Financial Analysis Handbook, rather than a stand-alone examination standard.
- It adds materiality guidance tied to the Financial Condition Examiners Handbook and allows the regulator to specify a threshold or the company to disclose and support its own.
- It explicitly defines generalized linear models (GLMs), adds GLMs as an Exhibit C model type, and gives regulators room to narrow later inquiries.
- It expands third-party-model oversight, explainability, and transparency questions, and broadens the SDLC question to include the wider AIS Program.

These changes should be preserved. They also allow the Working Group to focus the next revision on six operational questions.

Recommendations

1. Add a company-facing proportionality path for resource-constrained carriers

The pilot summary says states will apply proportionality, and version 5.0 tells regulators to customize the inquiry for a limited-scope examination. Add a short company-facing example showing what a proportionate response looks like for a carrier with a small AIS inventory, limited direct consumer impact, and no dedicated model-risk function. The example should identify the minimum evidence expected, when a narrative response is enough, and what facts trigger a deeper Exhibit C request.

2. Define a consistent unit for counting AI systems and AI models

Exhibit A will not produce comparable information unless respondents count the same unit. The Supplement now distinguishes AI Systems from AI Models, but practical boundaries remain unclear. Add examples for:

- an embedded fraud flag inside a claims platform;
- a general-purpose assistant deployed across several business functions;
- multiple models operating within one vendor product; and
- a model version change that does or does not create a new reportable unit.

A counting rule will improve both scoping and cross-state consistency without dictating a carrier's internal architecture.

3. Create a reusable third-party evidence package and a direct regulator-submission path

Version 5.0 appropriately asks whether a company can access and test purchased models and how it governs, monitors, tests, and ensures transparency for vendor models. Regional carriers may still be unable to obtain the detail requested because vendors restrict model information as confidential or proprietary.

The Working Group should publish a standard vendor evidence package covering intended use, validation, performance monitoring, fairness testing, explainability, change control, data categories, human oversight, and the carrier's contractual access rights. When specific detail cannot be given to the carrier, the Supplement should provide a uniform path for the vendor to submit it directly to the regulator under the confidentiality authority applicable to the examination or inquiry.

This proposal does not ask the Supplement to create new confidentiality rights. It asks the Supplement to standardize how existing authority is used so a vendor can respond once and a carrier can point to a complete regulatory submission. This problem has a working precedent: credit bureaus submit their score calculation methodologies to regulators under confidentiality, protecting the intellectual property while giving the regulator what it needs.

4. Address enterprise generative AI and shadow AI as operating risks

Version 5.0 adds definitions for generative and agentic AI, but the exhibits remain oriented toward identified models selected for regulatory review. Add targeted questions asking how a company:

- discovers unsanctioned AI use;
- prevents nonpublic personal information from entering unapproved tools;
- governs enterprise assistants that support many low- and medium-risk uses; and
- decides when a general-purpose tool becomes a high-risk use because of the workflow in which it is deployed.

This would capture a major present-day exposure without turning every employee productivity tool into a full Exhibit C response.

5. Clarify the boundary for AI used by professional-service providers

Exhibit B asks about material AI Systems used by third-party providers, including actuarial, claim, MGA, audit, and other professional services. Clarify when an insurer is expected to diligence AI used by a provider on the insurer's behalf and what evidence is proportionate when the provider controls the tool. A useful boundary would turn on whether the provider's AI can materially influence a regulated insurance decision, consumer communication, financial statement, or control assessment.

The same guidance should explain how the Supplement applies to independent agents and MGAs that participate in the workflow but do not operate as the insurer's internal business unit.

6. Clarify explainability evidence for consumer-impacting decisions

The Model Bulletin requires insurers to govern AI-supported consumer decisions so that they comply with applicable law, and version 5.0 asks how a company ensures explainability and transparency. The Supplement should clarify what evidence is sufficient to reconstruct an individual consumer-impacting decision: the inputs used, model or rule version, output, human review, and the reasons communicated when applicable law requires them.

The expectation should be technology-neutral and tied to applicable law, rather than imposing a universal reason-code format through the Supplement.

Targeted drafting clarifications

- Exhibit D, row 20: define "Reasonable Accommodations or Policy Modifications (granted or requested)" and add one line-of-business example.
- Previously submitted material: state a presumption that a current response may be incorporated by reference when the regulator can obtain it under existing handbook coordination procedures, subject to a request for updates or jurisdiction-specific information.

- Exhibit C vendor access: distinguish evidence the carrier must possess from evidence a vendor may provide directly to the regulator.
- Implementation guidance: retain the current flexibility for regulators to narrow requests after Exhibit A, but give respondents a short scoping worksheet so the basis for that narrowing is visible.

Conclusion

These recommendations come from trying to answer the Supplement as a regional carrier would, while recognizing the substantial changes already made in version 5.0. They would improve comparability, reduce duplicative requests, and give regulators better evidence about the AI uses most likely to create consumer or financial harm.

I appreciate the Working Group's open process and would be glad to discuss these recommendations or help test the vendor evidence package against information regional carriers can realistically obtain.

Respectfully submitted,
 Dr. Kaenan Hertz
 Insurtech Advisors
info@insurtechadvisors.com
 +1.516.218.1240

Appendix A: submission context

Current exposure: AI Risk Evaluation Supplement, version 5.0.

Pilot states: California, Colorado, Connecticut, Florida, Iowa, Louisiana, Maryland, Pennsylvania, Rhode Island, Vermont, Virginia, and Wisconsin.

Pilot period: March through September 2026. NAIC materials caution that individual state pilots may finish on different schedules and may not all be completed by September 30.

Public-comment deadline: Close of business Tuesday, September 29, 2026, according to the current NAIC exposure page. Comments are directed to Scott Sobel or Miguel Romero.

Current timeline: Version 5.0: 30-day September exposure. Version 6.0: planned 14-day October exposure. Version 7.0: planned consideration for adoption at the November 2026 Fall National Meeting.

Appendix B: proposed third-party AI evidence package

Purpose: a vendor completes one evidence package per model version. The carrier may attach the information it is permitted to retain and identify any regulator-direct submission. The regulator determines confidentiality and production treatment under applicable authority.

A. Vendor and model identification

- Vendor legal name and regulatory contact.
- Model/product name, version, and version date.
- Deployment model and approved insurance uses.
- Attestation date and covered period.

B. Intended use and constraints

- Approved and prohibited uses.
- Required inputs, known limitations, and intended users.
- Consumer, financial-reporting, or control impacts.

C. Validation and performance

- Most recent validation date and validator independence.
- Methods, metrics, thresholds, and summary results.
- Monitoring cadence, drift detection, and remediation triggers.

D. Fairness and compliance

- Tests performed, population or proxy attributes considered, thresholds, results, and remediation.
- Applicable legal or regulatory requirements mapped to controls.

E. Explainability

- Decision reconstruction evidence available to the carrier.
- Reason information available when required by applicable law.
- Limits on explanations and known failure modes.

F. Change control

- Definition of a material change.
- Version control and advance-notice commitment.
- Retesting obligations after changes.

G. Data

- Training, test, and operating data categories.
- Data provenance, retention, deletion, and nonpublic-personal-information controls.

H. Human oversight and authority

- Required human review and override.
- Tool, communication, payment, or decision authority granted to the model or agent.
- Kill switch and incident escalation.

I. Attestation

- Vendor officer attests that the package is complete and accurate to the vendor's knowledge and will be updated after a material change.

Appendix C: proposed handling categories for proprietary model information

These categories are a workflow proposal only. They do not create or limit confidentiality, privilege, trade-secret protection, or a regulator's production authority.

1. Category 1 - Carrier-held: information the carrier needs to govern the model and answer the Supplement.
2. Category 2 - Regulator-direct: additional validation, fairness, training-data, or methodology detail that the vendor submits directly under applicable confidentiality authority, with a summary supplied to the carrier.
3. Category 3 - Escalation required: information the vendor withholds. The vendor identifies the legal and factual basis for withholding, and the regulator determines whether a summary is enough or whether further production is required.

A carrier should not be deemed to have validated a model merely because a vendor made a confidential submission. The carrier must still hold enough information to govern use, monitor performance, respond to incidents, and meet its own legal duties.

COMMENT PAPER

From Model Review to Enterprise Solvency: Closing the Aggregation, Recovery-Rights, Data- Governance, and Regulatory-Asymmetry Gaps in the NAIC AI Risk Evaluation Supplement v5.0

A supervisory-clarification proposal for AI, agentic AI, ORSA, Digital Premises Liability, adversarial data intelligence, state-law asymmetry, third-party opacity, data lineage, and subrogation impairment

The Marie Curie Warning

Marie Curie helped reveal radioactivity before the scientific and occupational-control systems surrounding radiation were mature. Her death from aplastic anemia was likely caused by years of radiation exposure. The analogy for insurance supervision is not that AI is radiation; it is that useful capability can become economically indispensable before society has adequately characterized cumulative exposure, dose, concentration, or control failure. Insurance supervision should not wait for the equivalent of mature radiation epidemiology before asking where AI exposure accumulates and how it can threaten policyholders, recoveries, reserves, and capital.

**Prepared by Gerald (Gerry) Kennedy of
Observatory Strategic Management**

For consideration by the NAIC Big Data and Artificial Intelligence (H) Working Group
September 2, 2026

Formal comment deadline: close of business Tuesday, September 29, 2026

Executive Abstract

The NAIC has now formally exposed AI Risk Evaluation Supplement version 5.0 for a 30-day comment period ending September 29, 2026. The exposure is procedurally important because v5.0 moves the supervisory conversation beyond consumer-facing model review toward concepts that can support a more complete enterprise-risk inquiry: material financial impact, inherent risk before controls, third-party oversight, agentic AI, model inventories, model limitations, explainability, and data dictionaries. Those additions create a timely opportunity to close a second-order gap: the Supplement should expressly require regulators and insurers to evaluate AI not only model-by-model, but also as an aggregated enterprise dependency capable of altering recovery rights, evidentiary integrity, causation analysis, cross-line loss correlation, and prospective capital needs.

This paper recommends that the NAIC use the v5.0 consultation to add explicit examination language in seven areas: (1) aggregate common AI systems, model families, vendors, cloud providers, data sources, and failure modes across legal entities and insurance lines; (2) connect financially material AI dependencies and concentrations to ORSA risk categories, prospective capital, liquidity, reserves, and reinsurance; (3) evaluate AI-related impairment of salvage, subrogation, evidentiary integrity, causation analysis, and recovery rights; (4) prevent vendor opacity, contractual restrictions, or inability to test a third-party system from reducing insurer accountability; (5) measure inherent exposure before crediting human review, insurance, contracts, or other controls; (6) evaluate data not only by privacy sensitivity but by adversary utility, AI parseability, operational consequence, intelligence half-life, and aggregation potential; and (7) identify regulatory asymmetries created when state privacy statutes impose activity-based obligations on much of the data economy while categorically exempting insurance institutions or insurance-support organizations whose data activities may create comparable or greater enterprise exposures.

The core supervisory proposition is straightforward: a model can be individually well-governed while the enterprise remains systemically exposed. A common foundational model, cloud dependency, data source, or vendor can produce correlated underwriting, claims, fraud, legal, operational, and recovery-rights failures across otherwise separate portfolios. In a multiline insurer, that is not merely model risk. It can become solvency risk.

A further supervisory signal now comes from outside insurance regulation. New York Unified Court System Part 161, effective June 1, 2026, preserves the accountability of attorneys who use AI in court submissions and, through its model rule, requires careful independent review for fabricated or fictitious material. In May 2026, the Munich I Regional Court separately treated false statements generated by Google AI Overviews as statements attributable to Google rather than as merely third-party search content. Together, these developments illustrate an emerging accountability principle relevant to insurance supervision: deployment of AI does not extinguish the responsibility of the professional or enterprise that selects, configures, relies upon, or communicates the output. For purposes of v5.0, asserted human review should therefore be evaluated as an evidenced control, not accepted as a label.

Requested NAIC action

Adopt a cross-enterprise “AI Aggregation and Recovery-Rights” examination layer within v5.0, or direct examiners to obtain equivalent information through existing handbooks. The supervisory sequence should be: identify gross/inherent exposure; aggregate common dependencies and failure modes; test legal and operational recoverability; then evaluate controls and residual risk; finally, determine whether the remaining exposure is material to ORSA, reserves, reinsurance, liquidity, or prospective capital.

1. Procedural Posture and Why v5.0 Matters

The NAIC Big Data and Artificial Intelligence (H) Working Group states that, following its August 31, 2026 meeting, it is exposing the AI Risk Evaluation Supplement version 5.0 for written comments and feedback for a 30-day period ending September 29, 2026. Comments are directed to Scott Sobel and Miguel Romero. The NAIC's published Summary of Changes identifies several revisions that are highly relevant to financial-condition supervision: materiality aligned to the Financial Condition Examiners Handbook; a clarified concept of inherent risk as risk considered before mitigating controls; separate attention to models with Direct Consumer Impact and models with Material Financial Impact; a Model Inventory; new third-party-model oversight questions; model-risk and model-limitation fields; data-use and data-dictionary enhancements; and a definition of agentic AI.

Those revisions are constructive. They also expose the point at which model-centric supervision should connect to enterprise risk. The Supplement now has the vocabulary needed to ask whether a model is financially material. The next question is whether multiple individually non-material AI uses share a common dependency or failure mode that becomes material in aggregate. Financial-condition supervision cannot stop at the unit of analysis chosen by the technology team if the economic exposure is larger than that unit.

v5.0 development	Supervisory opportunity created
Inherent risk before mitigation	Require gross exposure measurement before contracts, human review, insurance, warranties, or other controls are credited.
Material Financial Impact	Map material AI dependencies to ORSA, reserves, liquidity, reinsurance, and prospective capital.
Model Inventory	Extend inventory to common model families, data sources, cloud/providers, agents, tools, and shared failure modes.
Third-party model oversight	Clarify that opacity or inability to test is an exposure, not a risk-reduction factor.
Agentic AI definition	Examine delegated authority, tool permissions, autonomous actions, and machine-speed accumulation.
Data dictionary / data-use fields	Add provenance, original purpose, recursive use, downstream propagation, evidence integrity, and operational/OT sensitivity.
Explainability and transparency	Connect explainability to causation, coverage, claims defensibility, and subrogation.

2. The Marie Curie Analogy: Capability Can Mature Before Exposure Science

Marie Curie's work transformed science and medicine. She discovered polonium and radium with Pierre Curie and became the only person to receive Nobel Prizes in two different scientific disciplines. Yet the Curies worked at a time when the harmful effects of radioactivity were not fully understood and precautions were rudimentary. Nobel Prize materials state that Marie Curie died from aplastic anemia likely caused by radiation exposure; the Bibliothèque nationale de France explains that radioactive contamination remains in portions of the Curies' archival materials and that later generations imposed increasingly sophisticated protective controls as knowledge evolved.

The analogy is deliberately limited. AI is not ionizing radiation, and current AI risks should not be asserted to have the same causal certainty as radiation injury. The governance lesson, however, is directly relevant: society can recognize extraordinary utility before it understands cumulative exposure. Early absence of complete causal knowledge is not evidence of zero exposure. A responsible supervisory regime therefore measures concentration, duration, dependency, and consequence while the science and loss experience continue to mature.

Supervisory principle

Do not make “we have not yet observed the catastrophe” equivalent to “there is no material exposure.” The appropriate insurance response to an emerging peril is controlled observation, exposure measurement, stress testing, and capital-aware governance and not retrospective recognition after correlation has already reached the balance sheet.

3. From Model Risk to Cognitive Concentration Risk

Insurance supervision has long distinguished between a single risk and an accumulation. AI requires the same discipline. A proprietary LLM may appear to be one technology asset, yet it can simultaneously influence underwriting, claims, special investigations, risk engineering, legal research, premium audit, customer service, actuarial analysis, reserve support, and executive decision-making. The common model, data layer, cloud platform, identity system, tool framework, or vendor may therefore act as a correlation mechanism across functions that traditional organization charts treat as separate.

This paper uses the term Enterprise AI Cognitive Concentration Risk to describe the possibility that common AI reasoning infrastructure or shared AI dependencies produce correlated decision errors, operational failures, legal exposures, or recovery impairment across multiple insurance functions, legal entities, lines of business, policy years, or jurisdictions. The concept does not require the AI to be “bad,” autonomous in a science-fiction sense, or even consumer-facing. Concentration can arise simply because many material processes depend on the same system or supply chain.

The NAIC’s own 2026 insurance-topic materials recognize widespread AI use and continued regulatory focus on third-party systems, data quality, cybersecurity, explainability, and insurer accountability. A 2026 Journal of Insurance Regulation article also notes that insurers are using or considering AI in claims functions including subrogation and detection of first- and third-party liability. These facts make the recovery-rights dimension especially important: AI is not only changing loss generation; it is already participating in the mechanisms used to identify responsible parties and recover paid losses.

4. Five Supervisory Clarifications Recommended for v5.0**4.1 Aggregate Common AI Dependencies Across Entities and Lines**

The Model Inventory should be supplemented by an Enterprise Dependency Inventory. Regulators should be able to identify common model families, model versions, cloud providers, model hosts, vector stores, data vendors, identity providers, retrieval systems, agent frameworks, fine-tuning pipelines, synthetic-data generators, and critical tools used across multiple functions or entities.

The examination objective is correlation identification. A model used in two lines may not be material within either line, but the same failure mode may be material to the group. Similarly, a third-party foundation model may be hidden behind multiple vendor products. The insurer should demonstrate that it can look through product names to identify common upstream dependencies.

- Require a group-wide map of common AI systems, model families, vendors, cloud/service providers, data sources, and agent/tool dependencies.
- Require identification of single points of cognitive or operational failure and the business functions they support.
- Require an aggregation view across underwriting, claims, operations, legal, finance, risk engineering, distribution, and customer servicing.

- Extend the scope beyond the carrier to material insurance supply-chain participants: agencies, brokerages, wholesalers, MGAs/MGUs, program administrators, TPAs, independent adjusters, claims-service organizations, loss-control and engineering firms, reinsurers, model vendors, cloud providers, and other critical service providers where their AI/data dependencies can propagate into the insurer's risk profile.

4.2 Connect Material AI Dependencies to ORSA and Prospective Capital

Where an AI system, model family, or common dependency has Material Financial Impact or where aggregated exposure is financially material, the Supplement should direct examiners to determine whether the exposure has been mapped into the insurer's enterprise risk management and ORSA processes. The objective is not to impose an automatic AI capital charge. It is to ensure that financially material AI dependency is assessed with the same prospective discipline applied to other emerging enterprise risks.

The ORSA Implementation (E) Subgroup's 2026 charge includes ongoing review of the Model Act and ORSA Guidance Manual. The v5.0 Supplement can therefore function as an upstream discovery mechanism: it identifies AI concentrations; ORSA determines whether those concentrations could threaten capital adequacy under normal and stressed conditions.

- Require severe-but-plausible scenarios for common-model failure, vendor outage, model corruption, data poisoning, agentic misuse, regulatory prohibition, or loss of access to a critical model/provider.
- Require scenario consideration of correlated underwriting error, claims error, litigation defense cost, restitution, market-conduct remediation, operational interruption, model reconstruction, and reputational lapse or retention effects.
- Require consideration of reinsurance responsiveness, exclusions, aggregation clauses, and potential disputes regarding occurrence and causation.
- Require management to identify whether prospective capital, liquidity, reserves, or risk limits need adjustment when residual exposure remains material.

4.3 Examine Salvage, Subrogation, Evidentiary Integrity, Causation, and Recovery Rights

AI risk can impair both sides of the insurance equation: it can contribute to loss and it can weaken the insurer's ability to recover after paying the loss. This is a supervisory blind spot if AI examination focuses only on model accuracy or adverse consumer outcomes.

Subrogation economics are often influenced by technology contracts containing arbitration clauses, indemnity provisions, liability caps, consequential-damage exclusions, class waivers, forum provisions, confidentiality duties, audit restrictions, or limitations on access to logs and technical evidence. A carrier may underwrite an insured expecting ordinary recovery rights, only to discover after loss that a critical vendor contract has materially constrained the recovery path. Observatory's August 2026 NAIC warning describes this as an accumulation problem when common technology contracts recur across large policyholder populations and multiple insurance lines.

Agentic and generative AI add further complications. AI-generated records, synthetic communications, mutable prompts, incomplete logs, retrieval-layer changes, and model updates can make causation reconstruction more difficult. If an insurer cannot reconstruct what model version acted, what data it received, what tool it invoked, or what evidence it relied upon, the result may impair coverage analysis, product-liability allocation, contribution, salvage, and subrogation.

- Ask whether AI-generated or AI-mediated evidence is preserved in a manner sufficient to reconstruct material decisions and loss causation.
- Ask whether contracts preserve the insurer's and regulator's rights to obtain logs, model/version information, incident artifacts, audit evidence, and technical cooperation after loss.

- Ask whether policyholder technology contracts or insurer vendor contracts can materially impair subrogation or contribution rights.
- Ask whether AI use can impair salvage value, chain of custody, provenance, authenticity, or admissibility of evidence.
- Require financially material recovery impairment to be reflected in pricing, reserving, reinsurance, and ORSA assumptions where appropriate.

4.4 Vendor Opacity Must Not Reduce Insurer Accountability

The NAIC Model Bulletin already states that insurers should address due diligence for third-party AI systems and data, seek appropriate contractual audit rights, require cooperation with regulatory inquiries, and perform contractual audit or other confirmation activities. The Bulletin further states that an insurer's use of AI remains subject to regulatory examination. The Third-Party Data and Models Working Group continues to address transparency, governance, and accountability for outside vendors.

The v5.0 Supplement should state the corollary explicitly: inability to obtain information from a vendor, trade-secret restrictions, model-provider opacity, contractual limitations, or practical inability to test a third-party model are themselves components of inherent risk. They should not be accepted as reasons to lower the insurer's accountability or reduce the examination scope. If the insurer cannot independently validate a material third-party dependency, the examination should capture that uncertainty as risk.

Proposed rule of interpretation

Opacity is not mitigation. A contractual promise is not a control until its practical enforceability, auditability, incident cooperation, and recovery value are demonstrated. Human review is not mitigation unless the human has the competence, authority, time, evidence, and independence necessary to challenge the system.

4.5 Measure Inherent Exposure Before Crediting Controls

The v5.0 Summary of Changes expressly emphasizes inherent risk before mitigating controls. This principle should be operationalized. Examiners should first establish the gross exposure created by the AI use case and its aggregation characteristics; only then should they consider governance, human review, contracts, insurance, indemnity, diversification, circuit breakers, or other controls.

Step	Required supervisory question
1. Gross / inherent exposure	If no mitigation existed, what could this AI system, common dependency, or failure mode affect?
2. Aggregation	Which entities, lines, processes, policyholders, vendors, and policy years share the dependency?
3. Recoverability	Would evidence, contracts, salvage, contribution, or subrogation remain available after loss?
4. Control effectiveness	Are human, technical, contractual, and governance controls demonstrably operative—not merely documented?
5. Residual exposure	What material risk remains after tested controls?
6. Financial-condition mapping	Does residual exposure affect reserves, reinsurance, liquidity, ORSA scenarios, risk appetite, or prospective capital?

5. Data Is Not Merely an Input: It Is an Exposure, a Commodity, and an Evidence Chain

The Supplement's expanded attention to data use and data dictionaries is an opportunity to move beyond conventional PII classification. Insurers accumulate vast quantities of operational and strategic intelligence through applications, underwriting files, loss-control surveys, engineering inspections, claims files, recorded statements, audits, photographs, financials, telematics, vendor data, and policyholder technology systems. Some of that information may be personally sensitive; some may be strategically sensitive because it reveals how physical systems operate, where controls are weak, or how losses propagate.

Digital Premises Liability a term coined by Observatory, is used here as an analytical framework: the duty of care should not be conceptualized only as preventing disclosure of personal information. It should include stewardship of accumulated operational intelligence where exfiltration, aggregation, AI-enabled parsing, or unauthorized secondary use can create downstream physical, economic, product-liability, or systemic losses. This is particularly relevant when compromised data can be mined to identify OT architectures, engineering deficiencies, business-interruption dependencies, or critical-system vulnerabilities.

For examination purposes, the practical questions are provenance, purpose, authority, transformation, propagation, retention, and downstream consequence. A data dictionary that names a field is useful; a supervisory data-lineage record that explains why the data was collected, how it was transformed, where it propagated, which models depend on it, and what happens if it is corrupted or exfiltrated is more valuable.

- Original source and purpose of collection.
- Legal/contractual authority and restrictions on secondary use.
- Transformations: embeddings, vectorization, enrichment, labeling, synthetic generation, fine-tuning, distillation, or feature derivation.
- Propagation into models, agents, third parties, backups, and downstream systems.
- Operational/strategic sensitivity, including OT/ICS, engineering, safety, catastrophe, and recovery-related information, not only PII.
- Retention, deletion, recoverability, and forensic reconstruction.
- Use of model outputs as future training inputs and governance of recursive or synthetic-data loops.

6. TravelersLLM as a Public Governance Case Study, Not an Allegation

Travelers publicly announced on June 30, 2026 that it developed TravelersLLM, a proprietary large language model tailored to its property-casualty business. Travelers states that the model was trained on millions of company documents and is intended to enhance underwriting analysis, accelerate research and model development, facilitate access to decades of institutional knowledge, improve workflows, and support its broader AI strategy.

That announcement should not be read as evidence that TravelersLLM is unsafe or noncompliant. It is valuable to supervisors because it illustrates the type of enterprise deployment that the v5.0 Supplement must be capable of evaluating. A proprietary model built on millions of insurance documents may concentrate decades of institutional judgment in a common reasoning layer. If the model or its successors influence multiple business functions, the appropriate supervisory inquiry extends beyond accuracy on individual questions.

Gerry Kennedy's July 2026 Travelers POV argues that insurance files contain not only facts but also interpretations, omissions, heuristics, historical practices, and institutional assumptions. It raises a governance question: does an enterprise LLM learn "insurance," or does it learn how a particular institution historically interpreted insurance? The POV further argues that if model outputs later become training

inputs, latent patterns can become self-reinforcing and the common model can become a concentration of decision authority. These are hypotheses to examine, not conclusions about Travelers.

Why the example matters to NAIC v5.0

A regulator should be able to determine (1) what common functions depend on the model; (2) what data and model lineage created it; (3) how its outputs are independently validated; (4) whether recursive/synthetic training is used; (5) how third-party components are governed; (6) whether a common failure could affect multiple lines or legal entities; and (7) whether the residual exposure is represented in ORSA and capital planning.

7. Human Review Is Not Automatically a Mitigating Control

A recurring assumption in AI governance is that a “human in the loop” reduces risk. That may be true, but only when the human performs a real control function. If AI becomes the primary interpreter of underwriting files, claims evidence, engineering reports, or legal materials, the human reviewer may become increasingly dependent on the system being reviewed. The result can be automation bias, reduced independent verification, or skill atrophy.

For supervisory purposes, human review should therefore be treated as a control that must be tested, not a label that automatically reduces inherent risk. Examiners should ask whether reviewers periodically demonstrate independent competence; whether they see primary evidence rather than only AI summaries; whether challenge and override rates are monitored; whether time and incentives permit meaningful review; and whether the organization tests performance when the AI is unavailable.

This is another application of the Marie Curie analogy: exposure can be cumulative. An organization may not notice declining independent human capability on any single decision. It may discover the dependency only when the AI is wrong, unavailable, compromised, or prohibited and the human layer expected to catch the error is no longer functioning as an independent control.

7A. Responsibility Cannot Be Delegated: Courts Are Converting AI Governance into a Standard-of-Care Question

Gerry Kennedy’s July 2026 article, “The Hidden Insurance Consequences of New York’s AI Court Rule. What Is in YOUR POLICY?”, identifies a development that should be made explicit in the NAIC examination framework: AI is changing not only how professional work is performed, but also how responsibility for that work will be reconstructed after a loss. The article’s central insurance proposition is that the inquiry following an AI-assisted professional error will not end with whether the professional made a mistake. It will extend to the AI platform used, independent verification, confidentiality controls, governance policy, supervision, and whether the professional standard of care was satisfied.

7A.1 New York Part 161: The Human Signature Remains the Accountability Point

New York’s Part 161 does not prohibit AI-assisted preparation of court papers and does not impose a blanket AI-use disclosure requirement. Instead, it preserves pre-existing duties and responsibilities regardless of whether AI was used. Appendix A goes further where adopted by a court: the user is expected to understand the tool’s capabilities and limitations, carefully review AI-assisted work, independently ensure that the filing contains no fabricated or fictitious authorities or material, and accept sanctions or other remedial consequences for failure.

For insurance supervision, this matters because “human in the loop” can become an empty control if the human is unable, unqualified, insufficiently resourced, cognitively dependent on the tool, or procedurally discouraged from challenging it. A signature, click, or nominal approval should not be credited as risk mitigation unless the insurer can demonstrate what was independently checked, which sources were consulted, what discrepancies were identified, what changes were made, and what authority existed to reject the AI recommendation.

7A.2 The Munich AI-Overview Ruling: Synthesis Can Create Direct Attribution

On May 28, 2026, the Munich I Regional Court ruled that Google could be directly responsible for incorrect statements generated by its AI Overview feature. The court distinguished conventional search results from generative synthesis: AI Overviews reformulated, structured, and directly answered user questions in Google’s own system, rather than merely providing links to third-party material. The Law Library of Congress summarizes the court’s reasoning as treating the generated statements as directly attributable to Google because Google developed and operated the system and controlled its algorithms. The decision is not a universal rule of U.S. insurance law, but it is an important comparative-law signal: where an enterprise transforms source material into new machine-generated assertions, courts may look to the operator rather than permit responsibility to disappear into the model.

7A.3 The Professional-Liability Feedback Loop

This creates a direct connection between AI governance and E&O/professional liability. When AI participates in coverage analysis, legal strategy, claim evaluation, medical review, engineering, actuarial work, fraud referral, or loss-control advice, the post-loss inquiry can become a verification-trail inquiry. A claimant, regulator, reinsurer, or subrogating carrier may ask: Which model and version was used? What data entered the prompt or retrieval system? What sources supported the output? What was independently verified? Who had authority to override the model? Was privileged or confidential information exposed to a third party? Did model terms restrict forensic access, auditability, indemnity, or recovery? Did the insurer’s own professional-liability or cyber program contemplate AI-assisted work?

The supervisory implication is important: an AI control should not be evaluated solely by model accuracy. It should be evaluated by the defensibility of the decision process after a claim. That requires provenance, source verification, versioning, logs, exception handling, escalation records, and evidence that professional judgment remained effective rather than ceremonial.

Supervisory proposition: Responsibility follows deployment. AI may alter the method of performance, but it should not be permitted to create an accountability vacuum between model developer, insurer, professional user, vendor, and policyholder.

7B. The Threat Is Already in the Wild: AI-Enabled Data Beneficiation Changes Breach Severity

The supervisory framework should evaluate artificial intelligence from both sides of the loss equation. Insurers are deploying AI internally, but adversaries can also use AI to interrogate, classify, correlate, and operationalize data already exfiltrated into uncontrolled environments. Gerry Kennedy’s August 5, 2026, article, “Mining for Ferrous or Non-Ferrous Data: What Could Happen When a ‘Data Magnet is Applied’,” frames this as commodity beneficiation: the breach is the extraction event, the stolen corpus is the ore body, and AI is the refinery that separates strategically valuable operational intelligence from high-volume conventional records.

7B.1 Ferrous and Non-Ferrous Data: A Supervisory Taxonomy

Traditional breach response is heavily oriented toward high-volume personally identifiable information: Social Security numbers, dates of birth, addresses, credentials, payment information, and similar records. Those exposures remain serious. For supervisory purposes, however, this paper uses the analogy “Ferrous Data” for abundant, readily identifiable, identity-focused information whose principal harms are traditionally associated with privacy, fraud, account takeover, and notification obligations.

“Non-Ferrous Data” describes lower-volume but potentially higher-consequence operational and strategic information: loss-control reports, engineering surveys, inspection photographs, catastrophe-model information, business-continuity plans, electrical one-line diagrams, fire-protection deficiencies, OT/ICS documentation, PLC configurations, SCADA architecture, building-management systems, data-center cooling and power design, critical-vendor dependencies, claims investigations, expert reports, and similar materials. The value of these categories is adversary-dependent. A fraud ring may prize identity data; a nation-state, ransomware operator, or sophisticated criminal group may derive substantially greater strategic value from operational intelligence.

The Supplement should therefore avoid equating sensitivity with privacy classification. A document can contain little regulated PII yet still be capable of materially increasing the probability or severity of bodily injury, property damage, business interruption, product failure, infrastructure disruption, or systemic loss.

7B.2 The “Data Magnet”: AI Changes the Economics of Stolen Corpora

Before modern AI, the cost of manually reviewing millions of heterogeneous stolen documents constrained adversarial exploitation. LLMs, embeddings, semantic search, entity extraction, knowledge graphs, vector databases, retrieval systems, and agentic workflows can reduce that constraint. They can identify relationships among facilities, vendors, engineers, technologies, deficiencies, and operational dependencies that may never appear together in a single document. The resulting intelligence can be created by inference and correlation rather than direct disclosure.

This changes the cyber kill chain. A breach may progress from steal -> parse -> classify -> enrich -> correlate -> target -> weaponize -> automate -> scale. The insured event is no longer adequately described by “records exposed.” Record count measures quantity removed; it does not measure the intelligence that can be manufactured from the corpus.

7B.3 Intelligence Half-Life and Recursive Criminal Capital Formation

Some compromised information depreciates quickly: credentials can be reset, payment cards replaced, and access tokens expired. Operational intelligence may have a much longer useful life. A building layout, industrial process, critical supplier relationship, historical engineering deficiency, OT architecture, or infrastructure dependency can remain useful for years. Its adversarial value may increase when later breaches, public sources, or improved analytical tools make correlation easier.

The resulting accumulation is recursive. Dataset A can improve targeting of Organization B; a subsequent compromise of B produces Dataset B; A and B together improve targeting of Organization C. Each loss can therefore add to a persistent secondary intelligence economy that improves the capability of future attacks.

This is an ORSA concern because the adversary’s information base can compound across insureds, lines, and years while the original insurer may continue to model each breach as an isolated event.

7B.4 The Insurance Supply Chain Is Also an Intelligence Supply Chain

The exposure does not reside only at the carrier. Policyholder information flows through agencies, retail and wholesale brokerages, MGAs/MGUs, program administrators, TPAs, independent adjusters, claims-service organizations, loss-control and risk-engineering firms, experts, counsel, reinsurers, modeling companies,

cloud providers, AI vendors, and other service organizations. Different participants often possess different fragments of the same risk. When those fragments are aggregated or later correlated after separate breaches, they can reconstruct a digital representation of the insured's physical and operational premises.

A statutory or regulatory exemption applicable to one participant does not eliminate the enterprise risk created by the broader chain. Examiners should therefore identify which parties possess high-consequence operational intelligence, what downstream rights and restrictions govern that data, how long it is retained, whether it is incorporated into AI systems, and whether contractual terms preserve incident cooperation, evidence access, audit rights, and recovery rights.

7B.5 Digital Bailment, Product-Enterprise Liability, and the Subrogation Feedback Loop

The data-custody problem should not be treated solely as privacy liability. Where an insurer or insurance-service entity receives operational information for underwriting, claims, engineering, loss control, or safety purposes, the relationship can resemble a digital bailment: valuable information is entrusted for a defined purpose and is expected to be safeguarded and faithfully stewarded. If that stewardship fails and the data later enables exploitation of a product or physical system, the original custody failure can intersect with product liability, negligent undertaking, agency, apparent-manufacturer, manufacturer-representative, component-supplier, professional-liability, and contractual theories depending on jurisdiction and facts.

Accordingly, a data custodian should not be presumed to stand outside the product-liability chain merely because it did not physically fabricate the ultimate product. If it functions as a manufacturer's representative, supplies an essential informational component, controls or validates a safety function, induces reliance, distributes AI-derived recommendations, or otherwise becomes an integral participant in the product or service enterprise, direct or vicarious liability pathways may arise. The examination objective is not to adjudicate those theories in advance; it is to determine whether the insurer has preserved the evidence and recovery rights necessary to pursue or defend them after loss.

This creates a feedback loop: entrusted data -> aggregation -> exfiltration -> AI-enabled beneficitation -> operational intelligence extraction -> exploitation of a product/system -> physical or economic loss -> insurance payment -> causation investigation -> subrogation/product-liability recovery. If logs, provenance, contracts, audit access, model lineage, or vendor cooperation are missing, the insurer may pay the loss and then discover that its recovery rights were impaired by the very digital-governance failure that helped create the loss.

7B.6 Proposed Supervisory Principle: Data Severity Must Include Adversary Utility

For v5.0 examination purposes, data risk should be evaluated across at least nine dimensions: (1) privacy sensitivity; (2) adversary utility; (3) correlation potential; (4) operational consequence; (5) intelligence half-life; (6) AI parseability; (7) aggregation potential; (8) recovery-rights significance; and (9) systemic/ORSA significance. This would allow examiners to distinguish a high-volume privacy event from a lower-volume release of strategically important operational intelligence whose downstream consequences could be substantially larger.

Supervisory proposition: The NAIC should assume that machine-scale parsing and correlation of stolen heterogeneous data is already economically accessible. The relevant question is not whether adversaries may someday acquire the capability; it is whether insurers have governed their own data custody, supply chains, incident response, recovery rights, and ORSA assumptions in light of that capability.

7C. State-Law Carve-Outs and Regulatory Asymmetry: Exemption Is Not Risk Mitigation

New Jersey P.L.2026, c.25 (A.5328), approved June 30, 2026, creates a useful supervisory case study. The statute regulates data brokers and data collectors, imposes purpose and data-minimization duties, prohibits covered data brokers and data collectors from selling or licensing sensitive data, and requires registration for covered participants. At the same time, the statute expressly provides that key provisions do not apply to an “insurance institution” subject to New Jersey’s Insurance Information Practices Act or to an “insurance-support organization” as defined by that Act. It also contains a GLBA-based exemption.

That structure should not be described as proof of unlawful discrimination by the legislature. The more precise supervisory concern is regulatory asymmetry: materially similar data-use, aggregation, AI-training, or monetization activities can receive different statutory treatment depending upon the legal identity or sectoral classification of the entity performing them. The uploaded governance analysis characterizes this as a potential bias in the architecture of state data-use laws because insurers and parts of their support ecosystem may be carved out even as their data holdings and downstream AI uses become increasingly technology-like.

7C.1 The Same Data Can Cross Different Legal Regimes

The problem is magnified by the insurance supply chain. The same policyholder information can move from an applicant to an agency, retail broker, wholesale broker, MGA/MGU, carrier, TPA, independent adjuster, claims-service provider, insurance-support organization, reinsurer, cloud provider, analytics vendor, and AI/model vendor. New Jersey’s Insurance Information Practices Act defines “insurance institution” separately from “agent,” and separately defines “insurance-support organization” to include certain entities that assemble or collect information for insurance transactions. Accordingly, whether a particular agency, brokerage, MGA, TPA, claims organization, vendor, or affiliated service company falls within an exemption is fact-specific; the carrier’s exemption cannot safely be treated as a blanket exemption for the entire chain.

The supervisory consequence is straightforward: the same data can become subject to different duties as it crosses organizational boundaries, while the operational risk created by misuse, AI training, exfiltration, correlation, or downstream weaponization does not change merely because the legal custodian changes. An entity-based carve-out therefore can create a mismatch between legal coverage and economic exposure.

7C.2 Insurance Regulation Should Follow the Activity and the Exposure

The NAIC should not assume that sector-specific privacy regulation makes the underlying AI/data exposure immaterial. A statutory exemption may be appropriate as a matter of legislative design because insurers are regulated elsewhere. But for ORSA, examination, and solvency purposes, exemption is not mitigation. The risk remains if the insurer retains, enriches, vectorizes, embeds, recursively reuses, commercializes, or trains AI on the data; if the information is shared with vendors or affiliates; if it enters a common enterprise model; or if it later becomes an adversarial intelligence asset after exfiltration.

The appropriate supervisory question is therefore not “Is the insurer exempt from this privacy statute?” It is “What activity is the insurer performing, what data is being used, what economic and operational value is being created, what legal basis governs each downstream use, and what correlated liabilities could arise if that use is later challenged or the data is lost?” That activity-and-exposure lens is more consistent with ORSA than a formalistic entity-label approach.

7C.3 The Carve-Out Can Create a Policyholder-Equity Problem

The asymmetry becomes more consequential where policyholder data is compelled or practically necessary to obtain insurance. Applications, recorded statements, financials, medical or prescription information

where permitted, telematics, photographs, loss-control reports, claims investigations, SIU material, engineering information, and other records are often supplied for a defined insurance purpose. If an insurer is exempt from a general data-broker or data-collector regime yet later converts that information into an enterprise AI asset, predictive model, behavioral inference system, or other source of corporate value, the exemption does not answer the separate questions of purpose limitation, contractual authority, legitimate interest, data equity, or whether the policyholder's expectations have been respected. The uploaded analysis emphasizes precisely this collision between FOMO-driven AI deployment, continuing legitimate interest, and the policyholder's insurable interest.

That issue belongs in the NAIC record because it can migrate from data governance into market conduct, unfair discrimination, contract litigation, D&O oversight, model remediation, and ultimately ORSA if the challenged practice is common across a book of business or embedded in a foundational enterprise model.

7C.4 Statutory Asymmetry Can Become an Aggregation Event

A carve-out can create latent concentration risk when management interprets legal exemption as permission for broad reuse. Consider a severe-but-plausible scenario in which an insurer has trained a foundational LLM on decades of policyholder and claims data under an assumed sectoral exemption. A later statute, regulatory interpretation, court ruling, or multistate examination narrows that exemption or determines that certain secondary uses fall outside the protected insurance activity. The resulting remediation could require model retraining, data deletion or segregation, re-performance of affected decisions, consumer notice, restitution, litigation defense, vendor remediation, and changes across multiple legal entities and lines at once.

That is not a conventional privacy fine. It is an enterprise-correlation scenario. The model, the corpus, the governance record, and the contractual basis for data use all become common causes of loss. In a sufficiently large multiline carrier, the consequences can implicate reserves, liquidity, reinsurance recoverability, capital planning, operational continuity, and reputation simultaneously.

7C.5 Proposed Supervisory Principle: Carve-Outs Must Be Mapped, Not Assumed

For v5.0 and related NAIC examination work, insurers should be required to maintain a state-by-state legal and operational map identifying: (1) the statutory basis for each material data use; (2) each entity and supply-chain participant relying on an exemption; (3) whether the exemption is entity-level, data-level, activity-level, or transaction-specific; (4) secondary and recursive AI uses not contemplated when the data was collected; (5) cross-border or cross-state transfers that change the governing regime; (6) contractual representations made to policyholders and vendors; and (7) the ORSA consequence if an exemption is narrowed, repealed, or judicially interpreted against the insurer.

Supervisory proposition: A privacy-law or data-broker exemption is a legal classification, not an exposure-reduction control. NAIC examination should treat the underlying data-use, AI, operational, discrimination, litigation, and recovery-rights risks on their economic substance and inherent severity before crediting any statutory carve-out.

Insurance-chain role	Illustrative New Jersey treatment	Potential asymmetry	NAIC implication
Carrier / insurance institution	Expressly exempt from relevant A.5328 data-broker/data-collector provisions if subject to the Insurance Information Practices Act	High-value data and AI uses can remain economically material even where the general statute does not apply	Do not credit exemption as risk mitigation; examine use, lineage, aggregation, and ORSA impact
Insurance-support	Expressly exempt where it	Large-scale collection for	Map function, data flow,

Insurance-chain role	Illustrative New Jersey treatment	Potential asymmetry	NAIC implication
organization	meets the statutory definition	insurance transactions may be carved out even though the same data could be highly valuable for AI and adversaries	downstream reuse, and vendor dependencies
Agent / broker / MGA / TPA / claims service	Treatment depends on statutory status, function, and other exemptions; agents are separately defined under the Insurance Information Practices Act	The same record can cross exempt and potentially non-exempt participants	Require entity-by-entity legal mapping across the supply chain
Cloud / analytics / AI / non-insurance vendor	Often governed by processor, controller, GLBA, contract, or other statutory analysis rather than an insurance carve-out	A carrier can outsource processing without outsourcing accountability or systemic exposure	Examine contract rights, auditability, provenance, incident cooperation, and recovery rights

8. Agentic AI: Delegated Authority, Tool Use, and Machine-Speed Accumulation

The addition of an agentic AI definition in v5.0 is consequential. Once a model can invoke tools, access systems, send communications, manipulate files, initiate workflows, or act on behalf of the insurer, the risk is not limited to the quality of generated text. The unit of supervision becomes delegated authority.

- What actions may the agent take without approval?
- What financial, contractual, coverage, claims, or operational consequences can follow from those actions?
- Can one agent invoke another agent or third-party model?
- Can an agent modify data that later becomes training or retrieval context?
- Can the system create, alter, or destroy evidence relevant to claims or recovery?
- Are there tested circuit breakers, kill switches, transaction limits, and identity/permission controls?
- How quickly could an error replicate across policyholders, claims, vendors, or lines before human detection?

Machine speed transforms frequency and aggregation. A low-probability error at the level of one recommendation can become material when thousands of parallel agents use the same model, data source, or flawed rule. Accordingly, the Supplement should ask about maximum plausible propagation before containment, not merely per-transaction error rates.

9. The Legal Feedback Loop: Digital Premises, Product Liability, Bailment, and Subrogation

The supervisory importance of recovery rights can be illustrated as a feedback loop. An insurer or insurance-service entity receives operational information for underwriting, loss control, claims, engineering, or other insurance purposes. That information is held within a digital environment. If stewardship fails and strategically valuable operational data enters an uncontrolled environment, AI-enabled parsing can make the data more useful to adversaries. A later physical or economic loss may then implicate product manufacturers, software vendors, integrators, professional-service providers, data custodians, or manufacturer representatives. The insurer pays the covered loss and seeks subrogation. The recovery path

then depends on evidence integrity, causation, contracts, indemnity, liability caps, waivers, audit rights, and the ability to reconstruct how data and AI systems contributed to the event.

The legal inquiry should not begin with a presumption that a data custodian stands outside product liability. Depending on jurisdiction and function, a custodian may also act as a manufacturer's representative, apparent manufacturer, data-component supplier, distributor, licensor, certifier, safety-service provider, or integral participant in the product enterprise. Product-enterprise doctrines, negligent-undertaking theories, agency principles, warranty, professional liability, and subrogation can therefore intersect when a party supplies critical data, safety services, AI recommendations, product representations, or operational intelligence. The examination issue is not to adjudicate tort liability in advance. It is to determine whether the insurer understands the liability and recovery assumptions embedded in pricing, reserves, reinsurance, contracts, and capital.

The supervisory question

After a material AI- or data-enabled loss, can the insurer prove origin of loss, cause of loss, responsible-party allocation, evidence authenticity, and contractual recovery rights, or has its own digital architecture or vendor contracting impaired the very subrogation economics assumed when the risk was priced?

10. Proposed ORSA Stress Scenarios

The following scenarios are not predictions. They are examples of severe-but-plausible exercises that can reveal whether common AI dependencies create material financial exposure.

Scenario	Stress mechanism	Financial-condition questions
Foundational model defect	A shared enterprise LLM produces a systematic interpretation error for months across underwriting and claims.	Restitution; reserve reopening; litigation; reinsurance; market-conduct remediation; capital impact.
Third-party model opacity	A material model produces anomalous outcomes, but the vendor restricts access to logs, weights, testing, or causal evidence.	Can the insurer validate, defend, remediate, and recover? Is opacity reflected in inherent risk?
Agentic authority failure	An authorized agent executes unauthorized or unintended actions across claims, payments, communications, or records.	Transaction limits; evidence preservation; fraud/crime/cyber interaction; occurrence aggregation.
Data-lineage contamination	Synthetic or inherited data defects propagate into successive model versions.	Can lineage be reconstructed? What decisions must be reopened? How many model generations are affected?
Cognitive dependency event	A critical AI system is unavailable or prohibited after personnel have become operationally dependent on it.	Business continuity; manual capability; claims/underwriting throughput; service levels; operational loss.
Recovery-rights impairment	A common vendor contract across many insureds caps liability or restricts discovery after a systemic technology loss.	Expected subrogation vs actual recovery; pricing/reserving mismatch; multi-line accumulation.
AI asset compromise	Model, vector store, embeddings, training data, or retrieval systems are corrupted/encrypted.	Replacement/retraining cost; BI; evidence integrity; third-party dependencies; cyber/reinsurance response.

8. State-law carve-out reversal scenario: A foundational insurer LLM or AI ecosystem has been trained on data under a broad interpretation of insurance-sector exemptions. A later state law, regulator, court, or multistate examination narrows the exemption or distinguishes traditional insurance use from secondary AI training and monetization. Stress-test model retraining/replacement, re-performance of underwriting or claims decisions, restitution, consumer remediation, legal defense, vendor disputes, operational interruption, reinsurance impact, and prospective capital consequences.

11. Proposed Examination Language for the Supplement

The following language is offered as a drafting proposal. It is intentionally designed to fit the Supplement's existing concepts of inherent risk, materiality, AI systems/models, third parties, and data use.

11.1 Proposed General Instruction: Enterprise Aggregation

Proposed text

For AI Systems or AI Models with Direct Consumer Impact, Material Financial Impact, or common dependencies reasonably capable of producing material aggregate impact, regulators should consider whether the insurer has identified and assessed concentrations arising from common models, model families, vendors, cloud/service providers, data sources, agent frameworks, tools, or failure modes across legal entities, lines of business, functions, and policyholder populations. Where such concentration may be material, regulators should consider whether the exposure is incorporated into the insurer's enterprise risk management and ORSA processes, as applicable.

11.2 Proposed Inherent-Risk Instruction

Proposed text

Inherent risk should be assessed before credit is given for human review, contractual indemnity, vendor warranties, insurance, audit reports, governance controls, or other mitigation. Where the insurer is unable to obtain information, testing access, audit rights, logs, provenance, or other evidence necessary to evaluate a material third-party dependency, that limitation should be considered in the inherent-risk assessment and should not, by itself, reduce the insurer's accountability for its use of the system.

11.3 Proposed Recovery-Rights Inquiry

Proposed text

For financially material AI uses or dependencies, describe whether the insurer has assessed the effect of AI systems, third-party contracts, data provenance, evidence preservation, model/version logging, arbitration provisions, indemnification terms, liability limitations, waivers, or other restrictions on salvage, subrogation, contribution, causation analysis, coverage determination, and recovery rights. Identify material limitations and explain how they are reflected in underwriting, claims, reserving, reinsurance, enterprise risk management, or ORSA, as applicable.

11.4 Proposed Data-Lineage Inquiry

Proposed text

For material models, identify the principal data sources and their original purposes; significant transformations or derived features; use of synthetic, AI-generated, or distilled data; downstream propagation into models or agents; retention and deletion controls; and any operational, engineering, OT/ICS, safety, catastrophe, or other strategically sensitive data whose compromise or misuse could create physical or systemic consequences. Where model outputs are reused as future training data, describe lineage and controls designed to detect recursive amplification of error or bias.

11.5 Proposed Human-Oversight Inquiry

Proposed text

Where human review is identified as a mitigating control, describe the reviewer's qualifications, access to primary evidence, authority to override, time and incentive structure, challenge/override monitoring, periodic independent competency testing, and business-continuity procedures for operation without the AI system. Explain how the insurer evaluates automation bias, overreliance, or degradation of independent review capability.

12. Suggested Placement Within v5.0 Exhibits

Existing area identified in v5.0 summary	Recommended addition
Exhibit A - Model Inventory / Material Financial Impact	Add common-dependency fields: model family, upstream provider, cloud/host, shared data source, agent/tool framework, enterprise functions, legal entities, lines, aggregation rating, ORSA mapping.
Exhibit B - AIS Program / third-party oversight	Add accountability-through-opacity, human-review effectiveness, board reporting on common dependencies, and recovery-rights governance.
Exhibit C - use case, development, risk, limitations	Add maximum propagation before containment, delegated authority, evidence impacts, failure-mode correlation, recovery impairment, and business-continuity dependency.
Exhibit D - data sets / data use / dictionary	Add provenance, original purpose, secondary/recursive use, synthetic/distilled data, operational/OT sensitivity, propagation, retention, and incident reconstruction.
Definitions	Add Enterprise AI Concentration, Common AI Dependency, Recovery-Rights Impairment, Model/Data Lineage, and Effective Human Oversight.

13. Proposed Definitions

These definitions are offered for discussion and are not presented as existing NAIC definitions.

Term	Proposed supervisory definition
Common AI Dependency	A model, model family, vendor, provider, data source, cloud/service, agent framework, tool, identity system, or other component relied upon by multiple material AI systems, business

Term	Proposed supervisory definition
	functions, legal entities, or lines of insurance.
Enterprise AI Concentration	The aggregate exposure arising when common AI dependencies or failure modes can produce correlated consumer, operational, legal, financial, or solvency effects across otherwise separate uses.
Recovery-Rights Impairment	A material reduction in expected salvage, subrogation, contribution, indemnity, evidentiary, or other recovery rights caused by technology architecture, data provenance, AI use, contractual terms, missing evidence, or third-party restrictions.
Model/Data Lineage	The documented chain of model versions, training/fine-tuning inputs, synthetic or distilled data, transformations, retrieval sources, and downstream dependencies sufficient to reconstruct material provenance and inheritance.
Effective Human Oversight	Human review performed by a person with adequate subject-matter competence, access to relevant primary evidence, authority and time to challenge or override, and periodic demonstration of independent capability.
Digital Premises Liability - analytical concept	Potential liability arising from failure to maintain a reasonably safe digital environment for entrusted data, systems, users, and connected physical operations, including foreseeable downstream consequences of data exfiltration, AI-enabled misuse, or unsafe autonomous action.

14. Examiner Question Set

1. Identify every AI model or system with Direct Consumer Impact or Material Financial Impact, and identify any non-material systems sharing a common dependency whose aggregate failure could be financially material.
2. Which model families, vendors, cloud providers, data sources, identity providers, vector stores, agent frameworks, or tools are common to more than one material use case?
3. What is the maximum plausible number of policyholders, claims, transactions, legal entities, or lines affected before a common AI failure could be detected and contained?
4. Has management quantified gross exposure before crediting human review, contracts, indemnity, insurance, audit reports, or other controls?
5. Where the insurer relies on a third party, what information can it not obtain? How is that opacity reflected in inherent risk?
6. Do contracts preserve audit rights, regulator cooperation, incident evidence, logs, model/version information, and technical assistance necessary for claims and recovery?
7. Has the insurer assessed whether vendor or policyholder technology contracts materially impair subrogation, salvage, contribution, or causation analysis?
8. Can the insurer reconstruct which model version, prompt/context, data source, tool call, and human action materially contributed to a decision or loss?
9. What AI-generated outputs are reused for training, fine-tuning, retrieval, or synthetic data? How is recursive error or bias propagation detected?
10. Does the insurer classify strategically sensitive operational/OT/engineering information separately from conventional PII for governance and incident response?
11. Where human review is a control, how is independent competency tested and how does the insurer measure automation bias or overreliance?

12. What would happen to critical operations if the AI system or provider became unavailable for 30, 90, or 180 days?
13. Which AI risks or concentrations are included in ORSA scenarios, risk appetite, risk limits, liquidity planning, reinsurance strategy, or prospective capital assessment?
14. How does the board receive information about enterprise AI concentration rather than only individual model performance?
15. What events would trigger mandatory escalation, circuit breaking, suspension, or model retirement?
16. For AI-assisted legal, claims, engineering, actuarial, medical-review, or other professional functions, what evidence demonstrates independent verification rather than nominal human approval?
17. Can the insurer reconstruct the verification trail for a challenged AI-assisted decision, including model/version, inputs, authoritative sources, changes made by the professional, exceptions, and escalation?
18. Have third-party counsel, TPAs, claims vendors, MGAs, agencies, brokers, consultants, and other insurance-supply-chain participants been assessed for AI-use terms that could impair confidentiality, privilege, audit rights, evidence preservation, indemnity, or subrogation?
19. Does the insurer classify data by adversary utility and operational consequence in addition to privacy sensitivity?
20. Can the insurer identify “Non-Ferrous” operational or strategic data - including OT/ICS, engineering, loss-control, infrastructure, catastrophe, claims, and business-continuity information - across its own systems and insurance supply chain?
21. Does breach severity analysis consider AI-enabled parsing, correlation, inference, intelligence half-life, and the possibility that stolen information becomes more useful as additional datasets are compromised?
22. Has the insurer modeled a scenario in which previously exfiltrated operational intelligence is later used to target insured physical systems, causing property, casualty, product-liability, or business-interruption losses?
23. Are agencies, brokerages, MGAs/MGUs, TPAs, claims-service organizations, loss-control/engineering firms, counsel, reinsurers, cloud providers, and AI vendors contractually required to preserve provenance, incident evidence, audit access, and cooperation needed for causation and subrogation?
24. For each material AI or data use, has the insurer identified whether applicable privacy/data-use exemptions are entity-level, activity-level, data-level, or transaction-specific?
25. Which agencies, brokerages, MGAs/MGUs, TPAs, claims-service organizations, insurance-support organizations, reinsurers, cloud providers, analytics firms, and AI vendors rely on different statutory exemptions or legal bases while handling the same policyholder data?
26. Has management evaluated whether secondary AI training, recursive model development, behavioral analytics, data enrichment, or cross-company sharing remains within the insurance purpose supporting the original collection and any claimed exemption?
27. Does ORSA include a scenario in which a material state-law exemption is narrowed, repealed, or judicially construed not to cover a material secondary data use?
28. Does the board receive reporting that distinguishes “legally exempt” from “low inherent risk,” and are statutory carve-outs prohibited from being treated as risk-mitigation controls in model or enterprise assessments?

15. Insurance Is Now in the Governance Business

The insurance industry has always done more than transfer risk. It has influenced safety through underwriting discipline, engineering, inspections, claims learning, and loss-prevention standards. AI turns that governance function inward. Insurers are not merely underwriting the AI risks of policyholders; they are increasingly embedding AI into the mechanisms by which they select risk, investigate loss, interpret evidence, identify responsible parties, preserve recovery rights, and allocate capital. Gerry Kennedy's August 2026 NAIC warning argues that data, AI, Digital Premises Liability, technology contracts, subrogation, reserves, and capital are no longer separate problems; they are becoming connected through correlation. The Travelers POV adds a second warning: a proprietary insurance LLM can convert decades of institutional judgment into a common enterprise reasoning layer. The purpose of this paper is not to conclude that such systems are unsafe. It is to ensure that the examination framework can recognize when their common dependencies become financially material.

Marie Curie's history supplies the closing discipline. Humanity did not stop using radiation when its dangers became apparent. It learned to measure dose, establish exposure limits, shield workers, monitor contamination, and build a governance system around a powerful phenomenon. Insurance should approach AI with the same intellectual humility. The absence of a complete loss history does not relieve supervisors or boards from measuring emerging exposure. The appropriate response is neither panic nor FOMO. It is governance.

Bottom line

The NAIC v5.0 exposure should be used to make AI aggregation examinable before a common model, vendor, data source, cloud dependency, agentic failure, or recovery-rights impairment becomes the insurance equivalent of an unmodeled catastrophe. The industry should measure the exposure while it still has the opportunity to govern it.

16. Formal Recommendations to the Working Group

1. Add an explicit enterprise-aggregation instruction to v5.0 for common AI dependencies and failure modes across lines, entities, and functions.
2. Require Material Financial Impact analysis to identify ORSA/ERM mapping where aggregate AI exposure is material.
3. Add recovery-rights inquiries covering salvage, subrogation, contribution, evidence integrity, causation, and contractual impairment.
4. State that vendor opacity, proprietary restrictions, or inability to test a material third-party system are components of inherent risk and do not reduce insurer accountability.
5. Require inherent/gross exposure assessment before mitigation is credited.
6. Expand data inquiries from field-level data dictionaries to lineage, purpose, secondary/recursive use, synthetic/distilled data, propagation, and operational/OT sensitivity.
7. Require testing of human oversight as a control, including competence, primary-evidence access, override authority, and operational capability when AI is unavailable.
8. For agentic AI, examine delegated authority, tool permissions, inter-agent dependencies, maximum propagation, and tested circuit breakers.

9. Coordinate v5.0 implementation with the ORSA Implementation (E) Subgroup, Financial Examiners Handbook work, Market Conduct Examination Guidelines work, and Third-Party Data and Models (H) Working Group.
10. Encourage examiners to treat common AI concentration as an accumulation problem even where individual use cases fall below stand-alone materiality thresholds.
11. Add a verification-trail requirement for AI-assisted professional and quasi-professional decisions, including evidence of independent human review, authoritative-source checking, version traceability, and override/escalation authority.
12. Require third-party AI governance reviews to include external counsel, agencies, brokerages, MGAs/MGUs, TPAs, independent adjusters, claims-service organizations, loss-control and engineering firms, reinsurers, cloud providers, model vendors, and other participants whose AI use can affect confidentiality, evidence, causation, professional liability, or recovery rights.
13. Add an adversary-utility dimension to data-risk examination so privacy sensitivity is not the sole measure of materiality.
14. Require insurers to identify and govern strategically sensitive operational, engineering, OT/ICS, claims, loss-control, infrastructure, and catastrophe information across the full insurance supply chain.
15. Require breach and ORSA scenarios to consider AI-enabled parsing, correlation, inference, long-lived intelligence value, and recursive accumulation of stolen operational data across events.
16. Require examination of whether data-custody, vendor, and technology contracts preserve evidence, provenance, model/data lineage, technical cooperation, and subrogation or contribution rights after an AI-enabled loss.
17. Add a state-law asymmetry inquiry requiring insurers to map insurance-sector, GLBA, insurance-support-organization, processor, and other material exemptions across the AI/data supply chain.
18. State expressly that a statutory exemption or carve-out is not a mitigating control for inherent-risk scoring and does not reduce the insurer's ORSA obligation to evaluate the underlying economic exposure.
19. Require stress testing of legal-basis and exemption failure, including model retraining, data segregation/deletion, re-performance of decisions, restitution, litigation, vendor disputes, and capital effects.
20. Encourage coordination between insurance regulators and general privacy/data regulators so materially similar data-processing activities do not become invisible merely because they occur inside a sectorally exempt entity.

References and Source Notes

Web sources were reviewed as of September 2, 2026. The Kennedy LinkedIn and governance materials are included as advocacy/source material for the supervisory thesis; they are not presented as NAIC positions. The New York court rule and Munich ruling are presented as external legal signals, not controlling insurance law. The Ferrous/Non-Ferrous Data taxonomy and "Data Magnet" are analytical metaphors proposed for supervisory discussion, not established NAIC classifications. The discussion of "bias" in state privacy law is framed as statutory or regulatory asymmetry created by categorical exemptions; it is not an allegation of unconstitutional discrimination or legislative misconduct.

[1] **Big Data and Artificial Intelligence (H) Working Group - Exposure Drafts.** National Association of Insurance Commissioners (2026). <https://content.naic.org/committees/h/big-data-artificial-intelligence-wg>

[2] **AI Risk Evaluation Supplement - Summary of Changes.** National Association of Insurance Commissioners (2026). <https://content.naic.org/sites/default/files/inline-files/summary-of-changes.pdf>

- [3] **Model Bulletin: Use of Artificial Intelligence Systems by Insurers.** National Association of Insurance Commissioners (Dec. 4, 2023). https://content.naic.org/sites/default/files/inline-files/2023-12-4%20Model%20Bulletin_Adopted_0.pdf
- [4] **Travelers Advances AI Strategy with Award-Winning Insurance-Specific Large Language Model.** The Travelers Companies, Inc. Investor Relations (June 30, 2026). <https://investor.travelers.com/newsroom/press-releases/news-details/2026/Travelers-Advances-AI-Strategy-with-Award-Winning-Insurance-Specific-Large-Language-Model/default.aspx>
- [5] **Artificial Intelligence - Insurance Topic.** National Association of Insurance Commissioners (2026). <https://content.naic.org/insurance-topics/artificial-intelligence>
- [6] **Third-Party Data and Models (H) Working Group.** National Association of Insurance Commissioners (2026). <https://content.naic.org/committees/h/third-party-data-models-wg>
- [7] **Own Risk and Solvency Assessment (ORSA) Implementation (E) Subgroup.** National Association of Insurance Commissioners (2026). <https://content.naic.org/committees/e/own-risk-solvency-assessment-implementation-sg>
- [8] **Artificial Intelligence and Insurance Regulation.** Cassandra R. Cole, Stephen G. Fier & Chad Marzen, Journal of Insurance Regulation (May 5, 2026). <https://content.naic.org/research/jir/artificial-intelligence-and-insurance-regulation>
- [9] **Marie Curie - Facts.** Nobel Prize Outreach (2026). <https://www.nobelprize.org/prizes/physics/1903/marie-curie/facts/>
- [10] **Marie Curie - Women Who Changed Science.** Nobel Prize Outreach (2026). <https://www.nobelprize.org/womenwhochangedscience/stories/marie-curie>
- [11] **Les cahiers de Marie Curie conservés à la BnF sont-ils radioactifs?.** Bibliothèque nationale de France (2026). <https://www.bnf.fr/fr/les-cahiers-de-marie-curie-conserves-la-bnf-sont-ils-radioactifs>
- [12] **#NAIC: This Is the Warning.** Gerry Kennedy, LinkedIn (Aug. 13, 2026). <https://www.linkedin.com/pulse/naic-warning-gerry-kennedy-hwmvc>
- [13] **A POV from a Old Travelers Underwriter who arrived every morning and said to his colleague, “What pile of lies do I have to go through today?”.** Gerry Kennedy, LinkedIn (2026). <https://www.linkedin.com/pulse/pov-from-old-travelers-underwriter-who-arrived-every-morning-kennedy-jymyc>
- [14] **NAIC Members Approve Model Bulletin on Use of AI by Insurers.** National Association of Insurance Commissioners (Dec. 4, 2023). <https://content.naic.org/article/naic-members-approve-model-bulletin-use-ai-insurers>
- [15] **Market Conduct Examination Guidelines (D) Working Group - 2026 Charges.** National Association of Insurance Commissioners (2026). <https://content.naic.org/committees/d/market-conduct-examination-guidelines-wg>

[16] The Hidden Insurance Consequences of New York's AI Court Rule. What Is in YOUR POLICY? Gerry Kennedy, LinkedIn (2026). <https://lnkd.in/p/g97-WNqY>

[17] 22 NYCRR Part 161 - Use of Artificial Intelligence Technology, including Section 161.3 Policy. New York State Unified Court System (effective June 1, 2026). <https://www.nycourts.gov/rules/part-161-use-artificial-intelligence-technology>

[18] Section 161 Appendix A - Model Rule Regarding Use of Artificial Intelligence Technology in Preparing Court Papers. New York State Unified Court System (2026). <https://www.nycourts.gov/rules/rule/section-161appendix-model-rule-regarding-use-artificial-intelligence-technology>

[19] Germany: Court Holds Google Liable for Incorrect AI Overviews. Law Library of Congress, Global Legal Monitor (July 17, 2026), discussing LG Munchen I, May 28, 2026, 26 O 869/26. <https://www.loc.gov/item/global-legal-monitor/2026-07-17/germany-court-holds-google-liable-for-incorrect-ai-overviews/>

[20] Mining for Ferrous or Non-Ferrous Data: What Could Happen When a “Data Magnet is Applied”. Gerry Kennedy, LinkedIn (Aug. 5, 2026). <https://www.linkedin.com/pulse/mining-ferrous-non-ferrous-data-what-could-happen-when-gerry-kennedy-dn3pc>

[21] New Jersey P.L.2026, c.25 (A.5328), approved June 30, 2026, regulating data brokers, data collectors, and sensitive-data sales/licensing. Official New Jersey Legislature text.

[22] N.J.S.A. 17:23A-2, Insurance Information Practices Act definitions, including “agent,” “insurance institution,” and “insurance-support organization.”

[23] New Jersey A.5328 First Reprint / enacted text, exemptions for GLBA-regulated financial institutions, insurance institutions subject to P.L.1985, c.179, and insurance-support organizations.

[24] Kennedy governance analysis: “FOMO, Data Legitimate Interest, and the Next ORSA Risk for Insurance,” including analysis of New Jersey A.5328, insurance-sector carve-outs, downstream AI use, Data Equity, legitimate interest, and ORSA implications (2026).

Appendix A - AI Aggregation and Recovery-Rights Examination Matrix

Domain	Minimum evidence / inquiry
Model / system	Purpose; version; owner; model family; materiality; consumer impact; financial impact.
Common dependencies	Vendor; cloud/host; data source; vector store; identity; agent framework; tools; APIs; upstream foundation model.
Enterprise reach	Legal entities; functions; lines; jurisdictions; policyholder populations; claims populations; policy years.
Authority	Read/write/send/pay/approve/deny/modify/delete/code/deploy privileges; human approval thresholds.
Data lineage	Source; original purpose; legal/contractual authority; transformations; synthetic/distilled data; recursive reuse.
Strategic data sensitivity	PII; underwriting; claims; engineering; OT/ICS; critical infrastructure; catastrophe; legal/privileged; recovery evidence.
Failure modes	Hallucination; bias; drift; prompt injection; data poisoning; tool misuse; deception; compromise; availability; vendor restriction.
Propagation	Maximum plausible transactions/claims/policyholders/functions affected before detection and containment.
Evidence integrity	Logging; provenance; versioning; prompts/context; tool calls; human actions; immutable records; chain of custody.
Recovery rights	Subrogation; salvage; contribution; indemnity; liability caps; arbitration; waivers; forum; discovery/audit restrictions.
Human control	Qualifications; primary-evidence access; override authority; challenge metrics; independent competency; no-AI continuity.
Financial-condition mapping	Reserves; reinsurance; liquidity; operational loss; ORSA scenarios; risk appetite; prospective capital.
Governance evidence	Board/risk committee reporting; AIS Program; independent validation; internal audit; incident escalation; circuit breakers.

Appendix B - Digital Premises / Subrogation Loss Chain

1. Entrusted or acquired data enters insurer or insurance-supply-chain custody.
2. Data is aggregated, transformed, embedded, vectorized, enriched, or used in AI/model workflows.
3. Common AI or data dependency becomes operationally material across functions or insureds.
4. Failure, compromise, exfiltration, hallucination, agentic misuse, or contract limitation occurs.
5. Downstream physical, economic, consumer, product, professional, or operational loss manifests.
6. Insurer pays defense, indemnity, first-party, business-income, cyber, workers' compensation, or other covered amounts.
7. Claims investigation attempts to establish origin of loss and causes of loss.
8. Salvage, subrogation, contribution, product-liability, negligent-undertaking, contract, or other recovery theories are evaluated.
9. Recovery is impaired if evidence, provenance, contractual rights, audit access, or technical cooperation were not preserved.
10. Unrecovered amounts feed back into loss cost, reserves, reinsurance economics, capital, and future underwriting.

Appendix C - Suggested Transmittal Language

Subject: Comments on AI Risk Evaluation Supplement v5.0 - Enterprise AI Aggregation, ORSA, and Recovery Rights

To the Big Data and Artificial Intelligence (H) Working Group:

Thank you for the opportunity to comment on the formally exposed AI Risk Evaluation Supplement version 5.0. The revisions concerning inherent risk, material financial impact, third-party oversight, agentic AI, model inventories, model limitations, and data use create a valuable foundation for examination. The attached paper respectfully recommends clarifying that AI supervision must also address enterprise aggregation and recovery-rights impairment. A model may be well-governed in isolation while a common model family, vendor, cloud dependency, data source, agent framework, or contractual architecture creates correlated exposure across legal entities and insurance lines. We therefore recommend that the Supplement require gross/inherent exposure assessment before mitigation, map material aggregated dependencies into ORSA/ERM processes, and examine whether AI use or third-party restrictions impair evidence, causation, salvage, subrogation, contribution, or other recovery rights. These additions would convert emerging cross-line risk into examinable supervisory language before it becomes a capital event.

September 10, 2026

Via email: ssobel@naic.org; maromero@naic.org

Scott Sobel, AI Policy Advisor

Miguel Romero, Director of Innovation, Cybersecurity and Technology

Big Data and Artificial Intelligence (H) Working Group

National Association of Insurance Commissioners

Re: Written Comments on AI Risk Evaluation Supplement v5.0

Dear Mr. Sobel and Mr. Romero:

On behalf of Legion Modular IP Holdings, Inc., please accept the attached Technical Comment on the NAIC Artificial Intelligence Risk Evaluation Supplement v5.0 and Attachment 1, a proposed Insurance AI Operational Assurance Profile. These materials are submitted by Legion in a private capacity and do not represent the views of any governmental employer or agency.

Legion generally supports the direction of Version 5.0. Our comments recommend targeted refinements intended to make the Supplement more useful for material and agentic AI systems while preserving its flexible, risk-based character. In particular, we recommend a clearer distinction between model architecture and system agency; proportionate mapping of material system dependencies, tools and actions; more explicit authority and human-intervention boundaries; material-change controls; decision-and-action reconstruction; and a consumer-risk scoping concept that complements financial materiality.

Attachment 1 translates those recommendations into an implementation-neutral operational assurance profile. It sets out outcome-based control objectives and representative evidence that could be considered, in whole or in part, as an optional pilot, additional exhibit, examiner resource, appendix, module, or other tool the Working Group finds useful. The Profile is not tied to a particular vendor, model, cloud environment, or implementation architecture.

The Profile is informed by Legion's broader public Institutional AI Governance Standard, but the full Standard is not included with this submission. We have intentionally kept the NAIC package focused on insurance-specific examination and assurance issues and excluded proprietary implementation architecture, algorithms, runtime mechanisms, trade secrets, and client-specific work product.

We appreciate the Working Group's continued work on the Supplement and the opportunity to comment. Legion would welcome the opportunity to answer questions or discuss any portion of the proposed Profile as the Working Group continues its review.

Sincerely,

William C. Biddle

Chief Executive Officer

Legion Modular IP Holdings, Inc.

wcb@legionmodularip.com

Attachments:

- Technical Comment - NAIC Artificial Intelligence Risk Evaluation Supplement v5.0
- Attachment 1 - Proposed Insurance AI Operational Assurance Profile

Legion Modular IP

Technical Comment

NAIC Artificial Intelligence Risk Evaluation Supplement v5.0

Submitted to: NAIC Big Data and Artificial Intelligence (H) Working Group

Attention: Scott Sobel, AI Policy Advisor; Miguel Romero, Director of Innovation, Cybersecurity and Technology

Submitted by: William C. Biddle, Chief Executive Officer
Legion Modular IP Holdings, Inc.

Date: September 10, 2026

Document ID: LMIP-NAIC-AIRES5-2026.09-R1.0

Re: Exposure draft issued August 31, 2026

EXECUTIVE POSITION Version 5.0 is a meaningful advance. Legion recommends adoption with targeted refinements that preserve the Supplement's flexible, risk-based character while making the review of material agentic AI systems more technically coherent, operationally testable, and proportionate.

Introduction

Legion Modular IP Holdings, Inc. appreciates the opportunity to comment on Version 5.0 of the Artificial Intelligence Risk Evaluation Supplement. Legion is a privately held Kentucky corporation focused on institutional AI governance, risk management, and assurance. These comments express the views of Legion in a private capacity; they do not represent the views of any governmental employer or agency.

Version 5.0 materially improves the Supplement through its treatment of inherent risk, materiality, the distinction between AI Systems and AI Models, direct consumer and material financial impact, model inventories, explainability and transparency, third-party oversight, and agentic AI. Legion also supports the Supplement's express status as an optional aid to existing examination and analysis procedures, its reliance on regulator judgment, and its direction that requests be customized to the scope and risk of the inquiry.

The principal opportunity for further refinement arises from the Supplement's new definition of agentic AI. A model can be well documented and appropriately validated while the larger system in which it operates creates risk through multi-step planning, tool use, delegated permissions, interactions with other models or services, and execution of actions. For material uses, the regulatory object is therefore not only the model, but the relevant decision-and-action system.

1. Separate model type from system agency

Exhibit C, item 3 currently requests the model type used in the AI System and gives "GLM, generative/agentive AI, or other AI/ML algorithm" as examples. The Supplement's own definitions properly describe agentive AI as an AI System that pursues objectives across multiple steps. Agentive behavior is therefore not, by itself, a model architecture or algorithm. A generalized linear model, neural network, or large language model may be a component of an agentive system; conversely, the same model may operate in a non-agentive application. [2]

Legion recommends separating this field into distinct attributes:

- Model architecture or technique: for example, GLM, tree-based model, neural network, large language model, or other statistical or machine-learning technique.
- System function: support, augment, or automate, using the Supplement's existing operational definitions.
- Agency and autonomy: whether and how the system plans or acts across multiple steps; the limits on autonomous action; required human intervention; and any permitted delegation. Bounded systems may also be multi-step.
- Tools and actions: whether the system can invoke tools, APIs, databases, communications, workflows, or other agents, and the categories of actions available.

This correction would improve comparability across companies and help examiners distinguish model risk from system orchestration and action risk without prescribing a particular technology design. Apply the same distinction to the generative/agentive model categories in Exhibit A so that inventory and detailed-review responses use consistent attributes.

2. Add a targeted material-agentive-system dependency and action map

Exhibit A appropriately makes the Model Inventory explicit, and the NAIC Model Bulletin contemplates inventories of both Predictive Models and AI Systems. For a material or high-risk agentive use, however, a model inventory alone may not show how the components combine to affect a regulated decision or financial process. [2, 3]

Legion recommends a narrowly scoped system record, requested only when Exhibit A identifies direct consumer impact, material financial impact, high inherent risk, or material agency. That record could identify:

- the AI System or use-case identifier, accountable business owner, and affected insurance process;
- material models, versions, providers, data sources, retrieval sources, tools, APIs, and external services;
- the decisions the system may recommend and the actions it may initiate or execute;
- the relevant human roles, downstream systems, third-party dependencies, and change owner; and
- the logging, monitoring, escalation, and incident-management points applicable to the system.

This would not require a broad enterprise architecture submission for low-risk back-office tools. It would give an examiner a proportionate view of the material dependency chain when a component-by-component review could miss the operative risk.

3. Operationalize authority, action boundaries, and reversibility

Exhibit B already asks how an insurer assesses autonomy, reversibility, and reporting-impact risk. That question is well directed. For agentic systems, it could be made more examiner-ready by adding targeted subquestions addressing: [2]

- which actions are authorized, prohibited, or subject to human approval;
- how those boundaries are technically enforced, including service-account, tool, and API permissions;
- whether the system may invoke another agent, change a workflow, or obtain expanded permissions;
- who may approve, override, suspend, or terminate operation, and how attempted boundary violations are recorded; and
- whether an action can be reversed, rolled back, or compensated for, and what fallback process applies when it cannot.

The phrase "human in the loop" can describe materially different controls. A person may merely observe, review after the fact, approve before execution, or possess an effective ability to stop the action. The Supplement would benefit from distinguishing those roles and asking whether authority boundaries are actually enforceable in operation.

4. Treat operational change as part of ongoing assurance

Exhibits B and C appropriately address validation, testing, ongoing monitoring, and the effectiveness of the AIS Program. Agentic systems can change materially even when the underlying model version does not. Relevant changes may include system instructions, prompts, workflow logic, retrieval sources, connected tools, routing, permissions, vendor behavior, or the composition of multiple agents.

For material systems, Legion recommends asking whether the insurer has risk-proportionate criteria for identifying a material change and for triggering approval, revalidation, or enhanced monitoring. The inquiry could cover predeployment testing, periodic or event-triggered reassessment, changes in operational outcomes, incidents and near misses, drift or degradation, third-party changes, and safe retirement or fallback. The Supplement need not prescribe a single methodology or frequency; the central question is whether the company can show that meaningful changes are detected, evaluated, authorized, and monitored.

5. Add proportionate decision-and-action traceability

Version 5.0 strengthens documentary traceability by asking companies to identify document titles and page numbers supporting checklist responses. Legion supports that change. For selected material systems, the Supplement should also ask whether the insurer can reconstruct a representative decision or a significant adverse event at a level appropriate to the risk. [2]

A reconstruction inquiry could address the relevant system and model version; material data or evidence received and its source; the applicable business rule, policy, or control; the output or recommendation; any human or technical authorization and override; the action taken; the downstream result; and the remedial response, if any. This need not require source code, proprietary algorithms, unrestricted access to trade secrets, or retention of every internal computation. A sampled or incident-based approach, subject to applicable confidentiality protections, can provide evidence that governance operated in practice rather than only on paper.

6. Preserve financial materiality while adding consumer-risk significance

Version 5.0 defines materiality by reference to a dollar amount that would influence an examiner's view of financial position. That is appropriate for financial examination. The Supplement, however, also serves a consumer-protection purpose. A practice may have modest aggregate dollar impact while producing repeated denials, delays, differential treatment, privacy harm, or other significant effects for consumers. [2]

Legion recommends preserving the Financial Condition Examiners Handbook concept for financial materiality while adding a separate consumer-risk scoping concept. Relevant factors could include the number or proportion of consumers affected, severity, duration, reversibility, impact on vulnerable consumers, recurrence or systemic pattern, potential discriminatory effect, legal significance, and complaint or adverse-outcome signals. This would guide the scope of inquiry without creating a new substantive legal standard or displacing existing handbook guidance.

Implementation principles

These recommendations can be implemented without changing the Supplement's role. Legion recommends that any additional questions remain optional, targeted, and proportionate; be used primarily for systems with direct consumer impact, material financial impact, high inherent risk, or material agency; permit narrative or alternative evidence where appropriate; coordinate with existing examination handbooks; and expressly respect applicable confidentiality, proprietary-information, and trade-secret protections.

Attachment 1 supplies a proposed Insurance AI Operational Assurance Profile that converts these recommendations into implementation-neutral control objectives and examiner evidence categories. Legion recommends that the Working Group consider the Profile, in whole or in part, through a limited pilot, additional exhibit, examiner resource, appendix, module, or other mechanism it considers appropriate, while avoiding duplicative submissions or disproportionate burden.

Conclusion

Version 5.0 is a constructive and timely advance. With the targeted refinements described above, the Supplement would more clearly distinguish models from the systems that use them, connect agentic capability to authority and action risk, and help regulators evaluate whether governance remains effective during operation. Legion appreciates the Working Group's consideration of these comments.

Respectfully,
William C. Biddle
Chief Executive Officer
Legion Modular IP Holdings, Inc.

© 2026 Legion Modular IP Holdings, Inc. All rights reserved. This comment and Attachment 1 are offered as implementation-neutral policy materials and do not grant any license to proprietary Legion implementation technology.

For more information
wcb@legionmodularip.com

Attachment 1

Proposed Insurance AI Operational Assurance Profile

Optional System-Level Governance and Assurance Framework for Material AI Systems

Implementation-neutral framework informed by the Institutional AI Governance Standard

Relationship to the generic Standard: Legion's public Institutional AI Governance Standard uses ten governance principles, five descriptive maturity levels, and eight core governance domains as a sector-neutral compatibility framework. This Insurance Profile translates that public framework into insurer-specific operational-assurance controls. It does not require NAIC to adopt Legion's generic maturity model or any proprietary Legion technology.

Prepared for: NAIC Big Data and Artificial Intelligence (H) Working Group

Submitted by: William C. Biddle, Chief Executive Officer
Legion Modular IP Holdings, Inc.

Date: September 10, 2026

PROPOSED CONSIDERATION Legion respectfully offers this Profile as an implementation-neutral framework for NAIC consideration. The Working Group could use it, in whole or in part, as an optional system-level supplement for selected material or high-risk AI Systems, including through a pilot, additional exhibit, examiner resource, appendix, module, or other mechanism the Working Group considers appropriate.

1. Purpose and recommended use

This Profile supplies a system-level operational-assurance layer for the NAIC Artificial Intelligence Risk Evaluation Supplement. It is designed for the point at which Exhibit A identifies an AI System that warrants focused review because it has Direct Consumer Impact, Material Financial Impact, material financial-reporting impact, high inherent risk, material agentic capability, or another regulator-designated risk characteristic.

The Profile complements the Supplement, the NAIC Model Bulletin on the Use of Artificial Intelligence Systems by Insurers, and the Market Regulation, Financial Condition Examiners, and Financial Analysis Handbooks. It does not create an independent legal duty, redefine financial materiality, or require every control for every AI use. A regulator may invoke all or part of it, and should tailor evidence depth to the nature, scale, complexity, autonomy, reversibility, and potential impact of the selected system.

1.1 Conformance convention

Within this voluntary Profile, 'shall' identifies a criterion for demonstrating conformance after a regulator or insurer elects to use the Profile. It does not state that the criterion is independently required by law. Conformance is assessed for a specified AI System or use case, not inferred for an entire enterprise, and may be demonstrated through any technology, process, or architecture that produces reliable evidence of the required outcome.

For each applicable control, the response record should identify: conformance status; responsible owner; supporting document and page; date last tested; evidence location; material exceptions; and remediation status. Acceptable status values are Conforms, Partially Conforms, Not Demonstrated, and Not Applicable with rationale.

1.2 Applicability triggers

- Direct Consumer Impact, potential Adverse Consumer Outcomes, or material consumer-risk significance;
- Material Financial Impact, material financial-reporting impact, or material effect on risk or control assessment;
- high inherent risk under the insurer's documented risk assessment;
- material agentic capability, including multi-step planning, tool or API use, delegated permissions, or the ability to initiate or execute actions; or
- a regulator-designated use case, incident, complaint pattern, third-party dependency, or other examination concern.

2. Core definitions

The Supplement's definitions continue to govern. The following terms add operational precision for this Profile.

Term	Profile definition
Authorized institutional effect	A decision, communication, transaction, workflow change, record change, payment, denial, escalation, filing, or other consequence that the insurer recognizes as an official act or result.
Material agentic capability	The ability of an AI System to pursue an objective across multiple steps in a manner that can materially affect a consumer, financial condition, financial reporting, regulated process, or control environment.
Operational assurance	Evidence-based assessment of whether an AI System operated within approved purpose, authority, risk, performance, and control boundaries for a specified scope and review period, supported by continuing monitoring.
System dependency and action map	A current record linking an AI use case to material models, data sources, retrieval sources, tools, APIs, vendors, human roles, downstream systems, decisions, and executable actions.
Material change	A change to a model, data source, system instruction, workflow, connected tool, permission, provider, control, or operating context that could alter risk, performance, compliance, or institutional effect.
Decision-and-action reconstruction	A risk-proportionate ability to reconstruct the system, evidence, control context, authorization, output, action, and outcome associated with a selected event.
Safe state	A predefined operating condition that restricts or suspends sensitive action when control confidence, data integrity, authorization, system availability, or legal compliance is uncertain.

3. Normative principles

- **System-level object.** Review the complete decision-and-action system when model-level review would omit material orchestration, tool, permission, human, or downstream risk.
- **Risk-based proportionality.** Scale controls and evidence to inherent risk, materiality, consumer impact, autonomy, reversibility, deployment scale, and third-party dependence.
- **Accountable ownership.** Assign named business, risk, technology, and control ownership, with escalation to senior management and the board as appropriate.
- **Authority before effect.** Define and enforce who or what may recommend, approve, modify, execute, override, suspend, or prohibit an action before it produces institutional effect.
- **Evidence by design.** Preserve sufficient records to demonstrate that governance operated in practice, not only that policies existed.
- **Lifecycle assurance.** Validate before deployment and reassess after material change, incident, degradation, or other risk-significant event.
- **Outcome accountability.** Observe consumer, financial, reporting, operational, privacy, and security outcomes and remediate material deviations.
- **Third-party parity.** Do not reduce governance expectations merely because a model, data source, tool, platform, or professional service is supplied by a third party.
- **Governed change.** Subject material changes to documented authority, evaluation, approval, deployment, monitoring, and withdrawal or reversion procedures.

4. Operational assurance control objectives

The controls below are outcome-based. They specify what a selected AI System should demonstrate and the evidence an examiner may request; they do not prescribe implementation mechanics. TEVV means test, evaluation, verification, and validation of requirements and fitness for intended use.

Domain A - Scope, purpose, risk, and accountability

Control	Minimum conformance expectation	Representative examiner evidence
IA-01 System inventory	The insurer shall identify the AI System or use case, accountable owner, affected process, material models, data sources, tools, APIs, vendors, downstream systems, recommendations, and executable actions.	System inventory entry; system dependency and action map; current owner attestation; links to model and data inventories.
IA-02 Purpose and use boundary	The insurer shall document intended purpose, approved users and processes, allowed and prohibited uses, affected decisions, applicable legal and policy requirements, and conditions for continued use.	Approved use-case charter; requirements or obligations map; prohibited-use rules; business-owner approval.
IA-03 Inherent risk and materiality	The insurer shall assess inherent risk before controls and document the materiality and impact rationale used to scope governance and testing. Financial materiality remains governed by applicable handbook guidance.	Risk assessment; threshold rationale; consumer, financial, reporting, operational, privacy, security, and third-party impact analysis.
IA-04 Accountability and challenge	The insurer shall assign accountable business and control owners, define review and escalation authority, and provide independent challenge commensurate with risk.	Responsibility matrix; committee or board reporting; approval records; independent review or audit workpapers; unresolved-exception log.

Domain B - Authority, action boundaries, and resilience

Control	Minimum conformance expectation	Representative examiner evidence
IA-05 Authority and action boundary	The insurer shall define actions the system may recommend, initiate, or execute; actions requiring approval; prohibited actions; permission limits; delegation rules; and boundaries for tools, APIs, service accounts, and other agents.	Authority and permission matrix; approval rules; least-privilege review; boundary and denied-action test results.
IA-06 Human authorization and override	For risk-significant actions, the insurer shall identify who may approve, reject, modify, override, or escalate; when intervention occurs; the authority basis; and the evidence retained.	Authorization workflow; role delegation; sampled approvals and overrides; exception rationale; reviewer training or competency records.
IA-07 Reversibility and fallback	The insurer shall assess whether material actions are reversible and maintain risk-appropriate rollback, correction, compensation, manual fallback, consumer-remediation, and continuity procedures.	Reversibility assessment; rollback or compensating-control test; manual procedure; recovery exercise; consumer correction plan.
IA-08 Operating state and suspension	The insurer shall define risk-significant operating conditions and the authority to continue, restrict, require authorization, suspend, recover, or retire the system. Sensitive execution shall cease when required control conditions are not met.	Operating-state criteria; suspension authority; incident decision log; safe-state or restricted-operation test; recovery approval.

Domain C - Lineage, validation, and controlled change

Control	Minimum conformance expectation	Representative examiner evidence
IA-09 Lineage and configuration	The insurer shall maintain traceability among the AI System, material model and configuration versions, data and retrieval sources, third-party services, connected tools, and applicable control profile.	Version record; data dictionary and source lineage; model-to-data relationship; provider and tool register; configuration record.
IA-10 TEVV and independent review	Before production and at risk-appropriate intervals, the insurer shall test, evaluate, verify, and validate fitness for intended use, limitations, adverse outcomes, reliability, security, and material agentic behavior. Independence shall be proportionate to risk.	Validation plan and report; test cases and acceptance criteria; adverse-outcome and boundary tests; independent review; unresolved limitations.
IA-11 Material change control	The insurer shall define material-change criteria covering models, data, prompts or system instructions, workflows, tools, permissions, providers, controls, and operating context, and shall trigger documented approval, revalidation, monitoring, and withdrawal or reversion as appropriate.	Change record; impact assessment; approval; regression or revalidation results; deployment and enhanced-monitoring evidence; withdrawal decision.
IA-12 Third-party assurance	The insurer shall obtain sufficient information and contractual cooperation to govern and examine material third-party AI, data, tools, and services, including changes and incidents.	Due diligence; contractual audit and cooperation terms where appropriate; assurance reports; service-change notices; contingency or exit plan.

Domain D - Runtime evidence, outcomes, and institutional assurance

Control	Minimum conformance expectation	Representative examiner evidence
IA-13 Runtime monitoring and incidents	The insurer shall monitor performance, drift, degradation, data quality, boundary violations, control failures, incidents, near misses, complaints, and material consumer or financial signals at a frequency commensurate with risk.	Monitoring design; thresholds and alerts; exception trends; incident and near-miss records; complaint linkage; escalation and disposition evidence.
IA-14 Decision-and-action reconstruction	For a risk-based sample and significant events, the insurer shall be able to reconstruct the relevant system and model version, material evidence, applicable control, output, authorization or override, action, downstream result, and remediation.	Sampled event package; trace or record linkage; document and page references; authorization record; outcome and remediation evidence.
IA-15 Outcome review and remediation	The insurer shall compare actual consumer, financial, reporting, operational, privacy, and security outcomes with approved expectations and shall investigate, correct, compensate, disclose, retest, or suspend as warranted.	Outcome metrics; adverse-outcome analysis; root-cause review; corrective action; consumer remediation; retest and closure approval.
IA-16 Assurance, retention, and attestation	Management shall periodically assess whether controls operated effectively, whether exceptions remain acceptable, and whether the system remains fit for use. Evidence shall be retained under applicable legal, examination, recordkeeping, and confidentiality requirements.	Control assessment; management attestation; internal audit or independent-assurance report; retention schedule; examination index; open-risk acceptance.

5. Minimum operating-state model

An insurer need not use these labels, but a material system should support equivalent control conditions. This table states outcomes, not transition logic.

Condition	Illustrative trigger	Minimum response
Approved operation	Purpose, authority, evidence, performance, and control conditions are satisfied.	Proceed within approved boundaries and preserve ordinary evidence.
Enhanced monitoring	Elevated sensitivity, emerging anomaly, minor degradation, or increased uncertainty.	Continue only within approved scope with additional monitoring and evidence.
Authorization required	A decision or action meets a predefined approval, exception, or escalation threshold.	Hold institutional effect until authorized; preserve the decision and authority record.
Restricted operation	Some functions remain acceptable but a permission, action, data source, or workflow presents elevated risk.	Narrow function, tools, users, or effects; communicate restrictions; monitor exceptions.
Suspended or safe state	Control uncertainty, data-integrity failure, material incident, unauthorized behavior, or unavailable supervision.	Stop sensitive action; preserve evidence; notify accountable authority; use approved fallback.
Recovery or retirement	Corrective action is complete, continued use is no longer justified, or the system is withdrawn.	Authorize recovery after testing or retire safely; preserve records and address downstream effects.

6. Assurance evidence stages

Stages A1-A4 describe evidence maturity for a specified system and review period. They are separate from the five governance-maturity levels in the general Standard. They are not certifications, accreditation, regulatory approval, or a substitute for control-by-control assessment.

Stage	Designation	Demonstrated condition
A1	Documented foundation	System inventory, purpose, ownership, risk assessment, baseline policy, and third-party identification are current.
A2	Managed controls	Authority, validation, change, monitoring, incident, and evidence controls are defined and operating for the selected system.
A3	Operationally assured	The insurer can demonstrate control operation through current evidence, event reconstruction, outcome review, remediation, and management attestation.
A4	Independently assessable	Evidence quality and control operation support risk-appropriate independent review, internal audit, examination, or third-party assurance without reconstructing governance manually after the fact.

7. Crosswalk to NAIC v5.0 and recognized anchors

NAIC or NIST anchor	Profile controls	Operational contribution
Supplement Exhibit A - inventory, inherent risk, consumer and financial impact	IA-01 to IA-04	Extends model inventory to the material system, dependencies, actions, ownership, purpose, and scoping rationale.
Supplement Exhibit B - AIS Program, board, ERM/ORSA, autonomy, reversibility, reporting impact	IA-03 to IA-08; IA-16	Converts governance narrative into accountable authority, action-boundary, fallback, state, and attestation evidence.
Supplement Exhibit B - testing, explainability, vendors, ongoing effectiveness	IA-09 to IA-13	Adds system-level lineage, TEVV, material-change triggers, third-party cooperation, monitoring, and incident evidence.
Supplement Exhibit C - high-risk models, limitations, validation, monitoring	IA-01; IA-09 to IA-15	Preserves model review while connecting the model to the operating system, authorization, actions, outcomes, and remediation.
Supplement Exhibit D - model data and data dictionaries	IA-09; IA-10; IA-13	Connects data lineage and permitted use to validation, monitoring, and selected-event reconstruction.
NAIC Model Bulletin - proportional AIS Program, lifecycle governance, third parties, examination	IA-01 to IA-16	Provides an optional evidence profile for demonstrating that relevant governance and risk controls operate in practice.
NIST AI RMF 1.0 - Govern, Map, Measure, Manage	IA-01 to IA-16	Maps organizational governance to system context, measurement, monitoring, response, and continuing risk management.
NIST TEVV-Athlon draft framework	IA-10; IA-13 to IA-15	Supports requirements-driven evaluation of real-world system performance, impact, and outcomes, including agentic systems.

8. Options for NAIC consideration

The Working Group could consider the following options:

1. Optional use. Select relevant controls after Exhibit A scoping and allow equivalent evidence, without requiring the full Profile for every AI system.
2. Limited pilot. Test the Profile on a small set of materially different AI uses and evaluate examiner clarity, evidence availability, duplication, burden, and confidentiality.
3. Evidence calibration. Refine examples and response fields using pilot results while preserving technology neutrality, proportionality, alternative evidence, and handbook authority. The Working Group could adopt, modify, incorporate, or use selected portions as an additional exhibit, module, appendix, examiner resource, or other form it considers appropriate.

Interoperability note. Insurers and technology providers may map their existing governance programs and controls to the public Standard or this Profile. Such mapping does not by itself establish conformance, certification, regulatory approval, Legion endorsement, or operating effectiveness.

IMPLEMENTATION-NEUTRAL PUBLICATION CLAUSE This Profile describes governance capabilities, control objectives, and evidence outcomes. Conformance may be demonstrated through any system, process, or architecture capable of satisfying the applicable criteria. The Profile neither requires a particular product nor discloses a proprietary implementation.

Primary authorities

[1] NAIC Big Data and Artificial Intelligence (H) Working Group - v5.0 exposure notice and comment instructions (August 31, 2026).

[2] NAIC AI Risk Evaluation Supplement v5.0, reproduced in August 31, 2026 meeting materials, PDF pp. 12-34 (Supplement pp. 1-23). See Exhibit A, Exhibit B narrative item 1(g), Exhibit C item 3, and definitions.

[3] NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers (December 2023), Sections 3-4.

[4] NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0), January 2023.

[5] NIST AI 600-1, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, July 2024.

[6] NIST AI 200-2, The TEVV-Athlon Framework for Evaluating AI Systems, Initial Public Draft, August 2026.

© 2026 Legion Modular IP Holdings, Inc. All rights reserved. This comment and Attachment 1 are offered as implementation-neutral policy materials and do not grant any license to proprietary Legion implementation technology.

For more information

William C. Biddle

Chief Executive Officer

Legion Modular IP Holdings, Inc.

wcb@legionmodularip.com

Comments on the AI Risk Evaluation Supplement

Version 5.0 | Big Data and Artificial Intelligence (H) Working Group

Yash Lundia | September 19, 2026

Members of the Working Group:

I am developing a Responsible AI Trust & Compliance Platform for community banks and smaller insurers, with a focus on documenting AI testing and human oversight. I am commenting in my personal capacity and have a prospective commercial interest in this area. These comments do not represent an employer or insurer.

I support V5.0's approach to tailoring inquiries and using existing documentation. My recommendations concern how a reviewer connects answers across exhibits to the model, assessment and review decision being described. A few clarifications would make those connections easier to follow.

1. Make model references consistent across exhibits

Exhibit A, Part Two, Model Inventory; Exhibit C, references 1 and 2; Exhibit D, system or model identification.

The draft already asks for model names and versions. I would add an instruction to use those references consistently across exhibits. A model can support several use cases, and a use case can continue under the same name after its underlying model changes. Without a clear link, an inventory entry and a testing report can appear to describe the same deployment when they do not.

The model version may not fully identify what was tested. For example, the same version could be deployed with different decision thresholds or input preprocessing. Results from one configuration may not apply to another. Where that distinction matters to the inquiry, an existing deployment or configuration reference would help establish the scope of the assessment.

Suggested addition to Exhibit A Part Two

Use a consistent model reference in related responses to Exhibits A, C and D. An existing inventory identifier or unambiguous model name may be used. Identify the relevant model version and, where needed to interpret the response, the deployment or configuration assessed. Explain the relationship where one model supports multiple AI Systems or use cases. If version information is unavailable, state that limitation and provide the deployment information known to the company.

The instruction could sit beside the existing inventory fields, preserving Exhibit A's counting categories and the company's existing identifiers.

2. Connect testing results to the review decision

Exhibit C, references 11 and 12; Exhibit B, Checklist reference 3 and document/page columns; Exhibit B, Narrative discussion of human involvement and testing.

Exhibit C asks about testing, validation and ongoing monitoring, including the last testing date. Exhibit B addresses governance and human involvement. These questions provide a useful basis for review. I suggest asking the company to link the assessment to the resulting review decision, so that the reader can follow what happened after a finding was raised.

Consider a hypothetical pricing model whose output distribution changes after a vendor update. A monitoring report records the change, but the reviewer still needs to know whether the insurer investigated it and what remains unresolved. The change alone does not establish unfair discrimination or a pricing error. A record of human review should explain the conclusion reached, including any limits on that conclusion.

Suggested addition after Exhibit C reference 12

For the testing or monitoring described in reference 11, identify the model version or deployment assessed, the testing date or period, and a reference to the supporting record. Where requested and proportionate to the inquiry, summarize relevant findings and limitations, the responsible review role, the review outcome or current status, and any follow-up action. Distinguish planned testing from completed testing, and completed review from resolution of a finding. Existing records may be cross-referenced where the regulator determines they remain current, applicable and sufficient for the inquiry.

A report identifier and section, or a reference to an existing review record, may be sufficient. The supporting record should make the assessment's scope clear, including the period and population tested where relevant. A policy explains how review should work. The record of a particular review shows what was considered and decided.

Suggested clarification to Exhibit B Checklist instructions

When citing an existing document, identify the applicable version or date and the relevant page, section or record reference. Where the regulator requests evidence of how a control operated, cross-reference the relevant operational record as well as the policy describing the control.

The level of detail should reflect the risk and scope of the inquiry. I would leave testing frequency to the applicable requirements and circumstances rather than introduce a universal schedule through these instructions. A summary can point to supporting records when needed, subject to the regulator's authority and applicable confidentiality protections.

A worked example would help establish the expected level of detail. I suggest tracing one finding from the test record through review to a documented resolution or an outstanding action, using records an insurer could already maintain.

3. Distinguish unavailable vendor information from confirmed nonuse

Exhibit D, Company Instructions; Exhibit B, Checklist reference 3p; Exhibit C, references 5, 6 and 9.

Exhibit D directs the company to leave a field blank if a listed data source is not used in model development. With a third-party model, however, the insurer may be unable to determine whether the source was used. The instructions should distinguish that uncertainty from a confirmed absence of the source.

For example, a vendor might document the inputs its service receives during use while withholding information about the data used to develop the model. Documentation of the service inputs would not answer the development-data question. The response should identify the basis for the insurer's answer and any information it still lacks.

Suggested replacement for the blank-field instruction in Exhibit D

Leave the field blank only where the company has determined that the listed data source is not used in development of the identified model. Where use of the source is unknown or the information is unavailable, state that explicitly and explain the basis. Identify any relevant request to a third party that remains pending or was not answered. If the company considers an item outside the scope specified by the regulator, explain the basis and seek clarification rather than leaving the field blank.

Exhibit B's third-party oversight response is a suitable place to explain how the company is handling the gap. Exhibit C's model-limitations response could be cross-referenced where it already contains that explanation.

Suggested explanatory note for Exhibit B reference 3p

Where information relevant to third-party oversight is unavailable, describe the limitation, the steps taken to obtain or assess the information, the role responsible for reviewing the limitation, and any follow-up action. Distinguish vendor statements from the company's own verification. The regulator may seek further information or determine that additional action is warranted.

Reporting information as unknown would make the limitation visible; it would not establish adequate oversight. The appropriate response depends on the significance of the missing information and the model's use. This clarification concerns the accuracy of the company's account of its data sources. It should not encourage collection of sensitive consumer data merely to complete the template.

These changes would help reviewers follow the supporting records while preserving the flexibility of V5.0. Thank you for considering my comments. I would welcome the opportunity to discuss the proposed wording with the Working Group.

Respectfully,
Yash Lundia

Reference: NAIC AI Risk Evaluation Supplement, version 5.0. Exhibit and reference numbers refer to the exposure draft. Indented passages are proposed wording. The examples are hypothetical.
<https://content.naic.org/committees/h/big-data-artificial-intelligence-wg>



September 29, 2026

Commissioner Houdek, Chair - Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106-2197

RE: AI Risk Evaluation Supplement version 5.0 comment period

Dear Commissioner Houdek and Members of the Working Group:

Monitaur appreciates the opportunity to submit written feedback on the 5th version of the AI Risk Evaluation Supplement. Over the course of the prior versions and execution of a Pilot of the Supplement, regulators and carriers have developed greater understanding of the opportunities and challenges to regulate AI's use across insurance.

We are pleased to see the significant progress that has been made since the initial version. We believe the large majority of the concepts in this version are consistent with fundamental AI governance, model governance, and broader risk management practices we expect of the industry, irrespective of regulatory guidance. However, there are several specific concepts either newly introduced, or continued from prior versions, which we recommend are refined to remove ambiguity and aid in the effective use of the Supplement.

Our comments first identify the revisions that materially improve the Supplement and then address one remaining issue: inconsistent use of "AI System" and "AI Model" across the instructions, exhibits, and definitions.

We are available for any clarifications or further input as you find it helpful.

Positive Changes

Several overarching changes materially reduce ambiguity for the industry and enable stronger cross-sectional comparisons of AI governance programs from a regulatory perspective. We particularly welcome the clearer statement of the Supplement's intent and scope within market conduct and financial examinations; the addition of materiality language; the incorporation of inherent vs residual risks; the use of the AI System as the primary unit of governance; and by naming generalized linear models (GLMs) within the



Supplement's taxonomy. Together, these revisions improve clarity, align the Supplement with inter-industry best practices, and demonstrate careful consideration of stakeholder feedback.

Areas for Improvement

Overarching recommendation: We do not agree that the introduction of the concept of an AI model is necessary. Through our customers and community, we have experienced strong alignment organizing everything from GLM's through LLM's and agents under the term of an AI System. A System takes an action, a model by itself does not take an action: it informs action taken by an individual or system.

The Background, Intent, and Instructions sections appropriately focus on AI Systems, including the goal of enhancing regulators' understanding of AI System use and potential risk to consumers and insurers. Exhibits A and B generally follow that approach, while Exhibits C and D shift toward model-level terminology without consistently explaining the purpose of that shift.

We recommend the following drafting hierarchy: an **AI System** is the primary unit of governance and reporting; a **model** is a component of an AI system; a **stand-alone model** may be examined separately when it independently creates material consumer or financial risk. Model-level information may be relevant to validation, drift, explainability, or financial examination, but it should be requested when relevant to evaluating a particular AI System or when a stand-alone model has a material impact.

We continue to believe that this *risk* evaluation Supplement should not be used as a general survey instrument. Model information may be relevant as supporting technical information in a market conduct or financial examination. When regulators determine that such information is necessary, the Supplement should request it in relation to the relevant AI Systems rather than allowing model type to determine the primary unit of inventory or risk assessment. Specific recommendations for Exhibit A appear below.

Aggregated Recommended revisions

1. Designate the AI System as the primary unit of inventory and risk evaluation.
2. Replace "AI Model" with "AI System" in the identified questions unless referring to a model or models which are components of an AI System.



3. State when and why examiners may request model-level information.
4. Standardize inventories as AI System inventories. Models used by carriers which may be material, but which are not a part of an AI System that performs actions should not be in scope of an AI focused evaluation.
5. Revise Exhibits C and D and their instructions to distinguish system-level review from component model review
6. Harmonize the definitions with these revisions.

Section Comments

Background, Intent, and Instructions

- We support defining the AI System as the primary unit of governance in the background section because adverse consumer outcomes and material financial impacts arise from the end-to-end operation of an AI System.
- Clarifying where the Supplement will be used is helpful to the industry, including the statement that the Supplement is supplemental to, and governed by, the applicable examination handbooks.
- The introduction of inherent risk is a welcome clarification. Distinguishing inherent from residual risk has been a recurring source of uncertainty, and inherent risk is an appropriate starting point for initial analysis.
- We appreciate the inclusion of GLMs within the Supplement's model taxonomy.
- The distinction between machine learning and AI in the Background section may be counterproductive unless it is necessary to a specific regulatory purpose. We recommend clarifying or removing that distinction to reduce ambiguity.

Exhibit A

- Exhibit A should consistently use the AI System as the primary unit of inventory and risk classification. Model-level information should appear as supporting information tied to the relevant AI System, which can be part of Exhibit C. Exhibit A, Part 2, item 3 is thus redundant, and we would recommend the section be removed.
- We recommend replacing “**Number of AI Models with Direct Consumer Impact**” and “**Number of AI Models with Material Financial Impact**” with AI System terminology. An AI System may contain one or more models, while the material consumer and financial outcomes under examination arise from the system's operation. Model-level information may be relevant as supporting information, but it



should not define the primary reporting unit. If the NAIC retains this information, it could request the proportion of AI Systems using each model category, including GLMs and other identified categories.

- Separate “generative AI” and “agentic AI” in the reporting taxonomy. Generative AI describes a class of outputs or underlying techniques, whereas Agentic AI is a type of AI System. An Agentic System can have many steps and component model types from GLM’s to LLM’s.

Exhibit B: (Narrative) AIS Program

- The prompt “Discuss the process by which the effectiveness of the *AIS Program* and individual models are assessed and modified” should refer to AI Systems unless the NAIC specifically intends to request model-level validation information. Question G and Section 2 already use AI System terminology.
- Section 3 consistently refers to AI Systems except in item (d): “Discuss how the company ensures explainability and transparency into any *AI models* recommending or making decisions.” We recommend using “AI Systems” here or explaining why model level information is required.
- This change would improve consistency while recognizing that explainability, interpretability, and transparency should be evaluated in relation to the material consumer or financial outcomes produced by the relevant AI System.

Exhibit B: (Checklist) AIS Program

- Stays consistent with AI Systems, which we agree with. No mention of AI Models.
- Our opinion is that the checklist and response version of Exhibit B is significantly more efficient for all parties.

Exhibit C: High-Risk AI Models

- We recommend using the AI System as the default unit of review in Exhibit C because the relevant regulatory outcomes arise from the operation of the AI System. Where a component model or stand alone model requires separate examination, the instructions should identify that circumstance expressly.
- The current instructions introduce additional ambiguity. We recommend focusing clearly on AI System level examination and review of stand alone models (not part of an AI System) with material consumer or financial impact.



- Since all AI Systems include some form of model, it would make sense to ask about the model as part of the System, but the edits have incorrectly elevated the model and therefore it excludes considerations like how security might have been implemented on the System or how the data used by the model and System was evaluated for appropriateness.

Exhibit D: AI Model Data

- If Exhibit D is to remain in this Supplement, we recommend AI System data for all of the reasons mentioned above; however, we remain of the opinion that the objectives of Exhibit D might be better handled through some other guidance or practice that is focused on permissible data and/or rating factors, versus this supplement focused on AI Systems.

Definitions and Appendix

- **“Adverse Consumer Outcome”** refers to an *AI System* decision (output) by an insurance company that is subject to insurance regulatory standards enforced by the Department that adversely impacts the consumer in a manner that violates those standards”.
- We agree with keeping the AI Model definition, but using AI System exclusively in the exhibits. Additionally, we recommend dropping “AI” from this model definition.
- We recommend revising the AI System definition to align with the proposed hierarchy:
 - **“AI System** is a machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, content (including text, images, videos, or sounds), or other outputs that influence decisions in real or virtual environments. An *AI System* may include one or more models and may operate with varying levels of autonomy.”
- **“Model Drift”** refers to the decay of a model’s performance over time arising from underlying changes such as the definitions, distributions, and/or statistical properties between the data used to train the model and the data on which it is deployed.”
 - We would recommend revising this definition to: AI System or Outcome Drift
- **Grammatical Recommendations:**
 - **“Augmentation”** refers an *AI System* that suggests an answer and/or advises a human who is making a decision.
 - Is missing a “to” in refers to an AI System...



- **“Neural Network Model”** refers to a machine learning model that mimic[sic] the complex functions of the human brain.
 - Is missing an “s” from mimics
- Direct consumer impact mentions system as well, which we agree with

As always, my team and I are at your service to assist in ensuring reliable, trustworthy, fair, and performant AI Systems for the insurance industry. If you have any questions on the contents of this submission, and if clarification, or draft definitions would be helpful, do not hesitate to ask.

Best regards,

Andrew Clark, Ph.D.

Co-founder and CTO

Monitaur

September 29, 2026

Commissioner Houdek (WI), Chair
NAIC Big Data and Artificial Intelligence (H) Working Group
c/o Miguel Romero, Director of Innovation, Cybersecurity and Technology
via email maromero@naic.org

Re: Big Data and AI (H) Working Group Version 5 of the AI Risk Evaluation Supplement

Chair Houdek and Members of the Working Group:

On behalf of the National Association of Mutual Insurance Companies (NAMIC)¹, we are writing in response to the exposure of the AI Risk Evaluation Supplement, Version 5 (hereinafter “Supplement”). Below we detail substantive and thematic concerns, and attached are suggested redlines for the Working Group to consider.

[1] GLMs Should be Removed from Scope of the Supplement

In an early version of the Supplement, GLMs were a defined term and explicitly excluded from scope through the Supplement’s definition of Artificial Intelligence. NAMIC supported this exclusion, and continues to recommend that the Working Group exclude GLMs and other predictive models that result in the use of static fixed formulas or rules. In version 5, under General Guidance, the Supplement states that GLMs are a machine learning technique, and “as with other machine learning techniques, GLMs are not without risk of causing unfair discrimination or other adverse consumer outcomes,” and are included in the Exhibit A quantification chart as a separate column.

Setting aside the technical debate about when or whether a GLM might start to have characteristics of a machine learning model or otherwise, the ultimate risks and uses of GLMs and other predictive models have been adequately addressed and handled with existing regulatory tools for decades. Including GLMs in the Supplement means that companies would submit information on a technology that has been in widespread use and subject to regulation since the 1990s. This would make the Supplement less valuable to regulators because the GLM data would make it harder to identify the truly novel uses of artificial intelligence that

¹ The National Association of Mutual Insurance Companies (NAMIC) is the foremost trade association representing the property/casualty insurance industry. Serving more than 1,300 member companies—including local and regional insurers as well as some of the nation’s largest carriers—NAMIC members collectively write \$467 billion in annual premiums, representing 61% of the homeowners and 53% of the automobile insurance markets. For more than 130 years, NAMIC has been the leading voice advancing public policy solutions and regulatory frameworks that promote a strong, competitive market and protect our members and their policyholders.

have arisen in recent years. Reporting on a 30-year-old technology deployment would distort the data and be unduly burdensome for companies.

The Supplement states that the concern with GLMs is ultimately “unfair discrimination” or “adverse consumer outcomes” which, as defined, are both areas of the law that insurance departments have been monitoring and reviewing GLM and predictive model use through existing tools. The original purpose of the Supplement was to address gaps in available tools for regulators to assist in review of AI Models. Because the Working Group has not established regulatory deficiencies over the last 30 years of GLM use and other predictive model use, their inclusion in the scope of the Supplement (focused on purported new risks that new AI models may represent) is overly broad. The inclusion of GLMs also risks duplicative and conflicting oversight and review metrics, and diverts regulator attention away from the original focus on purported new risks from emerging AI technologies.

GLMs result in static fixed scoring formulas which are not self-learning nor self-updating. Also, GLMs used for insurance pricing specifically are already being regulated, as parts of rate filing reviews. Some states even use special rate filing questionnaires to collect information about pricing models and their development in a consistent way. GLMs and other predictive models are simply tools that help insurers to make rates more closely aligned with expected cost, as is required by state laws and actuarial principles.

Within the set of intertwined definitions, the draft might be confusing the model development process with the model result. The process of developing a GLM or other type of predictive model uses data and iterative calculations to gradually improve an arithmetic formula or rules which will be used to determine estimates or predicted values. This iterative development process could be interpreted or viewed as “learning” or “updating”, although that only occurs during the singular and intentional model development process. But the resulting model itself (a final scoring formula or set of rules that is actually used to determine the estimate or prediction) is a static fixed formula or set of rules which does not change.

A revised model could be developed a year or month later using more recent data, to develop a new static fixed formula, to be able to calculate more accurate estimates or predictions. However, at least in the context of P&C pricing models, those models cannot be changed until they are filed with the state insurance departments. These models – which are used to help determine insurance premiums – do not change unless and until they are filed, and the companies and insurance departments know the formulas or rules that make up the models.

In addition to the foregoing, removing GLMs from scope of the Supplement would be consistent with approaches taken abroad by regulatory bodies like EIOPA. For many of the same reasons noted above, EIOPA excluded both GLMs and GAMs from the AI Act’s high-risk classification. From a risk standpoint, EIOPA concluded that human-supervised GLMs and GAMs do not pose the type of AI-specific risk that would justify their inclusion in scope, and that risks associated with GLMs and GAMs are already addressed through existing Union law and insurance sector supervision.² EIOPA further concluded that to include GLMs and

² Technical Annex on the exclusion of GLMs and GAMs from the AI Act high-risk classification – EIOPA’s further

GAMs in scope “would create material compliance burdens for widely used and well-understood and established statistical techniques,” as well as divert supervisory resources³ away from novel AI systems that may pose more specific AI-related risks. The NAIC’s risk-based approach to supervision has traditionally been a point of positive differentiation from European counterparts and has underpinned the growth of the world’s largest insurance market. NAMIC urges the Working Group not to adopt an approach that is even more burdensome than Europe’s.

Given the foregoing, NAMIC therefore recommends that GLMs be explicitly removed from scope of the Supplement.

[2] New Materiality Threshold

In the Regulator Instructions to Exhibit A, there is an addition that states “Regulators should either specify a materiality threshold to guide company responses to the Supplement or request that the company provide a threshold that it used in completing this Exhibit.” Because “materiality” is not a defined standard in existing law, NAMIC recommends that any materiality threshold be determined by the company rather than the insurance department. While we appreciate the attempt at narrowly tailoring scope to impacts that are material, we are concerned that the standard is not tied to existing law and authority that departments have today.

[3] Suggested Changes to Exhibit A Table – AI Systems vs. Models, “Other AI/ML Models”

The table in Exhibit A has been modified to separate out counts of “AI Systems” currently in use, and “AI Models” with Direct Consumer Impact and Material Financial Impact. The “AI Models” section is further broken down into Generalized Linear Models, Generative/Agentic AI Models, and “Other AI/ML Models.” Because “AI Models” comprise “AI Systems,” NAMIC is concerned with the use of both terms as quantification metrics in the chart. Including counts for both “AI Systems” and “AI Models” risks duplicative counting for purposes of completing the table. It is also unclear what regulators are looking for with this distinction. NAMIC members are unsure how to distinguish between AI Systems and AI Models. The current definitions do not sufficiently distinguish these concepts and instead create confusion. Moreover, the request for counts of AI Systems is not limited to systems with direct consumer or financial impact and is, therefore, outside the stated purpose of the Supplement. NAMIC therefore recommends removing columns two and three from Exhibit A, and that the Working Group remove the AI Systems definition and all uses of the term given the confusion that results with use of both AI Systems and AI Models.

Notwithstanding the concern over AI Systems vs. AI Models, NAMIC also suggests edits to the identified model groups. First, Generative AI models and Agentic AI models pose different risks, so grouping them together may result in a quantification that raises unnecessary alarm. Grouping them together also does not

input, EIOPA, 30 April 2026, p. 3, https://www.eiopa.europa.eu/publications/letter-and-technical-annex-exclusion-glms-and-gams-ai-act-high-risk-classification_en.

³ Id., 5.

give regulators an accurate view into how many of those models are actually agentic – much of the current conversation appears to overstate how much insurers are using agentic AI models. NAMIC’s suggestion (if AI Models are to be used for the table) is to utilize three categories: “Predictive AI Models,” “Generative AI Models,” and “Agentic AI Models.” These categories are the currently recognized AI Model types, thereby fostering consistency in how companies report the AI Model types, as well as greater clarity into the types of Models being used by each company.

From a procedural standpoint, NAMIC also notes that the substantial changes reflected in Version 5 Exhibit A have not been tested or piloted in this form. To the extent the Working Group wishes to vote on the Supplement for adoption, we would have reservations about these substantive changes. We believe that our suggested changes are more substantively aligned with what was piloted.

[4] Exhibits C and D

From a substantive perspective, NAMIC continues to suggest the removal of Exhibits C and D. Exhibit C is duplicative of content in Exhibits A and B. For instance, Exhibit A will give regulators a view into the types of models being used, and Exhibit B will provide information to the regulator about how a company is managing the risk of its AI model use. These aspects seemingly cover the bulk of the substance sought in Exhibit C. For example, Exhibit C asks companies to explain how the model is reviewed for compliance with applicable state and federal laws, and Exhibit B (Ref. 3b.) asks companies to point to processes and procedures that “ensures AI Systems are compliant with applicable state and federal laws and regulations.” Further, “high-risk AI model” is not a defined term in existing law. Exhibit C relies on other terms that are also not found in existing law in an effort to define what constitutes “high-risk” – terms like “material impact” and “adverse consumer outcomes.” Therefore, NAMIC suggests the removal of Exhibit C in its entirety.

NAMIC also suggests removing Exhibit D in its entirety. This information would not be informative in the context of an exam assessing financial and solvency risk, and these models would already be assessed or subject to review when filed (as well as subject to reviews conducted in market conduct examinations when assessing compliance with rating laws and unfair trade practices). Further, Exhibit D creates new standards focused on data and third-party use, and the Third-Party Data and Models Working Group is currently undertaking an effort focused on third parties. As such, the inclusion of Exhibit D is also premature without the work product from the Third-Party Data and Models Working Group.

Separately, based on NAMIC member feedback, it does not appear that the majority of the piloting states utilized Exhibits C and D. Therefore, from a procedural standpoint alone, NAMIC does not believe that these Exhibits are ripe for adoption.

[5] Redline Suggestions Attached



NAMIC has attached redline suggestions to go along with our substantive comments. We look forward to continued collaboration with the Working Group to arrive at a product that meets the needs of the market.

Sincerely,

Lindsey Stephani

Lindsey Stephani
Senior Policy Vice President
Data Science, Artificial Intelligence, and Cybersecurity
NAMIC

Artificial Intelligence (AI) Risk Evaluation Supplement

Background, Intent, and Instructions

Background:

The rapid expansion of big data and the adoption of artificial intelligence and machine learning are significantly transforming insurance practices. These technologies can offer substantial benefits to both insurance companies and consumers by facilitating the development of innovative products, improving customer interface and enhancing service, simplifying and automating processes, and promoting efficiency and accuracy. However, without robust governance and effective controls, the use of *AI Systems* may lead to Adverse Consumer Outcomes or compromise the financial soundness of an insurance company. Insurers are responsible for managing the risks associated with the development and implementation of *AI Systems* and must demonstrate to regulators that appropriate risk-based oversight mechanisms are in place and are functioning effectively. Ultimately, insurers that make decisions or perform actions impacting consumers or financial soundness that are made or supported by advanced analytical and computational technologies, including from *AI Systems*, must comply with all applicable insurance laws and regulations.

Intent:

The NAIC's Innovation, Cybersecurity and Technology (H) Committee charged the Big Data and AI Working Group (BDAIWG) to create tool(s) that would assist regulators to identify and assess both financial and consumer risks from the use of *AI Systems* on an on-going basis.

The following exhibits are designed to supplement existing market conduct, financial analysis, and/or financial examination review procedures for reviewing *AI Systems*. As this supplements existing NAIC resources, regulators should continue to consider existing NAIC resources as authoritative but may consider drawing from this supplement to assist in understanding and assessing a company's use of *AI Systems*. Inquiries and information requests performed related to this supplement will be coordinated consistent with the guidance provided by the *Market Regulation Handbook*, *Financial Condition Examiners Handbook*, and the *Financial Analysis Handbook*.

With the intention of the document as a supplement acknowledged, there may be some important context that is not specified in this document, particularly related to deciding which companies are in scope of inquiry. This is to avoid the perception of conflicting guidance as many important decisions leading to the full use of this Supplement would be governed by the aforementioned Handbooks.

These optional exhibits allow regulators to determine the extent of *AI Systems* usage for an insurance company and whether additional analysis is needed focusing on financial and consumer risk.

This Supplement consists of the following exhibit questionnaires:

- **Exhibit A: Quantify Regulated Entity's Use of AI Systems**

- **Exhibit B: AIS Program (Two Options: Narrative or Checklist)**
- **Exhibit C: High-Risk AI Models**
- **Exhibit D: AI Model Data**

Instructions:

Information obtained from the responses to the following Exhibits may supplement guidance and tools used during an existing market conduct, certification, financial analysis, and/or financial examination review, to enhance the regulator’s understanding of the extent of *AI Systems* utilization and assessment of potential risk to consumers and the insurance company. Effective assessment requires regulators to maintain a fluent understanding of the applicable laws and regulations, including but not limited to those pertaining to unfair trade practices, unfair claims settlement practices, corporate governance annual disclosures, confidentiality, financial reporting, and rating.

Regulators using this Supplement may wish to first use Exhibit A to determine if further inquiry using Exhibits B-D is necessary. It is possible that while the company responding indicates usage of *AI Systems*, its use of AI is limited or low in inherent risk such that further inquiry as contemplated by subsequent exhibits is not warranted. For instance, if a regulator is using the Supplement to inquire about a company’s claims practices and based on responses to Exhibit A, the regulator determines that the company’s use of AI is out of scope of a claims-related targeted exam, the regulator may determine it is appropriate to thereafter rely on the Market Regulation Handbook to guide the remaining work.

If information requested through this Supplement has already been provided to this department or any other state department of insurance, the company’s response should so state, in which case the regulators may accept prior submissions if the prior response is still current and applicable.

The responses to these exhibits will be considered by regulators when identifying the inherent risks of the insurer. They should also affect the planned examination or inquiry approach, as well as the nature, timing and extent of any further procedures performed.

General Guidance

Machine Learning (ML) Models

Machine learning models are widely used to estimate the relationship between a dependent variable and one or more independent variables. For example, an insurer may want to estimate how a driver’s age relates to the probability of being involved in a car accident. There are many ML models that can be used for this purpose.

~~In particular, generalized linear models (GLMs) are one such technique. As with other machine learning techniques, GLMs are not without risk of causing unfair discrimination or other adverse consumer outcomes and require effective governance over data quality, model development, validation, implementation, ongoing performance monitoring, and compliance with applicable laws and regulations.~~

Commented [LS1]: Because NAMIC suggests removal of GLMs from the scope of the Supplement, and also suggests categorically grouping AI Models into 1) Predictive AI Models, 2) Generative AI Models, and 3) Agentic AI Models, the Working Group may also consider deletion of this Machine Learning Models section in its entirety.

Commented [LS2]: For the reasons noted in our substantive comments, NAMIC recommends removal of GLMs from scope of the Supplement.

DRAFT

AI Risk Evaluation Supplement

2

This supplement asks about a company's use of specific models in Exhibit A to give regulators flexibility in determining their scope of interest.

Materiality

~~This Supplement was designed to allow regulators the flexibility to either specify materiality or rely on the company's materiality calculations as companies respond to Supplement related inquiries.~~

The concept of *materiality* leverages the definition otherwise used to drive financial exam-related decisions and as defined in the FCEH. For purposes of completing responses related to this Supplement, regulators ~~may opt to~~ should rely on a company's internal materiality thresholds. ~~provided~~ Companies should provide the information supporting a company's *materiality threshold*, including an explanation for why that threshold was considered appropriate.

Commented [LS3]: Because "materiality" is not a defined standard in existing law, NAMIC recommends that any materiality threshold be determined by the company rather than the insurance department.

However, nothing in the guidance provided within this Supplement should be construed to supersede FCEH guidance.

Risk Assessment

This Supplement also, at times, relies on a company's Risk Assessment. The concept of Risk Assessment centers around the discussion of *Inherent Risk* which leverages the definition reflected in the FCEH. For purposes of completing responses to this Supplement, ~~regulators that Risk Assessment~~ information provided to regulators should include explicit indication of *Inherent Risk* Assessments (i.e., risk before the consideration of controls or mitigation measures).

Consistent with the discussion of *materiality*, nothing in the guidance provided within this Supplement should be construed to supersede FCEH guidance.

Confidentiality

The information requested via these Exhibits will be held confidential, consistent with the requesting state's relevant authority.

Regulators using any of these exhibits should cite examination or other authority as appropriate when requesting information from insurers. Regulators should cite all relevant confidentiality statutes or other specific protections related to documents, materials or other information in the possession or control of regulators that are obtained by or disclosed to the regulators or any other person in the course of a market conduct inquiry and all information reported or provided to the regulator pursuant to cited examination or other authority.

Exhibit A: Part One - Quantify Regulated Entity's Use of AI Systems

Purpose: To obtain information pertaining to the number of *AI Systems* and *AI Models* used by a company to help facilitate risk assessment. Based on the responses from the company, regulators may ask for additional information related to their *AI Program* (Exhibits B), high-risk models (Exhibit C), and the types of data (Exhibit D) where there is risk for *Adverse Consumer Outcomes* or material adverse financial impact.

Company Instructions:

- Provide the most current counts and use cases of the following, as requested.
- For purposes of responding to information requests related to this Exhibit, models that *augment* or *automate* decision-making related to consumers are considered to have direct consumer impact.
- “*AI System*”, “*Adverse Consumer Outcome*”, and “*Use Case*” are defined in the “Definitions and Appendix” section of this Supplement.
- Include all lines of business. If the governance differs by company, line of business, or state, work with your regulator to determine if multiple submissions are needed.
- If a company has an existing *AI Systems* inventory, it may suggest submission of the inventory in lieu of completing Exhibit A.
- If a company has questions about any of the fields, please submit them in writing to **<regulator contact email address>**.

Regulator Instructions:

- Regulators should customize this Exhibit to limit information requested to more targeted inquiries for use in a limited scope exam.
- Regulators should **either specify a materiality threshold to guide company responses to the Supplement or request that the company provide a threshold that it used in completing this Exhibit.**
- **For column 3, regulators should specify the number of months.**

Company Legal Name and Group Name: _____

NAIC CoCode and Group Code: _____

Company Contact Name: _____ Email: _____

Line(s) of Business for Which This Response Applies: _____

Materiality Threshold <regulator to provide or specify if the company will provide>: _____

Commented [LS4]: See NAMIC's comment above - we suggest that regulators request the materiality threshold and explanation from the insurer.

Commented [LS5]: NAMIC encourages a defined term/number of months be agreed upon for purposes of the Supplement.

Date Form Completed: _____

DRAFT

(1) Use of AI System in Operations or Program Area	(2) Number of AI Systems Currently in Use	(3) Number of AI Systems Implemented in the Past X Months	Number of AI Models with Direct Consumer Impact			Number of AI Models with Material Financial Impact			(10) Use Case(s)
			(4) Generalized Linear Models (GLMs) Predictive AI Models	(5) Generative/Agentic AI Models	(6) Other AI/ML Agentic AI Models	(7) Generalized Linear Models (GLMs) Predictive AI Models	(8) Generative/Agentic AI Models	(9) Other AI/ML Models Agentic AI Models	
Marketing									E.g., UC1: Identify potential consumers interested in product.
Premium Quotes & Discounts									
Underwriting/Eligibility									
Ratemaking/Rate Classification/Schedule Rating/Premium Audits									
Claims/Adjudication*									
Customer Service									
Utilization Management/Utilization									

Commented [LS6]: As NAMIC explains in greater detail in our substantive comments, we are concerned that inclusion of both “AI Systems” and “AI Models” will result in confusion and duplicative counting, as models comprise AI systems.

NAMIC requests clarification on what regulators are looking for in establishing this distinction, and how regulators will address the risk of duplicative counting. If neither can be clearly established, NAMIC suggests reverting the quantification table back to its form as it was in Version 4. As a separate option, the Working Group might consider deleting columns 2 and 3 and utilizing the AI Model Columns with the suggested edits by NAMIC.

Commented [LS7]: The three recognized categories of AI Models are currently: 1) Predictive AI Models; 2) Generative AI Models; and 3) Agentic AI Models. As such, NAMIC encourages adhering to these buckets if AI Models are kept in the Exhibit A chart.

Commented [LS8]: Given the Working Group’s apparent focus on higher-impact functions like ratemaking and underwriting, NAMIC suggests that “Marketing” be removed from the Exhibit A table.

(1) Use of AI System in Operations or Program Area	(2) Number of AI Systems Currently in Use	(3) Number of AI Systems Implemented in the Past X Months	Number of AI Models with Direct Consumer Impact			Number of AI Models with Material Financial Impact			(10) Use Case(s)
			(4) Generalized Linear Models (GLMs) Predictive AI Models	(5) Generative/Agentic AI Models	(6) Other AI/ML Agentic AI Models	(7) Generalized Linear Models (GLMs) Predictive AI Models	(8) Generative/Agentic AI Models	(9) Other AI/ML Models Agentic AI Models	
Review/Prior Authorization/Level of Care Determination									
Fraud/Waste & Abuse									
Investment/Capital Management									
Reserves/Valuations									
Catastrophe Triage									
Reinsurance									
Other Insurance Practices** (if applicable)									

* Includes Salvage/Subrogation, adjuster assignments, etc.
 ** Any additional operation or program area with the potential to result in either Direct Consumer Impact or Material Financial Impact.

Commented [LS6]: As NAMIC explains in greater detail in our substantive comments, we are concerned that inclusion of both “AI Systems” and “AI Models” will result in confusion and duplicative counting, as models comprise AI systems.

NAMIC requests clarification on what regulators are looking for in establishing this distinction, and how regulators will address the risk of duplicative counting. If neither can be clearly established, NAMIC suggests reverting the quantification table back to its form as it was in Version 4. As a separate option, the Working Group might consider deleting columns 2 and 3 and utilizing the AI Model Columns with the suggested edits by NAMIC.

Commented [LS7]: The three recognized categories of AI Models are currently: 1) Predictive AI Models; 2) Generative AI Models; and 3) Agentic AI Models. As such, NAMIC encourages adhering to these buckets if AI Models are kept in the Exhibit A chart.

Commented [LS9]: NAMIC suggests removal of this category; the Working Group should explicitly note the operational areas it has in mind.

Exhibit A: Part Two – AI Systems Request List

In Support of the regulatory inquiries related to this Supplement, the regulators request information or documentation related to the following:

Regulator Instructions: Regulators should customize this Exhibit to limit information requested to more targeted inquiries for use in a limited scope exam:

1. ~~Materiality~~ – Explanation related to the company’s *materiality* calculation (if applicable) used in responding to inquiries related to this Supplement.
2. ~~Risk Assessment~~ – Documentation explaining the company’s risk assessment process for purposes of its use of *AI Systems*.
3. ~~Model Inventory~~ – A listing of the company’s models, that specifies risk for the following program areas <regulator to specify program areas in scope of the request>. Additionally, regulators request the model inventory or accompanying documentation include:
 - a. ~~AI Model Name~~
 - b. ~~Description of how the AI Model is used~~
 - c. ~~Description of the broader Use Case in which each AI Model is used~~
 - d. ~~Operation or Program Area(s) the AI Model is Used Within~~
 - e. ~~Inherent Risk Level Based Upon the Company’s Risk Assessment~~
 - f. ~~What consumer impact does this AI Model have?~~
 - g. ~~What financial impact does this AI Model have?~~

Commented [LS10]: The original purpose of Exhibit A was to be used as an introductory inquiry providing background information for a regulator to help determine if additional inquiry using the remaining exhibits is necessary. This new part 2 expands the scope of Exhibit A and the burden associated with providing this information. Further, this portion was never piloted.

As such, NAMIC strongly recommends removal of this new section, from both a procedural and substantive perspective.

Commented [LS11]: Notwithstanding our concern over introduction of this new part overall, NAMIC requests more detail from the Working Group on why the Risk Assessment portion is necessary here, given that Exhibit B is meant for use in obtaining information on risk management and governance.

Commented [LS12]: Notwithstanding our concern over introduction of this new part overall, NAMIC sees the model inventory requirement as duplicative of the information and work a company would already have provided in the Exhibit A table. If an inventory like this is requested, a company should not also need to fill out the table, and vice versa.

Exhibit B: (Narrative) AIS Program

Purpose: To obtain the Company's *AIS Program*, including how the company handles governance, risk identification, mitigation, and management framework and internal controls, and third-party *AI Systems* and data for *AI Systems*; and the process for acquiring, using, or relying on third-party *AI Systems* and data. Market and financial regulators should coordinate to gain access to the relevant section of the policies governing the use of *AI Systems*.

Company Instructions: Below are both Checklist and Narrative sections of Exhibit B of the NAIC's AI Risk Evaluation Supplement. If the responses to the checklist indicate that an explanation of a policy, procedure or process can be found on a given page of a document, indicate the page number and the document title and include the document referenced in the submission. If other sections of the document are irrelevant, the relevant sections of the document may be submitted in lieu of the entire document. Please see definitions at the end of this document.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam.

The references to, and questions about, elements of a company's *AIS Program* are not intended to be interpreted as creating new requirements for *AIS Program*.

Group and Company Legal Name: _____

NAIC Group and Company CoCodes: _____

Company Contact Name: _____ Email: _____

Date Form Completed: _____

1. Provide your *AIS Program*. [Click or tap here to enter text.](#)
 - a. What role maintains the *AIS Program*? [Click or tap here to enter text.](#)
 - b. Discuss the governance structure, Board reporting and frequency. [Click or tap here to enter text.](#)
 - c. Discuss the process by which the *AIS Program* is integrated throughout the organization, assessed and remediated. [Click or tap here to enter text.](#)
 - d. Discuss the process by which the effectiveness of the *AIS Program* and individual models are assessed and modified. [Click or tap here to enter text.](#)

Commented [LS13]: The Working Group has previously stated that regulators would use either the narrative portion or the checklist portion of Exhibit B, but the company instructions appear to suggest the intention is for companies to complete both. NAMIC therefore suggests adding clarification here that companies are not expected to complete both.

- e. Discuss how responsibility for governance within the organization is assigned and how the organization ensures consistency and alignment. [Click or tap here to enter text.](#)
 - f. Discuss the integration of the *AI Systems* in the Own Risk and Solvency Assessment (ORSA) and Enterprise Risk Management (ERM) assessments. [Click or tap here to enter text.](#)
 - g. How does the insurance company assess autonomy, reversibility, and reporting impact risk of *AI Systems*? [Click or tap here to enter text.](#)
2. Discuss the uses of *AI Systems* that:
- a. Generate a *material financial impact*. [Click or tap here to enter text.](#)
 - b. Generate a *direct consumer impact*. ~~The company's response should include, but not be limited to, whether there is a human in the loop, whether there are high volume transactions, how any exceptions or errors are found and resolved, and whether there were any staffing reductions as part of the *AI System* implementation.~~ [Click or tap here to enter text.](#)
 - c. Generate or impact material information reported in financial statements. [Click or tap here to enter text.](#)
 - d. Generate or impact risk or control assessment. [Click or tap here to enter text.](#)
 - e. Discuss the development, testing, and implementation of material *AI Systems* that the Company has implemented. If appropriate, include details regarding where any systems differ from established IT systems and data handling protocols. Discuss the basis for deviation from established practices. [Click or tap here to enter text.](#)
3. Provide the policy for, and discuss the use and oversight of, material *AI System* vendors, model design and testing:
- a. Discuss the validation and testing procedures performed on internally-developed *AI Systems*, including whether validation and testing are performed by someone that is independent from development. [Click or tap here to enter text.](#)
 - b. Discuss the validation and testing procedures performed on third-party vendor-supplied *AI Systems*. [Click or tap here to enter text.](#)
 - c. Discuss the testing and verification that has occurred including frequency, scope and methodology. [Click or tap here to enter text.](#)
 - d. Discuss how the company ensures explainability and transparency into any *AI models* recommending or making decisions. [Click or tap here to enter text.](#)
4. Provide the policy for, and discuss the use and oversight of, material *AI Systems* by third-party providers including actuarial, claim, MGA, audit, and/or other professional services. [Click or tap here to enter text.](#)
- a. Discuss the testing and verification that has occurred, frequency, scope, and methodology. [Click or tap here to enter text.](#)
5. Discuss additional aspects of the framework design and evaluation pertaining to *AI Systems*. [Click or tap here to enter text.](#)

Commented [LS14]: The Supplement interchanges use of "material impact" and "direct consumer impact." NAMIC suggests use of one term for consistency throughout.

Commented [LS15]: NAMIC suggests removal here. The new inclusion of human in the loop here is arguably beyond existing law. NAMIC is also unclear of the purpose of the staffing inquiry in the context of a market conduct or financial exam.

- a. Discuss the unit(s) responsible for the *AI Program*, assessment approach and frequency, and involvement with the program area to the extent it differs from that discussed previously. [Click or tap here to enter text.](#)

DRAFT

Exhibit B: (Checklist) AIS Program

Purpose: To obtain information about the Company's *AIS Program* including information about how the company handles governance, risk management and internal controls, and third-party *AI Systems* and data for *AI Systems*; and the process for acquiring, using, or relying on third-party *AI Systems* and data. The references to, and questions about, elements of a company's *AI Systems* Governance Risk Assessments are not intended to be interpreted as creating new requirements for *AI Systems* Governance Risk Assessments.

Company Instructions: Provide responses to the questions regarding how the governance of *AI Systems* fits within your company's system of supervision or Enterprise Risk Management program. Include all companies and lines of business. If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See [definitions](#) below.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam.

Group and Company Legal Name(s): _____

NAIC Group and Company Code(s): _____

Company Contact Name: _____ Email: _____

Date Form Completed: _____

Ref	AI Systems Use Questions for Company	Company Response
1	Has the company adopted a written <i>AIS Program</i> ? If yes, when was it adopted and what is the frequency of review for updating?	
2	Was the Board of Directors or management involved in the adoption of an <i>AIS Program</i> ?	
	2a. What is the role of the Board of Directors or management in the <i>AIS Program</i> ?	

Ref	AI Systems Use Questions for Company	Company Response	
3	Reference the processes and procedures of the Company <i>AIS Program</i> that addresses the following:		
	How the Insurance Company...	Document Name and Page #	If not specified in the <i>AIS Program</i> , provide details below:
	3a. Assesses, mitigates, and evaluates <i>inherent risks</i> related to the use of <i>AI System</i> including the possibility of the company engaging in unfair trade practices		
	3b. Ensures <i>AI Systems</i> are compliant with applicable state and federal laws and regulations		
	3c. Evaluates the risk of <i>Adverse Consumer Outcomes</i>		
	3d. Considers data privacy and protection of consumer data used in <i>AI Systems</i>		
	3e. Evaluates whether <i>AI Systems</i> are suitable for their intended use and should continue to be used as designed		
	3f. Considers <i>AI System</i> risks within its Enterprise Risk Management (ERM)		
	3g. Considers <i>AI System</i> risks within the Own Risk and Solvency Assessment Report (ORSA), as applicable.		
	3h. Considers <i>AI System</i> risks within the software development lifecycle (SDLC) or within the broader <i>AIS Program</i>		
	3i. Considers <i>AI System</i> risk impact on financial reporting		
	3j. Trains employees about <i>AI System</i> use and defines prohibited practices (if any)		
	3k. Quantifies <i>AI System</i> risk levels		
	3l. Provides standards and guidance for procuring and engaging <i>AI System</i> vendors		

	3m. Considers consumer complaints resulting from AI Systems and whether they are identified, tracked, and addressed		
	3n. Promotes consumer awareness of the use of AI Systems through disclosures, policies, and procedures for consumer notification, as appropriate		
	3o. Determined “materiality” and what threshold was used for assessing AI Systems risk.		
	3p. Governs, monitors, tests and ensures transparency of AI Systems developed by vendors.		

Commented [LS16]: The NAIC does not require the tracking of AI complaints and it may not be clear that AI has contributed to a complaint. As such, NAMIC suggest removal of this provision.

Exhibit C: High-Risk AI Models

Purpose: To obtain detailed information on in-production high-risk *AI models*, such as models making automated decisions that could cause *Adverse Consumer Outcomes*, *material financial impact*, *material consumer impact*, or material financial reporting impact. *AI model* risk criteria are set by the insurance company. Note that *materiality* should be a consideration in determining the level of risk. To assist in identifying models for which this information is requested, regulators may request information on the company's risk assessment and a model inventory if such information has not otherwise already been provided.

Company Instructions: Fill in the details for each of the *AI model(s)* requested. Include all companies and lines of business. If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See definitions below. The template below refers to both *AI Systems* and *AI Models* depending on the information being requested. There may be some instances where a company feels information should be provided in relation to the *AI System* and not the *AI Model* or vice-versa. This should be discussed with regulators as part of the submission process to avoid misunderstanding.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam.

Group and Company Legal Name(s): _____

NAIC Group and Company Code(s): _____

Company Contact Name: _____ Email: _____

Date Form Completed: _____

Ref	AI Model Information Requests	Company Response
1	AI model name and version number.	
2	Use cases and purpose(s) of model	
3	Model type used in the AI System (GLM, generative/agent AI, or other AI/ML algorithm)	
4	Model Implementation Date	
5	Model development (internal, "owned" third party, or true third party, or a mix of both – include vendor name if applicable)	

Commented [LS17]: For reasons noted in our substantive comments, NAMIC recommends the removal of Exhibit C in its entirety.

Commented [LS18]: Notwithstanding our overarching suggestion that Exhibit C be removed, NAMIC requests that the WG clearly distinguish what is meant by "owned" third party and "true" third party. Additionally, the WG should clarify whether insurers must report embedded vendor functionality and AI used by service providers acting on the insurer's behalf.

The WG should also collaborate with existing efforts across the NAIC that are looking at Third Parties to ensure there is no conflict between definitions.

6	Did the Company purchase the <i>AI model</i> , was the <i>AI model</i> developed specifically for the Company? Is the Company able to modify the <i>AI model</i> as necessary?	
7	Model risk classification (high, moderate, low, etc.)	
8	Model risk(s) (i.e., describe the potential for <i>adverse consumer impact</i>)	
9	Model limitations (i.e., limitations on <i>uses cases</i> or data inputs outside the domain of the training data, or other guardrails to manage model related risk)	
10	<i>AI model</i> type (automate, augment, support)	
11	Discuss testing model outputs (e.g. model drift, accuracy, unfair trade practices, unfair discrimination, performance degradation, etc.) and how the model was validated prior to being deployed as well as how its performance is monitored on an ongoing basis.	
12	Last date of model testing or validation.	
13	Discuss how the model affects the financial statements, risk assessment or controls.	
14	Discuss how the model is reviewed for compliance with applicable state and federal laws, including but not limited to the unfair trade practices act and unfair claims settlement laws.	
15	To the extent permitted by law, discuss if the company has had any actions taken against them for use of this model. Actions may include, but are not limited to, informal agreements, voluntary compliance plans, administrative complaints, ongoing third-party monitoring, cease and desist, remediation, restitution, fines, penalties, investigations, consent orders or other regulatory agency actions.	

Commented [LS19]: Notwithstanding our overarching suggestion that Exhibit C be removed, this field is unnecessary. Exhibit C is only for models the carrier has designated as high risk.

Exhibit D: AI Model Data

Purpose: To obtain detailed information of the source(s) and type(s) of data used in an *AI model(s)* to identify risk of *adverse consumer impact*, unfair trade practices, material financial impact, or material financial reporting impact.

Company Instructions: Provide details below for the data used in *AI model(s)*. If any of the data elements listed are used in the training or test data as part of the development of an *AI model*, provide information on whether the data element is sourced internally or whether the data element is sourced from a third party, in which case provide the name of the third-party vendor. Leave blank if a data source is not used in the development of *AI model* for the insurance operation. Include all companies and lines of business. If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See [definitions](#) below.

Regulator Instructions: Regulators should customize this supplement to limit information requested to more targeted inquiries for use in a limited scope exam. Regulators may also wish to customize this supplement for the line of business as some data types may or may not be pertinent based on the responding company's lines of business. Depending on responses received, a regulator may wish to ask for a data dictionary to gain a more complete understanding of a data used for a given model.

Group and Company Legal Name(s): _____

NAIC Group and Company CoCode(s): _____

Company Contact Name: _____ Email: _____

List the *AI System* or *AI models* which rely on the data described in response to this submission:

Describe the Line of Business for Which This Response Applies (complete one for each line of business):

Date Form Completed: _____

Commented [LS20]: For the reasons noted in our substantive comments, NAMIC recommends removal of Exhibit D in its entirety.

	(1)	(2)	(3)	(4)
Ref	Type of Data Element Used in AI Model	Describe how the AI Model uses this data including how the output is used within operations; or input data is used for training and testing	Internal Data Source	Third Party Data Source / Vendor Name
1	Aerial Imagery			
2	Age, Gender, Ethnicity/Race			
3	Consumer or Other Type of Insurance/Risk Score			
4	Crime Statistics			
5	Criminal Convictions (Exclude Auto-Related Convictions)			
6	Driving Behavior			
7	Education Level (Including school aptitude scores, etc.)			
8	Facial or Body Detection / Recognition / Analysis			
9	Geocoding (including address, city, county, state, ZIP code, lat/long, MSA/CSA, etc.)			
10	Geo-Demographics (including ZIP/county-based demographic characteristics)			
11	Household Composition			
12	Image/video Analysis			
13	Income			
14	Job History			
15	Loss Experience			
16	Medical, including Biometrics, genetic information, pre-existing conditions, diagnostic data, etc.			
17	Natural Catastrophe Hazard (Fire, Wind, Hail, Earthquake, Severe Convective Storms)			

18	Online social media, including characteristics for targeted advertising			
19	Personal Financial Information			
20	Reasonable Accommodations or Policy Modifications (granted or requested)			
21	Telematics/Usage-based insurance			
22	Vehicle-Specific Data including VIN characteristics			
23	Voice Analysis			
24	Weather			
25	Other: Non-Traditional Data Elements (Regulator to specify)			

DEFINITIONS AND APPENDIX

Where available, for the purposes of this evaluation terms are defined in accordance with the NAIC Model Bulletin on the Use of AI Systems by Insurers (https://content.naic.org/sites/default/files/2023-12-4%252520Model%252520Bulletin_Adopted_0.pdf):

“Adverse Consumer Outcome” refers to an *AI System* decision (output) by an insurance company that is subject to insurance regulatory standards enforced by the Department that adversely impacts the consumer in a manner that violates those standards.

“Agentic AI” refers to *AI Systems* that can pursue objectives across multiple steps without requiring human input at each stage, which could involve perceiving the environment, deciding what to do next, taking actions (often through tools or APIs), observing/analyzing the results, and/or adjusting the action accordingly. The difference is in agency – the system does not just respond to a single prompt or instruction, but actively executes a process.

“Algorithm” means a clearly specified mathematical process for computation; a set of rules that, if followed, will give a prescribed result.

“AI Model” refers to an integral component of an *AI System* that learns from data to perform tasks via computational, statistical and/or machine learning techniques. It combines algorithms, training data and learned parameters to produce outputs from a given set of inputs, such as making predictions or decisions on data it has never seen before. For purposes of this Supplement, Generalized Linear Models are not considered AI Models.

“AI System” is a machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, content (such as text, images, videos, or sounds), or other output influencing decisions made in real or virtual environments. *AI Systems* are designed to operate with varying levels of autonomy. For purposes of this Supplement, Generalized Linear Models are not considered AI Systems.

“AIS Program” refers to the controls and processes that an insurer adopts and implements to ensure the responsible use of *AI Systems*. An *AIS program* may be comprised of general guidelines, governance framework, risk management and internal controls, and third-party oversight.

“Artificial Intelligence (AI)” refers to a branch of computer science that uses data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement, or the capability of a device to perform

Commented [LS21]: NAMIC notes that “Adverse consumer impact” and “Adverse consumer outcome” are terms that are used interchangeably throughout the Supplement. The Working Group should uniformly use one term, and define that term.

Commented [LS22]: There is not enough distinction between “AI Model” and “AI System” to necessitate both terms for purposes of the Supplement. NAMIC suggests that the Working Group identify which is of greatest concern and why, and utilize only one of these terms consistently throughout the Supplement.

functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement. This definition considers machine learning to be a subset of artificial intelligence.

“Augmentation” refers an *AI System* that suggests an answer and/or advises a human who is making a decision.

“Automation” refers to an *AI System* that does not involve human intervention.

“Consumer Impact” refers to a decision by an Insurer **specifically impacting consumers** that is subject to insurance regulatory standards enforced by the Department.

“Direct Consumer Impact” refers to an *AI System* that *augments* or *automates* human decision-making whose output has the potential to create final consumer outcomes including but not limited to underwriting including tier and coverage eligibility determinations, premium determinations, ratemaking, and claims denial and settlement-related decisions. This has less applicability to *AI Systems* used for internal analytics, model monitoring, and back-office tooling with no consumer-visible output or other internal operational or administrative functions that do not determine or materially affect an outcome for a specific consumer.

“Externally Trained Models” refers to transferred learnings from pre-trained models developed by a third party on external reference datasets.

“Generative Artificial Intelligence (Generative AI)” refers to a class of *AI Systems* that generate content in the form of data, text, images, sounds, or video, that is similar to, but not a direct copy of, pre-existing data or content.

“Generalized Linear Model (GLM)” refers to a statistical modeling technique that extends linear regression to by relating predictive variables to an outcome variable using an assumed additive or multiplicative relationship through a link function and an appropriate probability distribution.

“Inherent Risk” refers to the risk of economic loss or inaccurate financial reporting before considering internal controls.

“Internally Trained Models” refers to models developed from data internally obtained by the company.

“Machine Learning (ML)” refers to a field within artificial intelligence that focuses on the ability of computers to learn from provided data without being explicitly programmed.

“Material Financial Impact” should be determined based on either the company’s materiality threshold which should be disclosed when responding to Supplement related inquiries or based on the threshold to be provided by regulators.

Commented [LS23]: NAMIC requests revision of the “augmentation” definition for clarity. As it currently reads, augmentation refers to an AI system; augmentation is not a system, but an aspect or attribute of a process.

Commented [LS24]: NAMIC requests revision of the “automation” definition for clarity. As it currently reads, automation refers to an AI System; automation is not a system, but an aspect or attribute of a process.

Further, “automates human decision making” does not align with the definition of “automation” above. That definition says automation does not involve humans.

Commented [LS25]: NAMIC suggests this inclusion because, as written, the definition is too broad and captures decisions that do not impact consumers specifically.

Commented [LS26]: NAMIC suggests clarifying the consumer impact definition as proposed here. Without clarification, “consumer impact” could include decisions about how to build a financial statement, for example, as these are subject to insurance regulatory standards enforced by a Department. Yet, the financial statement example does not have a “consumer impact” in the way the Working Group has discussed their concerns around “consumer impact.”

“Materiality” refers to the dollar amount above which the examiner’s perspective of an insurer’s financial position will be influenced.

“Model Drift” refers to the decay of a model’s performance over time arising from underlying changes such as the definitions, distributions, and/or statistical properties between the data used to train the model and the data on which it is deployed.

“Neural Network Model” refers to a machine learning model that mimic the complex functions of the human brain. These types of models consist of interconnected nodes or neurons that process data, learn patterns and enable tasks such as pattern recognition and decision-making, including but not limited to: Single/multi-layer perceptrons/fully connected networks (MLPs/FCs), Deep Learning (DL), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory Neural Networks (LSTMs), Sequence Models, Large Language Models (LLMs), and Reinforcement Learning Models (RLs).

“Predictive Model” refers to the mining of historic data using algorithms and/or machine learning to identify patterns and predict outcomes that can be used to make or support the making of decisions.

“Residual Risk” refers to an assessment of risk after considering risk-mitigation strategies or controls.

“Support” refers to an *AI System* that provides information but does not suggest a decision or action to a human.

“Third Party” for purposes of this Supplement means an organization other than the insurance company that provides services, data, or other resources related to AI.

“Validation Method” refers to the source of the reference data used for validation, whether Internal, External, or Both.

“Use Case” refers to a description of a specific function in which a product or service is used.

Commented [LS27]: As explained in greater detail in our substantive comments, NAMIC encourages the Working Group to delete the definition of “materiality” and ground the “materiality” determination or standard in company-based thresholds tied to existing examination guidance and authority.

Operations

Claims - Examples: Claim assignment, triage/fast-tracking, individual/bulk claim reserving including loss estimation, imaging/video analysis, fraud detection, litigation, estimation of closure rates, salvage/subrogation, examination/report ordering.

Customer Service - Examples: Agent/broker/internet/customer service interaction (chatbots), online/smart phone apps, loss prevention/risk mitigation advice, payment plans, complaints.

Marketing - Examples: market research, target advertising, market/coverage expansion, customer segment target marketing, demand modeling, agent/broker incentive plans, up/cross-selling.

Other: Cyber Security, Strategic Operations, Reserving, Investments, Capital Management, Financial Reporting, Reinsurance, Legal, Legal Exposure, Reputation Risk.

Ratemaking/Pricing - Examples: Development of overall/base rates, expense/loss loadings, estimation of trends and loss development, development of manual rating factors, tiering criteria, insurance credit scoring, territory boundary definitions, numeric/categorical level groupings and interactions, individual risk rating, telematics/UBI, price optimization, schedule rating factors.

Underwriting - Examples: Policy/coverage acceptance or eligibility, company placement/tiering, schedule rating, decisions based on telematics/UBI, report ordering, retention modeling, inspections, anomaly detection.

Comments on the Application of the NAIC Artificial Intelligence Risk Evaluation Supplement to Generalized Linear Models Used for Pricing

September 29, 2026

Background

The National Association of Insurance Commissioners (NAIC) Innovation, Cybersecurity and Technology Committee charged the Big Data and AI Working Committee to create a tool that would assist regulators with identifying and assessing financial and consumer risks from the use of AI Systems. The result of this charge is the proposed Artificial Intelligence (AI) Risk Evaluation Supplement, which is designed to supplement existing market conduct, financial analysis or financial examination review procedures.

Within the scope of this review are Machine Learning (ML) models, and more specifically Generalized Linear Models (GLM). As stated in the *General Guidance* of the AI Risk Evaluation Supplement, “GLMs are not without risk of causing unfair discrimination or other adverse consumer outcomes and require effective governance over data quality, model development, validation, implementation, ongoing performance monitoring, and compliance with applicable laws and regulations.”

We believe that GLMs used specifically for the purpose of ratemaking should be excluded from the scope of the AI Risk Evaluation. We believe this for two reasons. First, the characteristics of GLMs are such that they do not meet the definition of AI. Second, we believe that all the concerns outlined in the AI Risk Evaluation Supplement are already assessed and covered as of the current rate regulatory process. As a result, including GLMs used for the purpose of pricing would be a duplication of effort in most states.

Characteristics

We believe that the characteristics of a GLM are such that the need for review under an AI framework is not necessary. GLMs, by their very nature, are:

- **Transparent**: the formulas underlying a GLM are understandable, and the process by which data is input into a GLM, used in model development, and translated into model parameters can be easily traced from start to finish. Thus, the GLM process and the resulting output is explainable.
- **Deterministic**: because the formula is a set for a GLM, if one inputs the same data into a GLM, the same results will be obtained.

These characteristics of a GLM are inconsistent with the definition of AI in the AI Risk Evaluation Supplement. The specific applicable definitions are:

- **AI Model**: refers to an integral component of an AI System that learns from data to perform tasks via computational, statistical and/or machine learning techniques. It combines algorithms, training data and learned parameters to produce outputs from a given set of inputs, such as making predictions or decisions on data it has never seen before.
- **AI System**: machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, content (such as text, images, videos, or sounds), or other output influencing decisions made in real or virtual environments. AI Systems are designed to operate with varying levels of autonomy.
- **Artificial Intelligence (AI)**: refers to a branch of computer science that uses data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement, or the capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement. This definition considers machine learning to be a subset of artificial intelligence.

There are multiple aspects of these definitions that are not true of GLMs.

- A GLM does not learn from data. A GLM takes in data, performs calculations, and returns an output which is strictly the result of this calculation. However, there is nothing the GLM will “learn” as part of this process that will change the underlying GLM formulas. The GLM will apply the same formula every time to produce the same output.
- GLMs use a specific formula and parameters developed based on an analysis to generate predictions on a new dataset. These predictions are deterministic, and there is no room for the algorithm to produce anything other than the result of the prediction formula. The GLM will not

produce recommendations or content.

- GLMs do not operate with any autonomy. There is no room for systems to independently perform a different calculation other than what is specified by the GLM formula. GLM runs are initiated by a user and the output returned. The results of a GLM in a pricing analysis are then used to produce rating factors that are used as part of a rating formula. So, the results are not used directly, but used to justify selections that are then reviewed by a regulator. Therefore, there is no autonomy in this process, and the final application always requires multiple levels of human intervention.
- A GLM is formulaic and deterministic, therefore there is no reasoning, learning or self-improvement involved. It involves arithmetic functions that will be reproduced exactly every time a GLM is run.

Therefore, we believe there is a strong argument that a GLM violates multiple aspects of the AI definition, and therefore should not be covered under the AI Risk Evaluation.

Current Regulation

The use of GLMs in insurance pricing analyses began in the 1990's. GLMs have been used for pricing for over 30 years, and as a result, insurance companies have been filing GLMs as support for rate development for decades. This has led to a significant level of understanding amongst regulators regarding the development, interpretation, and evaluation of GLM models. This has included the development of a framework for reviewing GLM models by the NAIC, an adoption of GLM questionnaires and model review frameworks by a number of states, and the review of individually filed models by the NAIC model review team for individual states.

We believe that the concerns regarding a GLM outlined in the AI Risk Evaluation Supplement are already covered by existing regulation. Specifically, the items described as needing effective governance are already covered in the CASTF Predictive Model White Paper, which was released on September 9, 2020. The items needing effective governance and the applicable sections of the CASTF white paper are shown in the table below.

Effective Governance Over	CASTF Predictive Model White Paper
Data quality	A.1, A.3, A.4
Model development	A.2, B.1, B.2, B.3, B.6
Validation	B.4, B.5
Implementation	C.1, C.2, C.3, C.4, C.5, C.6, C.7, C.8
Ongoing performance	
Compliance with applicable laws and regulations	C.9

It is clear from this mapping that the majority of the concerns raised in the NAIC Risk Evaluation Supplement are more than adequately covered in the NAIC White Paper. Of the concerns listed, the only one that is not covered by the CASTF white paper is the concern related to the ongoing performance of the model. We believe this concern is addressed in practice by frequent company reviews of GLM and pricing plans and filings that occur by most companies at least annually, if not more frequently.

In addition to a number of states relying on the NAIC model review framework, there are also a number of states that have specific checklists or processes that are followed when companies file models, including GLMs. These states include, but are not limited to, Arizona, Michigan, Florida, Connecticut, California, Washington and New York.

Given the significant amount of existing regulation related to the use of GLMs in pricing, we believe that the additional requirements of the NAIC AI Risk Evaluation would be duplicative.

Conclusion

As a result of the characteristics of a GLM not being consistent with the definition of AI, and due to existing regulatory reviews covering the concerns highlighted regarding GLMs, we believe GLMs developed specifically for the purpose of supporting rating factors should be excluded from the NAIC AI Risk Evaluation. GLMs do not function as an AI System, and this additional regulation would be duplicative.

Respectfully Submitted,

A handwritten signature in blue ink that reads "Roosevelt C. Mosley, Jr." The signature is written in a cursive style with a large initial 'R'.

Roosevelt C. Mosley, Jr, FCAS, MAAA, CSPA
Managing Principal
Pinnacle Actuarial Resources, Inc.

From Pamela Hann (Point32Health)

Hello Scott and Miguel,

We have reviewed the **AI Risk Evaluation Supplement v5.0** and have several questions. Below, I've included the relevant requirements from the supplement, with our questions listed under each item.

Exhibit A: Part 2 - 3f

What consumer impact does this AI Model have?

- Please clarify the definition of "consumer". Does it include any external users of the model, such as customers, potential customers, or providers?

Exhibit B: Narrative - 2c

Generate or impact material information reported in financial statements.

- Does the regulator expect all AI Systems that support financial reporting processes to be included, or only those that impact financial statements meeting the established materiality threshold?

Exhibit C: Company Instructions

Fill in the details for each of the AI model(s) requested. Include all companies and lines of business. If the governance differs by entity, line of business, or state, work with your domestic regulator to determine if multiple submissions are needed. See definitions below. The template below refers to both **AI Systems** and AI Models depending on the information being requested. **There may be some instances where a company feels information should be provided in relation to the AI System and not the AI Model or vice-versa.** This should be discussed with regulators as part of the submission process to avoid misunderstanding.

- Should this be updated to refer only to the AI Model, since the rest of Exhibit C refers to the model entity?

Exhibit C: 5

Model development (internal, "owned" third party, or true third party, or a mix of both – include vendor name if applicable)

- How should companies distinguish between "Owned Third Party" and "Mix of Both"? Specifically, if a company retrains, fine-tunes, or otherwise modifies a third-party model, should the model be reported as "Owned Third Party" or "Mix of Both"?

Exhibit C: 7

Model risk classification (high, moderate, low, etc.)

- Does this risk classification apply only to inherent risks related to economic loss or inaccurate financial reporting?

Exhibit C – 11

Discuss testing model outputs (e.g., model drift, accuracy, unfair trade practices, unfair discrimination, performance degradation, etc.) and how the model was validated prior to being deployed as well as how its performance is monitored on an ongoing basis.

- Do regulators expect companies to provide only the methodology, the methodology with representative scores, or a full set of validation and monitoring metrics, including thresholds, for each reported model?

Please let me know if you have any questions or need further clarification.

PUBLIC COMMENT - TECHNICAL AND REGULATORY IMPLEMENTATION OBSERVATIONS

Project OV Comments on NAIC AI Risk Evaluation

Supplement Version 5.0

Decision-Level Governance, Progressive Friction, and Reconstruction

Date: September 28, 2026

To: Nathan Houdek, Chair, Big Data and Artificial Intelligence (H) Working Group, National Association of

Insurance Commissioners

Submitted to: Scott Sobel, AI Policy Advisor; Miguel Romero, Director of Innovation, Cybersecurity and

Technology

From: Brad Rothrock, Founder, Project OV

Project OV submits the following comments on Version 5.0 of the AI Risk Evaluation Supplement. The

Supplement appropriately increases attention to AI-system and model inventories, materiality, direct

consumer impact, explainability and transparency, third-party models, testing and monitoring, data, and

the distinction between model risks and model limitations.

These comments focus on a related implementation question:

For consequential AI-assisted decisions, how can an insurer - and, when appropriate, a regulator -

demonstrate that documented governance actually operated in the decision that occurred?

Policies, inventories, testing programs, governance frameworks and model documentation are necessary. For a

material decision, however, those records may not by themselves show what occurred at the point where

evidence, AI output, uncertainty, human judgment, institutional authority and actual execution came together.

Project OV is a research and engineering project focused on evidence governance and decision reconstruction

for consequential AI-assisted institutional decisions. It does not determine coverage, legal liability or ultimate

truth, and it does not replace regulators, insurers, actuaries, claims professionals, model developers, evaluators

or enterprise-risk systems. The narrower objective is to preserve enough decision-state evidence that

consequential decisions can later be reviewed, challenged and reconstructed.

In that architecture, AI may contribute evidence, analysis and recommendations; institutional authority must

remain separately established.

1. From Documented Governance to Demonstrable Governance

The Supplement asks insurers about AIS Program governance, materiality, consumer impact,

human

involvement, testing, model risks and limitations, third-party systems, data and related controls.

Those inquiries

help establish whether governance has been designed and maintained.

For selected consequential decisions, a complementary question is whether that governance can be

reconstructed in operation. A representative decision record could answer questions such as:

- What governing policy, contract, procedure or control applied?
- What material evidence existed at the time, and what evidence was actually used?
- Which model, evaluator or AI-assisted process contributed?
- What material uncertainty, warning or limitation was known?
- Who or what possessed relevant authority?
- Where did a consequential decision fork occur?
- What human review was required?
- What was recommended, approved, institutionally authorized, attempted and actually executed?
- What outcome followed?
- What changed when later evidence, a challenge or a new limitation arrived?

This does not require recording every model token, technical event or intermediate computation.

The objective

is a minimum sufficient governance record: enough information to reconstruct a consequential decision

without imposing unnecessary burden on ordinary activity.

2. Progressive Friction, Agility, and the Minimum Viable Governance Unit

Human oversight should not mean that a person must manually review every AI-assisted activity. That would be

impractical in many environments and could undermine legitimate efficiency and agility.

Instead, human involvement should be risk-proportionate and progressively escalated. Routine, low-

consequence and readily reversible activity may proceed with lightweight automated controls.

As materiality,

uncertainty, autonomy, irreversibility, authority mismatch, missing evidence, model limitations or third-party

dependence increase, the amount of governance friction should also increase.

To preserve agility, Project OV's architecture uses the concept of a minimum viable governance unit: the

smallest current decision state sufficient to determine whether progression remains permissible and to support

later reconstruction if the activity becomes consequential.

Depending on context, that unit may contain:

- actor or process identity and task;
- current authority;

- applicable obligation or governing rule;
- material evidence and uncertainty;
- known model or evaluator limitations;
- the current decision fork or gate;
- the contemplated action and its reversibility; and
- pointers to deeper retained evidence if later reconstruction is required.

The minimum viable governance unit is not a full case file and is not intended to turn every routine operation into

a heavyweight audit event. The operating principle is to capture the minimum sufficient state, escalate

selectively, and reconstruct deeply only when needed.

Progressive friction can therefore operate through several levels:

- Routine state: automated assistance, ordinary monitoring and lightweight controls remain sufficient.
- Heightened state: additional evidence, narrower permissions, further validation or specialist review may be required.
- Consequential fork: materially different consequential paths become explicit, and the system evaluates whether autonomous progression remains permissible.
- Qualified human review: where judgment, exception handling or institutional authority is required, the matter is routed to an appropriately qualified and authorized person.

- Pause or refusal: if required evidence, authority, review or another material precondition is absent, the affected transition can be held rather than silently treating the absence as approval. The objective is not to put humans everywhere. It is to concentrate meaningful human judgment at the consequential points where uncertainty, materiality and legitimate authority make it necessary.

3. Turning Progressive Friction Into an Operating Mechanism

A consequential workflow can maintain a compact governance state containing, as applicable, governing

obligations, identity, current authority, material evidence, uncertainty, model or evaluator findings, the current decision state, reversibility, required human involvement and the action currently being considered.

As that state changes, the level of friction can change with it.

Obligation -> Knowledge -> Authority -> Evidence -> Decision Fork -> Human Judgment -> Authorization ->

Execution -> Consequence -> Reconstruction

Several events often compressed into a single word such as "decision" are mechanically different:

AI recommendation != human approval != institutional authorization != attempted execution != actual execution != observed outcome.

Those distinctions become increasingly important as AI systems move from providing information toward initiating or materially advancing actions.

4. Model Risks and Limitations Should Connect to Operational Controls

Version 5.0 usefully distinguishes model risks from model limitations and separately addresses testing,

validation and monitoring. A practical follow-on question is:

What happened because the limitation or finding existed?

For a material limitation, validation problem, monitoring alert or other significant finding, regulators may

reasonably ask whether it:

- required additional review;
- restricted the model's permitted use;
- narrowed the action the workflow could take;
- required additional evidence or re-validation;
- escalated the matter to a qualified person;
- held a consequential transition;
- allowed an exception and, if so, under whose authority; and
- left evidence that the control actually operated.

There is an important difference between documenting a limitation and binding a limitation to the decision

process it is intended to constrain. Project OV testing has encountered that distinction directly.

5. Meaningful Human Approval Should Be Bound to the State Actually Reviewed

Meaningful human review requires more than the existence of a person somewhere in the workflow. For

consequential decisions, a reconstructable record should be capable of showing who reviewed the matter, what

authority existed at that time, what material information was presented, what uncertainty or limitation was

visible, what alternatives were available, what the reviewer decided, and what material state the approval covered.

Approval should not automatically become indefinite authority. If material conditions later change - for example,

amount, payment recipient, material evidence, governing policy, model result, model version, material limitation

or reviewer authority - the workflow should be capable of determining whether earlier approval remains valid or

whether review must reopen.

This is another practical expression of progressive friction:

Material state change -> renewed governance burden

6. Third-Party AI Should Remain Within the Insurer's Governed Decision Record

Insurers should not be expected to recreate a vendor's development environment or duplicate the work of

specialist model evaluators. But outsourcing a model should not require outsourcing the insurer's ability to

reconstruct its own consequential decision.

For material uses, an insurer should be able to preserve enough information to establish the third-party system

or model used, the relevant version or configuration, its intended role, important known limitations, material

testing or evaluation information, relevant changes, how its output affected the insurer's workflow, and who

retained institutional decision authority.

This allows specialized systems to remain specialized. Model developers, evaluators, identity providers,

cybersecurity systems, observability platforms, claims systems and data platforms may each provide relevant

evidence. The insurer's governed decision record can preserve the role and provenance of those contributions

without requiring one system to perform every specialty.

7. Representative Decision Reconstruction Could Complement Existing Examination

Project OV is not suggesting another mandatory exhibit or a universal reporting requirement.

For higher-risk

situations, however, regulators could consider requesting one or more representative decision-level

reconstruction records.

Possible inquiry: For AI Systems or Models Involving Direct Consumer Impact, Material Financial

Impact, elevated autonomy, or other higher-risk uses, regulators may request representative records

demonstrating how the company's AIS Program operated in practice. Such records may include, as

applicable, material evidence relied upon; relevant AI model or system identity; material risks, limitations or monitoring findings; applicable human review and authority; material exceptions or state changes; resulting authorization or disposition; downstream action; and subsequent challenge,

remediation or re-review.

Related inquiry: Where a material model limitation, testing result or monitoring finding is identified,

describe how that finding affects operational controls, escalation, human review, re-validation or restrictions on use, and provide an example demonstrating how the control operated in practice.

These inquiries would build on the Supplement's existing structure rather than create a competing governance framework.

8. Maria Chen - A Bounded Claims Test of Progressive Friction

Project OV has used a fictional commercial-property claims scenario involving Maria Chen as a bounded test

environment. The purpose is not to determine whether the fictional claimant is legally entitled to coverage or

payment. The test asks whether the evidence, governing obligations, uncertainty, authority, AI assistance,

human judgment, execution and later challenge surrounding a consequential claim decision can be faithfully

reconstructed.

The synthetic scenario involves a \$340,000 commercial-property claim under a \$500,000 policy. Approximately

\$185,000 is comparatively uncontested, while approximately \$155,000 depends materially on disputed

causation and coverage treatment. Relevant materials include policy terms, claims procedures, authority rules,

inspections, weather information, repair estimates, competing expert opinions, AI-assisted analysis, claimant

challenge and information arriving after an initial decision.

Routine assistance

The AI-assisted process may organize information, retrieve evidence, identify potentially relevant provisions,

compare records, surface discrepancies and prepare an initial recommendation. It does not receive institutional

authority to settle or deny the claim.

Escalation

As the workflow approaches the materially disputed portion of the claim, consequence and uncertainty increase.

The architecture therefore reaches a Decision Fork:

Proceed with a consequential disposition now, or pause/narrow the pathway pending additional evidence and/or qualified human review.

The governing question is not "Is the AI right?" It is: Is further autonomous progression permissible under

the exact evidence, uncertainty, authority and governing state that exists now?

The human handoff and exact-state approval

The intended human-review state shows applicable policy provisions, relevant procedures, reviewer authority,

evidence supporting competing interpretations, evidence weighing against them, unresolved uncertainty,

missing evidence, the AI recommendation and its limitations, open challenges, available

alternatives and material consequences.

The human disposition is separately recorded from the AI recommendation and from later institutional authorization and execution. The test then deliberately changes material conditions after approval - such as amount, payment destination, policy interpretation, causation evidence or authority - to determine whether the earlier approval still authorizes the next consequential action.

Later evidence and reconstruction

Later evidence is not inserted backward into the original rationale. The record is intended to preserve what was known at the time, what was missing, what was actually considered, what arrived later, what earlier conclusions were affected, what required reconsideration and what later decision superseded the earlier state.

The purpose is not to assume institutions always make the correct decision initially. It is to make the decision - and subsequent correction - reconstructable.

9. How Far the Maria Chen Testing Has Reached

Portions of these mechanics have been executed and tested. In a September 2026 bounded Chen adaptation,

the test was executed twice and completed 47 of 47 specified mechanism checks.

Those runs demonstrated, among other things, preservation of a disputed factual premise; retention of a

rejected correction rather than silent replacement; reviewed correction while retaining the predecessor;

identification of dependent interpretations requiring renewed review; refusal to allow changed scope to reuse an

earlier approval; refusal of selected unauthorized actions; durable recovery following a forced process stop; and

preservation of a completed receipt without duplicate delivery.

Those mechanism checks did not establish that the full institutional pilot was complete. The same work

exposed material bottlenecks:

- Affected-party access: a simulated claimant-side participant could access part of the record but could not

use the intended formal challenge pathway.

- Semantic separation: a causal statement could enter a lower evidentiary tier on one tested path even

though the architecture intended to distinguish observation, derivation and interpretation.

- Review independence: competing higher-order interpretations could be promoted by one

reviewer/

steward account without the intended review structure.

- Interpretive replacement: a tested higher-order replacement pathway refused because the fuller capability

remained unfinished.

- Complete consequential reconstruction: the bounded run did not yet produce the complete institutional

outcome and independently reconstructed decision packet required for full acceptance.

Subsequent engineering strengthened additional decision-state and scope-binding controls. It also exposed an

important authority boundary: in one preserved test, authority could be deactivated after final validation while a

decision still committed. That counterexample remains an open authority-binding problem and is preserved as a

failure rather than converted into a passing result.

10. A Separate Bounded Underwriting Challenge

A separate synthetic contractor-pollution-liability underwriting workflow tested ten bounded capabilities involving

source-linked application fields, completeness checks, source-referenced changes, consistency evaluation,

evaluator receipts, reopening when evidence changed, review routing, human approval, authorization, simulated

execution and restart/reconstruction.

The controlled run completed 176 of 176 specified checks. The same frozen capability set was then subjected

to a broader adversarial challenge that completed 28 of 36 expectations, leaving eight unmet expectations

across six finding areas. The challenge remains classified as findings open. It is not represented as production

underwriting, carrier validation, ACORD conformance, an actual quotation or binding transaction, or independent

validation.

The most consequential finding

One evaluator returned an unassessed result requiring review. The evaluator result was retained as evidence.

However, when the evaluator limitation had not been translated into a structured downstream obligation,

scripted human approval and authorization could still progress to one local simulated release.

When the

limitation was explicitly bound to the workflow, review and authorization correctly refused progression.

The system possessed the warning, but possession of the warning did not itself prove that the

warning controlled the consequential workflow.

This exposed a practical distinction between recorded governance and operational governance. The same challenge also exercised controls that rejected or constrained agent self-approval, authorization

before required review, stale revisions, changed-price execution, changed-recipient execution, modified

evaluator receipts, evidence tampering, truncated history, forged decision worksheets and duplicate simulated execution.

Other findings involved structurally present but substantively inadequate information, including visually blank

fields, ambiguous answers and material disclosures outside the limited mapped-field set. These findings

reinforce another distinction: structural completeness is not substantive adequacy.

11. Observations From the Tests

The claims test and the separate underwriting challenge address different parts of the same governance

problem. The claims test provides the broader consequential-decision path; the underwriting challenge shows

how a relatively small integration gap can break that path.

Together, the tests have exposed several distinctions that may be relevant to regulatory evaluation:

- recording authority is not the same as proving authority at decision time;
- recording a model limitation is not the same as binding it to the applicable workflow;
- identifying uncertainty is not the same as making uncertainty affect progression;
- having a human somewhere in a workflow is not the same as meaningful qualified human review;
- recording approval is not the same as binding approval to the exact material state;
- recording that an action was blocked is not necessarily proof that downstream execution was prevented;
- preserving a model output is not the same as making the model the decision authority; and
- preserving a final outcome is not the same as reconstructing how that outcome occurred.

These distinctions become visible when governance is tested mechanically rather than evaluated only from policy documentation.

12. A Decision-Level Lens for the Supplement

NAIC Area Complementary Decision-Level Question

Materiality What causes the workflow to move from routine controls to heightened governance?

Direct Consumer Impact Can the path to the consequential consumer outcome be reconstructed?

Human Oversight Was the appropriate human engaged at the consequential fork with sufficient information and actual authority?

NAIC Area Complementary Decision-Level Question

Autonomy What could the system do without additional authorization?

Reversibility Could the contemplated action be stopped, corrected or reversed?

Explainability Can a later reviewer understand the actual decision basis, not merely the model generally?

Model Risks What risk was identified, and how did governance respond?

Model Limitations Did the limitation change what the system or people were permitted to do?

Testing / Monitoring Did a material finding produce an operational consequence or merely a report?

Third-Party Models How did external model evidence enter the insurer's own governed decision record?

Human Approval What exact material state did the approval cover?

Change Over Time What material change reopens review or invalidates stale authority?

Execution What was recommended, approved, authorized, attempted and actually executed?

Challenge / Remediation Can later evidence or objections change the operative state without erasing history?

13. Boundaries of the Project OV Contribution

Project OV is not offered as a replacement for the Supplement or for existing insurance governance. It does not

perform every function the Supplement asks insurers to describe.

Project OV does not claim to substitute for model inventories, actuarial validation, model-performance testing,

unfair-discrimination testing, privacy programs, software-development controls, enterprise risk management,

legal interpretation, claims expertise, carrier underwriting standards or regulatory judgment.

Specialists and existing institutional systems should continue to provide those capabilities.

Project OV focuses

on the relationships among them when they become material to a consequential decision.

The architecture is intended to preserve distinctions among evidence and claims; claims and interpretation; AI

output and authority; recommendation and approval; approval and institutional authorization; authorization and

execution; attempted execution and actual effect; historical state and later knowledge; and current and

superseded decisions.

The objective is not to eliminate uncertainty. It is to make uncertainty, authority and consequential decision

history sufficiently visible that they can be reviewed, challenged and reconstructed.

Conclusion

Version 5.0 asks increasingly mature questions about how insurers identify, govern, test and oversee AI

systems and models. A complementary perspective is:

For selected consequential decisions, can the institution reconstruct how its governance actually

operated?

A useful decision record should make it possible to determine what was known, what was uncertain, what obligations and rules applied, what AI contributed, what risks or limitations were known, who had authority, where progressive friction increased, what the human decided, what was authorized, what actually occurred, what outcome followed and what changed afterward.

Project OV testing to date indicates that many of these relationships can be made mechanically visible. It has also shown that relatively small gaps between evidence, limitations, authority and execution can materially weaken governance if they are not tested.

Decision-level reconstruction and progressive, risk-proportionate friction may therefore complement the AIS

Program, model, risk, limitation and data inquiries already contained in the Supplement. Supporting no-code architecture descriptions, synthetic test methodology, preserved findings and limitations, and technical evidence can be made available if useful to the Working Group's evaluation.

Respectfully submitted,

Brad Rothrock

Founder, Project OV

Reference materials considered:

NAIC AI Risk Evaluation Supplement Version 5.0

NAIC AI Risk Evaluation Supplement - Summary of Changes from Version 4.0

Project OV synthetic test references are described only at the level necessary to explain the regulatory observations. The tests do not establish production insurance readiness or independent validation.

September 29, 2026

Via Electronic Submission

To: Scott Sobel, AI Policy Advisor; Miguel Romero, Director of Innovation, Cybersecurity and Technology

Re: AI Risk Evaluation Supplement Version 5.0 - Inspectable Evidence of AI Governance Controls

Dear Mr. Sobel and Mr. Romero:

Thank you for the opportunity to comment on Version 5.0 of the AI Risk Evaluation Supplement. I am Dwayne Clark, founder of Recon.AI. Our work focuses on evidence-based governance for AI and agentic systems, including the ability to establish what a system was permitted to do, what evidence supported an action or decision, what human oversight occurred, and what governance controls actually operated at the relevant time.

Version 5.0's increased emphasis on model inventories, identifiable supporting documentation, third-party model oversight, explainability, and agentic AI is a constructive development. I respectfully suggest one additional principle for consideration as the Supplement evolves:

Regulatory evaluation should distinguish evidence that a governance control is documented from evidence that the control actually operated as intended for a particular AI system, model, decision, or action.

From control design to evidence of operation

Policies, governance frameworks, model inventories, and control documentation provide important evidence of control design. For higher-risk AI uses, regulators may also benefit from contemporaneous, inspectable evidence demonstrating control operation. Such evidence could establish:

- which AI system and model or version was involved;
- the purpose and scope for which that system was authorized;
- the data, evidence, or other inputs materially relied upon;
- whether a human review or approval requirement applied and, if so, whether it occurred;
- whether material model, data, third-party, or operating-condition changes occurred after initial approval;
- whether monitoring detected drift, anomalies, control failures, or conditions outside approved boundaries;
- what intervention, escalation, remediation, or recovery actions occurred; and
- whether the resulting evidence can be independently reconstructed or verified during supervisory review.

Agentic AI and third-party models

This distinction may become particularly important for agentic AI. A governance framework may establish that an AI system is generally approved for a use case, while a separate evidentiary question remains: **Was this particular action within the system's authorized scope when it occurred, and what evidence demonstrates that?**

A similar issue arises with third-party models. An insurer may maintain appropriate vendor-governance documentation while still needing to demonstrate how changes in a third-party model, service, data source, or capability are detected and evaluated throughout the lifecycle.

A three-part evidence distinction

Evidence category	Supervisory purpose
1. Control design evidence	Policies, procedures, governance frameworks, assigned responsibilities, and documented control requirements.
2. Control operation evidence	Records demonstrating that required controls actually operated for relevant AI systems and material events.
3. Lifecycle evidence	Records demonstrating ongoing monitoring, material-change detection, human intervention, escalation, remediation, and recovery over time.

This distinction need not require a particular technology or implementation method. The objective would be to help regulators determine not only whether appropriate governance has been described, but whether sufficient evidence exists to reconstruct and evaluate how that governance functioned in practice. This could strengthen decision defensibility while reducing reliance on retrospective narrative reconstruction during examinations.

Potential application within the Supplement

NAIC may wish to consider whether Exhibit B, related regulator instructions, or future supervisory guidance could expressly distinguish these evidence categories. The Version 5.0 request for document names and page numbers is a useful mechanism for connecting responses to documentation. For material or higher-risk uses, an additional evidence-of-operation concept could help examiners determine whether documented controls were effective in the relevant operational context.

This suggestion is intended as an evidence architecture principle rather than a recommendation for any particular vendor, platform, or technical implementation. Nor would such evidence, standing alone, establish that an AI system is fair, actuarially valid, compliant, or appropriate. Rather, it can make the operation of governance controls more inspectable and independently reviewable.

Thank you for considering this comment and for the Working Group's continued development of practical supervisory tools for AI systems. I would welcome the opportunity to participate in further discussion regarding evidence architecture, agentic AI governance, third-party model oversight, or methods for making AI governance controls more inspectable during regulatory evaluation.

Respectfully,

Dwayne Clark
Founder, Recon.AI

September 14, 2026

Scott Sobel, AI Policy Advisor

Miguel Romero, Director

National Association of Insurance Commissioners

Submitted by email to ssobel@naic.org and maromero@naic.org

Re: Comment on the AI Risk Evaluation Supplement, version 5.0, exposed August 31, 2026

Dear Mr. Sobel and Mr. Romero:

I write in response to the Big Data and Artificial Intelligence (H) Working Group's exposure of the AI Risk Evaluation Supplement, version 5.0. I am an insurance and risk adviser in New Jersey. My practice serves private equity sponsors, their portfolio companies, and carriers, and my background is carrier underwriting at AIG, risk management at Goldman Sachs, and consulting at EY.

A note based on these comments. I worked through version 4.0 request by request earlier this year. For version 5.0, I relied on the Working Group's published summary of changes rather than the clean text, and where a point below rests on language I have not read in its current form, I have said so. I would withdraw any comment the current text has already answered.

In the interest of transparency, I also advise IndyKite, a technology company working on identity and authorization for Artificial Intelligence (AI) systems. This letter is submitted on my own behalf and on behalf of Sharer Consulting LLC. It recommends no product, proposes no technology, and has not been reviewed by any of my clients.

The changes described in the summary, including the renaming, the repositioning of the document as a supplement to existing handbooks, the explicit model inventory, and the separation of models with Direct Consumer Impact from those with Material Financial Impact, make the instrument better bounded than its predecessor. They reflect a 12-state pilot that has been running since March. My comment concerns one thing the revision appears not to have done.

Summary of the comment

Version 5.0 builds an explicit financial-impact track. It defines Material Financial Impact, it aligns Materiality with the Financial Condition Examiners Handbook, and it asks insurers to identify separately those models whose effect is financial rather than directly on consumers. Based on version 4.0 and the summary of changes, I see nothing in that track connecting the inquiry to the actuarial function that performs the underlying work, or to the accountability structure that already governs it.

I recommend four changes. None asks the Working Group to require a control that does not exist today, and one of them should return examination effort rather than consume it.

1. Record, by function, who prepared and who reviewed the response.

RE: : Comment on the AI Risk Evaluation Supplement, version 5.0
September 14, 2026

2. For models flagged Material Financial Impact, ask whether the model's output is directly relied upon in work supporting an actuarial opinion, memorandum, or report, and what the actuary responsible for that work did with the model.
3. Cross-reference the Handbook's existing guidance on actuarial work, given that the Supplement now takes its materiality definition from that Handbook.
4. In the new third-party model oversight item, distinguish vendor governance from professional review.

The substance is actuarial, and the accountability is not.

In version 4.0, the operational areas listed in Exhibit A included ratemaking, rate classification, schedule rating, and premium audits, and they included reserves and valuations. The high-risk model questions asked how a model affects the financial statements, the risk assessment, or the control environment. The governance questions reached the Own Risk and Solvency Assessment. Version 5.0 also adds a materiality standard drawn from financial examination practice. A substantial part of the instrument therefore addresses work that actuaries perform and that an existing body of professional standards already governs.

In version 4.0, the word "actuarial" appeared exactly once, in narrative question 4, as one item in a list of third-party professional services whose AI the company is asked to oversee. The summary of changes records no revision on this point. If that language has carried forward unchanged, the instrument's substance is actuarial throughout, while its only mention of the actuary is as a category of outside vendor to be supervised.

I do not read this as an oversight. The Supplement is company-directed by design; it addresses the insurer rather than a credentialed individual, and that is a defensible drafting choice. The choice has three consequences worth naming.

Attribution. A carrier can be fully responsive across every exhibit with no actuarial involvement anywhere in the record. Whoever owns the submission owns the answer, and in many carriers that owner will be compliance, legal, or technology.

Reliability of the hardest answers. Questions on validation, testing, and monitoring are most likely to be answered by a function that does not perform the work. Such a response describes intent rather than practice, and that difference is what a follow-up inquiry is meant to surface.

RE: : Comment on the AI Risk Evaluation Supplement, version 5.0
September 14, 2026

A missed opportunity to scope. The summary of changes indicates that regulators may limit further inquiry based on the responses received. Where a financially material model already sits inside work an actuary performed and is answerable for, that is a reasonable place to begin scoping. Nothing I can see in the instrument surfaces it.

Recommendations

One. Record who prepared and who reviewed the response, by function. One field, naming the function rather than the individual. The purpose is to show the regulator where the answer originated, not to place professional exposure on a named person who may not have drafted the document. I would ask that the field call for a single accountable function rather than a list, since an answer naming four contributing functions conveys nothing. The premise is modest. For a model that drives reserves or rates, whether the actuarial function saw the response before it was filed is a fact an examiner would want and cannot presently obtain.

Two. For models flagged as having Material Financial Impact, ask about reliance rather than sign-off. Suggested language: "For each AI model identified as having Material Financial Impact, indicate whether the model's output is directly relied upon in work supporting an actuarial opinion, memorandum, or report. If so, describe the scope of the actuary's involvement with the model, including whether the actuary evaluated the model or relied on the work of others, and identify the standards the actuary applied."

I would ask the Working Group to resist a simpler version of this question. Actuaries opine on reserves and on rates, not on models. A pricing model feeds an indication that feeds a filed rate, and an actuary rarely signs an opinion on that model as a model. A question asking whether an actuarial opinion covers a model therefore invites a false binary, and it will return inconsistent answers: negative for well-governed models and affirmative for lightly touched ones. Reliance is the concept actuaries work in and the one they can answer accurately. The word "directly" is doing work in the suggested language. Without it, a model several steps upstream of an indication is relied upon in one sense and not in another, and the question inherits the ambiguity it was meant to remove.

For the same reason, the question should ask for scope and basis rather than a yes or a no. A question answerable with a yes can, in time, be satisfied by a thin sign-off obtained for the purpose. A question asking what the actuary did makes a thin review visible.

Two notes on drafting. Opinion, memorandum, and report do not carry identical meaning across property and casualty, life, and health, so the instruction should let a company name the item applicable to its line rather than force a single label. And on the final clause, asking a company to identify the standards its actuary applied is a disclosure about that company's own practice. It is not an adoption by the NAIC of standards set by an independent board, and the instruction should say so if there is any risk of it being read the other way.

RE: : Comment on the AI Risk Evaluation Supplement, version 5.0
September 14, 2026

I believe this recommendation returns examination effort. A substantive affirmative answer gives a regulator a documented place to begin scoping, which is what the summary of changes already contemplates. A negative answer, for a model the company itself designated as financially material, is the finding.

Three. Cross-reference the Handbook's existing guidance on actuarial work. Version 5.0 aligns its materiality definition with the Financial Condition Examiners Handbook. Where that Handbook already addresses an examiner's use of actuarial work, a cross-reference in the definitions or the instructions would keep the Supplement consistent with the authority it now draws on, and it introduces no new expectation of any carrier.

Four. In the third-party model oversight item, distinguish vendor governance from professional review. Version 5.0 adds a checklist item on third-party model oversight and, per the summary of changes, adopts a definition of Third Parties drawn from the Third Party Registration Framework. For a third-party model used in ratemaking, rate classification, reserving, or valuation, contractual vendor governance and review by the company's own actuarial function are different controls with different failure modes. I recommend asking for both, or asking the company to state which of the two it relies upon.

I expect the answer will often be that no actuary at the company has examined the third-party model, and that contractual restrictions on access are the reason rather than any failure of diligence. That is the argument for asking. The question separates a vendor relationship the company can see into from one it cannot, and it identifies where access terms rather than governance policies are the binding constraint.

Objections I anticipate

That ASOP 56 already covers this. Actuarial Standard of Practice No. 56 applies to models an actuary designs, develops, selects, or uses, whether or not the NAIC says so, and it applies no less because a model uses AI. I agree, and that is why recommendations two and four are framed as questions about what already happened rather than as new requirements. The standard binds the actuary who touches the model. It says nothing about whether an actuary touched the model at all, and that is the fact an examiner cannot presently see.

Although the American Academy of Actuaries did not ask for this, the Academy commented on the predecessor exposure on September 8, 2025. In that letter, it addressed posture and principle: regulatory consistency alongside state flexibility, principle-based flexibility, consumer protection and fairness, and support for the pilot. It did not raise this point. That no one asked for something does not establish that it is unnecessary, and the most direct test available is to put the question to the Academy during this cycle, which would settle it better than my letter can.

This reaches beyond the Working Group's charge: the actuarial opinion, the appointed actuary, and the standards governing them sit with other NAIC committees and task forces, and recommendations two and three touch their territory. If the Working Group concludes the question is not its to answer, I would

SHARER CONSULTING LLC

Advisory · Innovation · Transformation

RE: : Comment on the AI Risk Evaluation Supplement, version 5.0
September 14, 2026

ask that it be referred rather than closed, so that it stays live for the version going to the Fall National Meeting.

That this adds burden to a revision intended to reduce it. Recommendation one is a single field. Recommendation three is a cross-reference. Recommendation four replaces one question with a better-specified version of the same question. Recommendation two adds a request, and it should return effort by giving a regulator a documented place to begin scoping. I should be candid about the limit of that claim. None of the four requires a carrier to create a control it does not have, but a question about actuarial involvement will, over time, encourage some carriers to obtain it. I regard that as a benefit rather than a hidden cost, and the Working Group may weigh it differently.

I appreciate the Working Group exposing the revision for comment, and I am glad to answer questions on any of the above.

Respectfully submitted,



Jeffrey Sharer
Founder, Sharer Consulting LLC
Jeffrey.sharer@sharerconsulting.com
914-649-9919



September 29, 2026

Commissioner Nathan Houdek, Chair
Doug Ommen and Mary Block, Co-Vice Chairs
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
By email to Scott Sobel (ssobel@naic.org) and Miguel Romero (maromero@naic.org)

Re: Comments on the AI Risk Evaluation Supplement, Version 5.0

Dear Chair Houdek and Members of the Working Group:

Verdict Systems Inc. appreciates the opportunity to comment on Version 5.0 of the AI Risk Evaluation Supplement, exposed on August 31, 2026.

Verdict builds technology that records automated and AI-influenced actions at the moment they occur, in a form a third party can check. We have a commercial interest in this subject and say so at the outset. The recommendations below are technology-neutral. None requires a particular vendor, product or standard, and each can be met with openly published methods.

1 Summary

Version 5.0 asks insurers to describe their AI controls. It does not yet ask whether an insurer can show that a control operated on a particular action. We propose one added clause in Exhibit B, one added field in Exhibit C and one definition. Together they would let a regulator ask that question in a form a company can answer and an examiner can check.

2 The Working Group has already heard the distinction

The Working Group's August 13 Summer National Meeting materials include a presentation from AM Best on AI governance across predictive, generative and agentic systems. It put the difference between two kinds of governance in one line:

“Documentary governance answers with policies. Operational governance answers with evidence.”

For agentic systems, the same presentation listed “full action logging” and “clear evidence of human review before consequential actions” among the signs that controls work in practice.¹ AM Best labeled the material as discussion only, with no change to its criteria. We cite it for the distinction, not as a standard.

Version 5.0 currently asks for the first kind of answer. Exhibit B asks for the written AIS Program and, in the checklist version, the document name and page number behind each response. Exhibit C asks, for each high-risk model, how it is tested and monitored and when it was last tested. Exhibit B also asks, “How does the insurance company assess autonomy, reversibility, and reporting impact risk of AI Systems?” These are the right questions. Each is answered with a document the company writes about its own controls. The words “record,” “log” and “evidence” do not appear anywhere in the draft.

3 Why Version 5.0 is the right place

Version 5.0 defines agentic AI for the first time: systems that “pursue objectives across multiple steps without requiring human input at each stage,” including by “taking actions (often through tools or APIs).” For a scoring model, a description of human review may be an adequate answer to Exhibit B question 2b. For a system that acts, an examiner needs to know what the system did, who approved it,

and whether the company can show both afterward in a form someone outside the company can check.

The Model Bulletin already contemplates this kind of request. Among the information a department may request about a specific AI System, Section 4 lists “Documentation of compliance with all applicable AI Program policies, protocols, and procedures.”² The Supplement is where that request takes a concrete shape.

4 What makes a record checkable

Three properties separate a record an examiner can rely on from a log the insurer controls:

- (a) It is created when the action occurs, not reconstructed later.
- (b) Any change made to it after creation can be detected.
- (c) A party other than the insurer can confirm that it existed at a stated time, without access to the insurer’s systems.

Open mechanisms already provide each property: cryptographic hashing, hash chaining, trusted timestamps under RFC 3161, and public append-only transparency logs of the kind web browsers rely on through Certificate Transparency (RFC 6962). The IETF’s SCITT working group is developing transparency services for signed statements. A regulator does not need to choose among these. The Supplement only needs to ask whether a company’s records have the three properties.

These records have limits, and they should be stated plainly. A record of this kind shows that it existed in a given form at a given time. It does not show that the recorded event happened as described, that every action was recorded, or that the underlying decision was correct. Examiners would still test completeness and accuracy by sampling. What changes is the starting point: the examiner samples records whose integrity and timing can be checked, instead of starting from a narrative.

5 Recommended additions

5.1 Exhibit B (Narrative Version), question 2b. After “whether there is a human in the loop,” insert the text below, and make the corresponding addition to the Checklist Version:

PROPOSED INSERTION

“and whether, for a sampled AI-influenced action, the company can produce a record of the action and of any human approval, override or escalation that was created when the action occurred, and whose integrity and time of creation a third party can verify without access to the company’s systems,”

5.2 Exhibit C, new field following Field 11.

PROPOSED FIELD

“Action records: whether the model’s actions, and any human approvals or overrides of those actions, are recorded when they occur in a form that shows any later change; how long those records are retained; and how a regulator could verify a sample of them.”

5.3 Definitions.

PROPOSED DEFINITION

“Contemporaneous Action Record” refers to a record of an action taken or recommended by an AI System, and of any human approval, override or escalation of that action, that is created at or near the time of the action, whose integrity can be verified, and whose time of creation can be confirmed by a party other than the company.

5.4 Scope. Consistent with the Supplement’s risk-based design, regulators could direct these questions first to agentic AI and to AI Models with Direct Consumer Impact or Material Financial Impact. They are questions, not mandates. A company that answers that it cannot produce such records has told the regulator something useful about inherent risk.

6 Related submissions

Gerald (Gerry) Kennedy of Observatory Strategic Management has filed two comment papers in this period: “From Model Review to Enterprise Solvency” (September 2, 2026) and “From Recognition to Operative Supervision” (September 23, 2026). The second recommends that mitigation credit reflect evidence grade, with independent validation receiving more credit than assertion or documentation (Recommendation 8); action logging and evidence preservation sufficient for coverage, causation, claims defense and regulatory review (Recommendation 13); evaluation of human oversight for demonstrated intervention (Recommendation 16); and a business provenance record that denies verified-evidence status to records whose origin or transformation cannot be reconstructed (Recommendation 27). We support those recommendations. This comment addresses a narrower question they depend on: what makes a record verifiable by someone other than the company that produced it.

Mr. Kennedy and Verdict have discussed an advisory role for him. No advisory agreement has been made, and Verdict did not draft or fund either of his papers. He has reviewed this reference.

7 Demonstration

Appendix A reproduces one record made this way, for a fictional claims-payment approval, so that staff can see what an examiner would receive. Two of the three checks it describes use only a standard SHA-256 tool and a public log that Verdict does not operate. The record was produced on Verdict’s public sandbox, which signs with a short-lived key generated for each request. It demonstrates the integrity and timing checks. It does not demonstrate the identity or key custody of a production signer.

We would welcome questions from members or staff, and would be glad to support any field validation the Working Group undertakes.

Respectfully submitted,

Shayne Beavan

Founder, Verdict Systems Inc.

founder@verdict.systems • +1 832 641 1892

Disclosure. Verdict Systems Inc. is a Delaware corporation based in Houston, Texas. It has filed U.S. patent applications relating to its products; none has been examined. It is not an insurer or producer. The recommendations above do not depend on any Verdict technology.

¹ Big Data and Artificial Intelligence (H) Working Group, Summer National Meeting materials, August 13, 2026, Attachment B: AM Best, “AI Governance in Insurance: Distinguishing Risk Across Predictive, Generative and Agentic Systems,” slides 14, 17 and 21.

² NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers (adopted December 4, 2023), Section 4.

Appendix A. A contemporaneous action record

DEMONSTRATION • FICTIONAL SCENARIO • PUBLIC SANDBOX TIER

Scenario (fictional). An AI claims agent recommended payment of \$18,400 on homeowners water-damage claim DEMO-CLM-0925. The company’s claims policy requires adjuster approval for any AI-recommended payment above \$10,000. Adjuster ADJ-114 approved, and the payment was released. The recommendation, the rule that required approval, the approval and the release were captured when they occurred and sealed as one record.

RECORD ID	ser_60599b88be0b957451470aed88ce1e4b
COMMITMENT	27049c8b8a6d6d57326e577ad2cd4b476d98351c6962d7e40570cc5290ec46e5
PUBLIC LOG	Sigstore Rekor, log index 2951501338
LOG TIME	2026-09-25 06:42:15 UTC
LOG VALUE	e8ab2c90962568e0e5d672c8bb9d128078ed9acc486ee7269bfc409da89496

How to check it

1. Look up log index 2951501338 in the public Rekor log, for example at search.sigstore.dev. The entry shows the log value above and the time the log recorded it. Verdict does not operate this log.
2. On any computer, compute the SHA-256 of the commitment string. The output equals the log value, which shows the commitment existed no later than the log time.

```
printf '%s' 27049c8b8a6d6d57326e577ad2cd4b476d98351c6962d7e40570cc5290ec46e5 | shasum -a 256
```

3. Re-deriving the commitment from the full record ties its content to the logged value; a change to any byte produces a different commitment that no longer matches the log. Verdict will provide the full record to staff on request. Its verifier at verdict.systems/verify performs this step in the browser without sending the record anywhere.

What this record shows, and what it does not

STATEMENT	STATUS
The content is the same content that was logged	Shown, steps 1–3
The record existed no later than 06:42:15 UTC on September 25, 2026	Shown, steps 1–2
Steps 1 and 2 need no access to the insurer’s or Verdict’s systems	Shown
The claim, the approval and the payment occurred	Not shown (fictional)
Every action in the workflow was recorded	Not shown
Who signed it	Not shown (sandbox key)

The last three rows are why examination would still sample and test. A record that states its own limits is easier for an examiner to rely on than one that does not.

From Brian Wroblewski (Individual Commenter)

Thank you for the opportunity to comment on Version 5.0 of the AI Risk Evaluation Supplement.

I support the Supplement's continued movement toward evidence that is specific, traceable, and capable of regulatory review. As insurers deploy increasingly capable AI systems, including agentic systems, I encourage the Working Group to consider whether the Supplement should more explicitly distinguish general staff training from evidence that the individuals responsible for operating, reviewing, overriding, or supervising a material AI use are actually prepared for that role.

Without prescribing a particular credential, course, testing method, or technology, the Supplement could clarify that examiners may consider evidence addressing:

1. the roles or functions responsible for human review, escalation, override, or supervision of material AI-supported activity; 2. the competencies the insurer has determined are appropriate for those roles in light of the AI use case, its risks and limitations, and the authority given to the individual; 3. how the insurer determines that relevant personnel have demonstrated those competencies, rather than relying solely on attendance or course completion; 4. how recently the evidence was established and whether material changes to the AI system, use case, policies, or operating procedures trigger reassessment; 5. how gaps identified through testing, incidents, complaints, monitoring, or model changes are translated into targeted remediation for the relevant personnel; and 6. how the insurer documents these determinations in a manner that can be reviewed without requiring a regulator to infer competence from a general training log.

A principles-based formulation would preserve flexibility across insurers, business lines, technologies, and risk levels. It would also avoid creating a new credentialing regime or implying that a single course, certification, or standardized score establishes readiness for every use of AI.

The broader purpose would be to connect the Supplement's strong focus on AI-system governance with the human oversight on which many controls ultimately depend. A well-designed control should be able to show not only that human review is required, but also why the designated reviewer is appropriately prepared for the particular function being performed.

Thank you for considering these comments.

Brian Wroblewski
Individual Commenter



NATIONAL ALLIANCE OF LIFE COMPANIES

An Association of Life and Health Insurance Companies

September 29, 2026

Mr. Scott Sobel
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106-2197
Email: maromero@naic.org

Subject: AI Risk Evaluation Supplement version 5.0

This letter is submitted on behalf of the National Alliance of Life Companies (NALC). NALC is a trade association of more than fifty life insurance companies and associates that represents the interests of smaller and mid-sized life insurers, specialty insurers, and their policyholders. Thank you for the opportunity to comment on Version 5.0 of the AI Risk Evaluation Supplement, and specifically on its treatment of generalized linear models (GLMs).

NALC supports the Working Group's work on the AI Risk Evaluation Supplement. We believe that it provides important guidance to regulators and carriers alike. We agree that GLMs, like any model used in insurance decisions, require sound governance. Our concern is not with that premise, but rather with Version 5.0's classification of GLMs as a machine learning technique and their inclusion within the Supplement's inventory of AI Models.

GLMs are the standard statistical method behind experience studies, pricing and valuation assumptions, and simplified issue underwriting. They are not a new or emerging technology. They have been in use of more than 50 years, and are already subject to a mature governance and supervisory framework including Actuarial Standards of Practice. Their outputs are reviewed by regulators through actuarial memoranda and, for principle-based reserving, VM-20 assumption-setting requirements and VM-31 reporting.

The NAIC's approach to insurers' use of AI, as reflected in the Model Bulletin and in the Supplement itself, is risk-based: governance and regulatory scrutiny are meant to be proportionate to the risk of adverse consumer outcomes that a given system presents. Treating long-established, fully transparent GLMs as AI Models inverts that approach, directing the Supplement's most intensive inventory and documentation requirements toward the models that present the least AI-specific risk and diluting regulators' ability to focus on the systems that warrant closer review.

A GLM's structure is specified by the actuary and its coefficients can be read and challenged, which is the opposite of what makes AI a distinct governance concern. A summary of our concerns includes:

- 1. Disproportionate burden.** A small carrier with no machine learning in production would still have to inventory, classify, and document its GLMs as AI, building governance infrastructure



NATIONAL ALLIANCE OF LIFE COMPANIES

An Association of Life and Health Insurance Companies

that larger carriers already maintain for models already documented under professional standards.

2. **Duplicative documentation.** The classification adds a second documentation requirement for the same models.
3. **Reduced transparency.** A company that uses only GLMs will report a footprint that reads as AI usage, and treating transparent and opaque models alike removes a reason to prefer interpretable methods.

For these reasons, NALC respectfully urges the Working Group to reconsider the inclusion of GLMs within the scope of the Supplement.

Thank you again for the opportunity to address our comments.

Sincerely,

Scott Harrison
Chief Executive Officer

Dear Mr. Sobel and Mr. Romero,

Please accept the written comment below on AI Risk Evaluation Supplement version 5.0. It makes four specific requests tied to this draft:

1. Keep the new "agentic AI" definition centered on irreversible action authority, not model-architecture labels.
2. Treat Exhibit B question 3p as a demand for provable oversight artifacts — not vendor self-attestation.
3. Require self-specified materiality thresholds to be enforceable and reconcilable to production records, not just disclosed.
4. Expect Exhibit A Model Inventories to be exportable and linked to real decision paths.

I would welcome the chance to discuss at the Oct 8 public Webex if useful.

Respectfully,
Demond Davis
Founder, Nisaba LLC
419-973-0274
hello@velaru.xyz

WRITTEN COMMENT OF NISABA LLC
AI Risk Evaluation Supplement, Version 5.0
National Association of Insurance Commissioners
Big Data and AI (H) Working Group
Date: September 14, 2026

I. Interest of the commenter

Nisaba LLC builds fail-closed control systems for high-stakes automated actions: clearance is granted only when a contemporaneous, attributable record can be produced for what was attempted, what policy applied, who authorized it, and whether the action was allowed or denied. We submit this comment because version 5.0 of the Supplement moves the exam framework onto that same terrain — agentic AI, third-party model oversight, disclosed materiality, and inventoryable model populations.

We are not asking NAIC to endorse a vendor. We are asking the Working Group to write the Supplement so that a company can prove the answers it will soon be required to give, rather than only assert them.

II. Support for scoping "agentic AI" — and one request for exam operability
Version 5.0's formal definition of "agentic AI" is the right move. Market-conduct work that stops at generative chatbots will miss irreversible actions already occurring in claims, underwriting referral, utilization management, and payment initiation.

For example: an agentic claims-processing tool that reviews a submitted claim and auto-approves a payout without human sign-off is exactly the kind of system this Supplement needs to reach — regardless of whether the vendor markets it as "an agent," "a workflow automation," or simply "AI-assisted processing."

Request: Keep the definition tied to action authority and irreversibility, not to model architecture labels. For exam purposes, the useful question is not "does the vendor call this an agent?" It is:

Can this system initiate or materially influence an irreversible outcome (payment, denial, binding decision, account restriction, coverage determination) without a contemporaneous, attributable control record?

If the Supplement's definition is read that way, Exhibit A inventories and Exhibit B oversight answers become testable. If it is read as marketing taxonomy, companies will inventory chat features and miss the write-paths that create consumer harm.

III. Exhibit B, question 3p — third-party AI oversight must be provable, not promissory

Exhibit B's new question 3p asks how a company oversees third-party AI models. That is the right question. The default industry answer today — "we reviewed the vendor's documentation and rely on their controls" — is not oversight. It is delegation without evidence.

For 3p to function in a market-conduct exam, a satisfactory answer should identify artifacts, not assurances. Minimally, a company should be able to produce, for each in-scope third-party model or agentic service:

1. Named accountability inside the licensee — a natural person (or named role with a sitting occupant) responsible for the third-party use, not a vendor contact.
2. A control description specific to the write-path — what irreversible actions the third party can trigger in the licensee's book of business.
3. An attestation trail a stranger to the vendor relationship can check — signed clearance records, deny/allow logs with policy versions, or independent verification channels — not only a vendor-controlled summary.
4. Fail-closed behavior when oversight inputs are missing — if inventory metadata, policy version, or attestation channel is unavailable, the high-stakes action does not proceed by default.

Concrete answer to the practical question 3p raises:

How would a company prove third-party oversight rather than assert it?

By retaining, for sampled transactions, a record that a regulator or independent examiner can verify without trusting the vendor's dashboard: what action was requested, which third-party model or agent was in path, which licensee policy version applied, who or what authorized clearance, the allow/deny outcome, and an independently checkable signature bound to that record. Oversight is the ability to produce that trail under exam. Anything less is narrative.

We encourage NAIC to state explicitly in 3p guidance or examiner notes that vendor self-attestation alone is insufficient for material third-party agentic uses, and that companies should expect sample-based testing of oversight artifacts.

IV. Self-specified materiality thresholds — disclose, then bind

Requiring companies to specify and disclose their own materiality threshold is realistic. It also creates a visible integrity gap: a firm can set a convenient threshold, disclose it, and never prove that production systems actually enforce it. Request: Pair disclosure with an exam expectation that the disclosed threshold is operational, not ornamental. A company should be able to show:

- the threshold value and effective date,
- the systems and third parties in scope above that threshold,
- at least one control that changes behavior at the threshold (review, dual control, deny-by-default, escalation), and
- a timestamped record that the threshold in force at decision time matches the disclosed threshold.

Without that binding, "self-specified materiality" becomes a paper control. With it, the Supplement creates a fair, company-specific bar that still supports examination.

V. Exhibit A Model Inventory — make it exportable and decision-linked

A Model Inventory requirement is only as strong as its freshness and its link to real decisions. We support Exhibit A and ask that examiner guidance prefer inventories that are:

- exportable in a machine-readable form,
- mapped to business write-paths (not only to model names),
- tagged for agentic / non-agentic uses using the Supplement's new definition, and
- reconciled to third-party entries that feed question 3p.

An inventory that cannot explain which irreversible consumer outcomes a listed model can influence will not help an examiner. An inventory that can will.

VI. Alignment with the Model Bulletin's accountability direction

We note separately that staff work converting the 2023 AI Model Bulletin into disclosure components — including board or senior-management attestation naming an executive responsible for AI governance — points in the same direction as Sections III–V above: named accountability, documented controls, verifiable attestation. Version 5.0 of the Supplement is the exam-facing place to make those ideas operational for agentic and third-party systems. We support that trajectory.

VII. Closing

Version 5.0 correctly expands the frame to agentic AI, third-party oversight, disclosed materiality, and model inventory. The Working Group's remaining drafting choice is whether those additions produce exam-ready proof or exam-ready prose. We urge the Working Group to:

1. keep the agentic-AI definition centered on irreversible action authority,
2. treat Exhibit B 3p as a demand for oversight artifacts, not vendor reassurance,
3. require self-specified materiality thresholds to be enforceable and reconcilable to production records, and
4. expect Exhibit A inventories to be exportable and linked to real decision paths.

Thank you for the opportunity to comment. I am available for questions before or during the Oct 8, 2026 public discussion.

EASYESSENCE

pkangari@easyessence.com | easyessence.com | 813-444-7830

September 8, 2026

VIA ELECTRONIC MAIL (ssobel@naic.org; maromero@naic.org)

Commissioner Nathan Houdek, Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106

Re: Comments on the AI Risk Evaluation Supplement, Version 5.0

Dear Commissioner Houdek:

Thank you for the opportunity to comment on Version 5.0 of the AI Risk Evaluation Supplement. I write as the founder of EasyEssence, an independent practice that evaluates insurers' consumer-facing AI outputs against their policy forms. We build none of the systems we review and take no role in tuning them.

The Supplement already asks the right questions about accuracy and model drift. Our comments concern where those questions sit and what triggers them. In short, the path from counting a consumer-facing AI Model to measuring its accuracy has gaps a company can fall through without meaning to. The evidence questions also stop a step short of what an examiner may need to reach a conclusion. Of the seven suggestions below, three are clarifications to existing definitions and instructions, three request new fields, and one asks that a current field be split. Suggested language for each appears in the appendix.

1. We suggest clarifying that an AI Model whose answer reaches a consumer has Direct Consumer Impact, whether the Model delivers that answer directly or a company representative relays it.

Carriers are increasingly deploying AI that drafts the answer to a consumer's coverage question for a company representative to read aloud. That configuration is difficult to place against the definitions as written. Direct Consumer Impact attaches to a Model that augments or automates a human decision, and Augmentation means advising a human who is making one. A representative reading a coverage answer aloud is conveying a decision rather than making it, so the Model sits between the two, and a company may reasonably place it in Support, which falls outside Direct Consumer Impact. Yet the answer is consumer-visible, which the Direct Consumer Impact definition treats as the mark of relevance, and the Supplement's Background covers decisions and actions impacting consumers that are made or supported by these technologies. One sentence in the definition would carry that through: an AI Model has Direct Consumer Impact where its output is communicated to a consumer, whether the Model delivers it directly or a company representative relays it. Without it, one insurer will count these Models and another will not, and the Exhibit A totals regulators rely on to scope inquiries will understate how much AI is speaking to consumers.

2. We suggest the Risk Assessment instructions state that a representative's review of a Model's answer before delivery, and automated checks within the Model's platform, are controls, reflected in Residual Risk rather than Inherent Risk.

Once a Model is counted, its rating decides what happens next. The Supplement distinguishes Inherent Risk, before controls, from Residual Risk, after them, and asks companies for the inherent view. What the instructions do not yet say is where the two safeguards carriers rely on most for consumer-facing AI fall. A company representative who reviews the AI's answer before reading it to the consumer is a control. So is a platform feature that checks the answer against policy language. A company that credits either when setting the inherent rating will rate the same Model lower than a company that does not. We suggest the Risk Assessment instructions name those two examples as controls belonging in Residual Risk, which would settle the point and make ratings comparable.

3. We suggest the Inherent Risk definition include Adverse Consumer Outcomes alongside economic loss and inaccurate financial reporting.

Inherent Risk is currently defined in financial terms. For a consumer-facing Model, the exposure the Background describes is an Adverse Consumer Outcome rather than a financial statement effect. A company applying the definition literally may find no financial loss to point to, even though the consumer exposure is real. We recognize the financial framing is new in Version 5.0, added to align the definition with the Financial Condition Examiners Handbook; the prior definition was neutral as to the type of exposure. We suggest inserting "or Adverse Consumer Outcomes" after "inaccurate financial reporting." The economic-loss and financial-reporting language stays as it is. Adverse Consumer Outcome is a term the Supplement already defines, so the change restores the prior neutrality with an existing term rather than a new concept. It also carries the Background's intent into the instruction that governs the rating.

4. We suggest the Model Inventory in Exhibit A add one item for each AI Model with Direct Consumer Impact: whether its outputs are tested for accuracy after deployment, and by what method.

Exhibit C holds the Supplement's questions about what a Model actually produced, and a Model reaches Exhibit C when it is classified as high-risk, a classification that rests on the company's own risk assessment. Nothing else in the Supplement leads a regulator there. A company may rate a consumer-facing Model below high risk by judgment, or for any of the reasons in points 1 through 3. When it does, there is no inquiry into accuracy at all: no question about whether the Model's answers were right follows. The AIS Program checklist's complaint-tracking item does not fill that gap, since it captures only the errors consumers noticed, and a wrong coverage answer is usually not noticed until a claim is denied. For the Models whose errors reach policyholders directly, one self-attested classification determines whether any question about output accuracy is ever asked.

The Model Inventory in Exhibit A is the place for a second signal alongside the company's own classification. It currently records each Model's name, use, Use Case, operational area, Inherent Risk level, and consumer and financial impact. We suggest one added item: for each Model with Direct Consumer Impact, "whether its outputs are tested for accuracy after deployment, and by what method." Without it, nothing in the company's responses tells a regulator which consumer-facing Models warrant a closer look. With it, the regulator has that information and keeps full discretion over what to do with it. The method is what makes the answer usable. For example, a company that samples the Model's outputs and checks them for accuracy is measuring the outputs, while a company whose representatives review each answer before delivery is describing a control around them. Both are legitimate, and with the method stated, an examiner can tell which one is in place.

5. We suggest separating Exhibit C's combined question on pre-deployment validation and ongoing monitoring into two fields.

The current field asks how a Model was validated before deployment and how its performance is monitored on an ongoing basis, and it names both explicitly. Because it is a single free-text response, the answer does not show which of the two was addressed, and answers are difficult to compare across insurers. The two are not interchangeable. An AI Model keeps changing after release in ways conventional software does not. The policy documents it draws on are amended, and the instructions steering it are tuned in operation. The mix of consumer questions shifts. Where the model version is not pinned, the model itself updates on its supplier's schedule. The Supplement recognizes this in defining Model Drift. A pre-deployment validation therefore describes the Model as it was at release, and only ongoing monitoring describes it during the period an examiner is reviewing. A company can have done the first and not the second, and a combined field will not show which. Separating them into two fields would.

6. We suggest a new Exhibit C field, following the monitoring field, asking who performed the validation and monitoring and their relationship to the Model's developer.

Exhibit C records how each Model was developed using four categories: internal, "owned" third party, true third party, or a mix. Exhibit B question 3, and 3c in particular, captures the company's validation and testing policy, including frequency, scope, and methodology, and that is the right place for policy. Exhibit C then asks two things about each Model: how it was validated before deployment, and how its performance is monitored on an ongoing basis. It does not ask who performed either.

We suggest one new field asking who performed the validation and monitoring described in the preceding fields, and how that party relates to the Model's developer. The answer would build on the categories Exhibit C already applies to development. Because the answer to who performed the work is sometimes another AI model, the choices would also include automated evaluation, with that model identified. That choice is worth capturing. Research presented at NeurIPS 2024 found that large language model evaluators give inflated scores to their own outputs, and that the bias strengthens as models become better at recognizing their own work.¹

The field would tell an examiner how much independent weight an accuracy result carries, using a distinction the Supplement has already drawn. Insurance regulation already draws that distinction in its core documents. Under the NAIC's Annual Financial Reporting Model Regulation (#205), an insurer's financial statements are certified by an accountant independent of the insurer. Larger insurers must also maintain an internal audit function that is organizationally independent. Banking supervisors' model risk guidance has expected validation to be independent of model development since 2011. In a presentation to the Working Group on August 13, AM Best described predictive AI governance as beginning with model risk controls, including independent validation.² The field asks only who did the work and how they relate to the developer.

7. We suggest a new Exhibit C field recording the date and period of the most recent check of a Model's live outputs for accuracy, with the Model's configuration at that time: policy forms in force, model version, instruction version, and number of outputs checked.

Exhibit C already asks for the last date of model testing. We suggest one parallel field for live outputs, recording as of the response when they were last checked for accuracy and over what period, along with the four configuration facts named above. An examiner could then see whether the Model was measured during the period under review, and those facts are what allow a later change in accuracy to be traced to its cause.

Separately from the seven suggestions above, we support retaining Exhibit B question 4's oversight questions for AI used by professional service providers, including actuarial, claim, MGA, and audit providers, and the new checklist item 3p on third-party models. The entity operating a consumer-facing Model is often not the insurer, and in jurisdictions that have adopted the Model Bulletin the insurer remains responsible for it.

Taken together, these suggestions connect the Supplement's existing questions to the AI Models consumers rely on, and none asks the Working Group to replace company self-assessment, which is the right foundation for this instrument. Each makes an existing question reach the exposure it was written for.

I would welcome the opportunity to discuss any of these points with the Working Group or NAIC staff, including at the October 8 meeting if time for interested-party comment is available.

Respectfully submitted,

Phillip Kangari

Founder, EasyEssence

pkangari@easyessence.com

cc: Commissioner Doug Ommen, Co-Vice Chair; Mary Block, Co-Vice Chair; Scott Sobel, NAIC; Miguel Romero, NAIC

¹ Panickssery, Bowman, and Feng, "LLM Evaluators Recognize and Favor Their Own Generations," NeurIPS 2024.

² Minutes of the Big Data and Artificial Intelligence (H) Working Group, August 13, 2026 (draft pending adoption), presentation by AM Best, offered for discussion purposes.

Appendix: Suggested Language

Point 1. Definitions, "Direct Consumer Impact," add after the enumerated examples: An AI Model has Direct Consumer Impact where its output is communicated to a consumer, whether the Model delivers that output directly or a company representative relays it.

Point 2. Risk Assessment instructions, add: A representative's review of a Model's answer before delivery, and automated checks performed within the Model's own platform, are controls and are reflected in Residual Risk rather than Inherent Risk.

Point 3. Definitions, "Inherent Risk," insert after "inaccurate financial reporting": or Adverse Consumer Outcomes

Point 4. Exhibit A, Part Two, Model Inventory, add item h, for each Model identified as having Direct Consumer Impact: Whether the Model's outputs are tested for accuracy after deployment, and by what method.

Point 5. Exhibit C, replace the combined testing field with two: (a) How the model was validated prior to deployment, including testing of outputs for factors such as model drift, accuracy, unfair trade practices, unfair discrimination, performance degradation, etc. (b) How the model's performance is monitored on an ongoing basis, including testing of outputs for factors such as model drift, accuracy, unfair trade practices, unfair discrimination, performance degradation, etc.

Point 6. Exhibit C, new field following the testing and monitoring field (or following (b), if point 5 is adopted): Who performed the validation and monitoring described above: internal staff responsible for the Model's development; internal staff not responsible for its development; "owned" third party; true third party; a mix of the above; automated evaluation (identify the evaluating model).

Point 7. Exhibit C, new field: As of this response, the date and period of the most recent check of the Model's live outputs for accuracy, and at that time: policy forms in force, model version, version of operating instructions in effect, and number of outputs checked.

Comments on the NAIC AI Risk Evaluation Supplement, version 5.0

Falsify OÜ · Reg. 17574308 · Narva mnt 5, 10117 Tallinn, Estonia · hello@falsify.dev · 9 September 2026 · Submitted to the Big Data and Artificial Intelligence (H) Working Group, 30-day exposure ending 29 September 2026

Dear Mr. Sobel and Mr. Romero,

Falsify OÜ (Tallinn, Estonia) submits two comments on version 5.0 of the AI Risk Evaluation Supplement exposed on 31 August 2026. Both concern a single point: the Supplement asks how a high-risk model was validated and when it was last tested, but does not appear to ask the company to identify the pre-deployment acceptance criteria against which validation was judged, or when those criteria were fixed.

Comment 1 — Exhibit C, new field after field 11

Exhibit C states that AI model risk criteria are set by the insurance company, and field 11 asks the company to discuss how the model was validated prior to deployment. A validation narrative is only as informative as the bar the model was measured against. Two models can report identical validation results while one was held to criteria fixed before testing and the other to criteria selected after the results were known. Field 12 records the last date of testing, but no field appears to record when the acceptance criteria were fixed, so an examiner cannot distinguish the two cases from the response.

Proposed field, designated optional, inserted after field 11:

"Pre-deployment acceptance criteria. State the metrics and thresholds the model was required to meet before deployment, when those criteria were fixed, and what record establishes that they were fixed before validation results were available. If the criteria were revised after validation results became available, state when and why. Where available, identify evidence whose chronology does not depend solely on the validation report itself, such as a version-controlled or independently time-evidenced record of the criteria."

This adds one row to a questionnaire the company already completes, and it parallels the Supplement's treatment of materiality, where the threshold used is disclosed to the regulator rather than left implicit.

Comment 2 — Exhibit B (Narrative), item 3a

Item 3a asks whether validation and testing are performed by someone independent from development. Independence of the tester addresses one source of bias; control over when the acceptance criteria are established addresses another.

Proposed addition at the end of 3a:

"...and whether acceptance criteria are established before validation begins, by whom, and how subsequent changes to those criteria are controlled and documented."

Interest statement. Falsify maintains an open specification for tamper-evident records of evaluation criteria and their timing. Neither proposal depends on PRML or on any particular tool. Evidence could come from an internal records system, an independently timestamped record, or another mechanism that preserves both the criteria and when they were fixed.

We consent to this submission being made public.

Kind regards,

Cüneyt Öztürk
Founder, Falsify OÜ
Tallinn, Estonia

**From Joseph Stephens
Founder, Spectralgraph LLC**

Members of the Working Group,

My name is Joseph Stephens and I run Spectralgraph, an independent testing laboratory in Enid, Oklahoma. We measure fabrication in large language models by reading the model's own internal signals instead of grading the text it writes. Our method is published, including a leave-one-model-out result of 0.852 AUC separating fabricated content from genuine on models it had never seen: <https://doi.org/10.5281/zenodo.21365654>

I'm not writing for any insurer or any vendor, and I'm not asking you to require anything of anyone. Just wanted to provide a few comments from my research.

1. Record how the model was being served, next to the result.

C-12 asks for the last date of model testing, and C-11 asks the company to discuss testing of model outputs. Both invite a company to report a result. I'd ask you to consider capturing the serving configuration alongside it.

This month we took one open weights model and a fixed set of prompts and ran our measurement. Then, one at a time, we changed four settings on the software serving the model. They're the kind engineers adjust to make a server run faster, like numeric precision and batch size. Each change moved the scores for individual subject areas, the biggest by about a quarter of their value. Some went up and some went down, so there's no single conversion that would put the new numbers back in line with the old ones. Upgrading to a newer version of the serving software, on newer hardware, moved those scores by up to a third.

Under all four settings, we saw no measurable loss in how well the measurement told fabricated answers from genuine ones. That means the instrument keeps working under each of those settings, and its readings compare only with readings taken under the same one.

So the consequence for the Supplement is narrow. If a company's number this period differs from its number last period, or from another company's, the form as drafted gives a reader no way to tell whether the model changed or the software serving it did.

2. Model limitations can take a number, not only a paragraph.

Item C-9 now asks about model limitations, and its parenthetical points at use cases and data inputs outside the domain of the training data. Most companies will answer it in prose. It doesn't have to be. Where a company can measure, the same limitation can be stated per subject domain with a figure beside it, on its own question types instead of a public benchmark. A blank box gets you a paragraph, and paragraphs are hard to compare across companies, or across years for one company.

3. Who is accounting for the vendor?

Where an AI system comes from a vendor, the insurer usually doesn't control the serving stack and may not hear about it when it changes. Anything handed over by a vendor is even less likely to stay

comparable than a result the insurer produced itself. Recording the configuration would help there too.

Those are the only three things I'm asking you to consider. I'll happily answer any questions you might have.

Respectfully submitted,

Joseph Stephens

Founder, Spectralgraph LLC

Enid, Oklahoma

contact@spectralgraph.ai

September 2026

The Honorable Nathan Houdek, Chair
The Honorable Doug Ommen, Co-Vice Chair
The Honorable Mary Block, Co-Vice Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners

Re: Comments on the AI Risk Evaluation Supplement, version 5.0 (exposure draft)

Chair Houdek, Co-Vice Chairs Ommen and Block:

I develop secure AI kernels. These comments focus on how the Supplement distinguishes systems that generate outputs from systems that execute actions, and what evidence is needed to assess human oversight of those actions.

Three definitions are new in version 5.0: Agentic AI, Automation, and Augmentation. The definitions preamble adopts terms from the NAIC Model Bulletin “where available,” and these three were not available there. The Working Group can therefore revise them without creating inconsistency with the Bulletin.

1. Distinguishing generation from execution

Rename “Automation” to “Autonomous Execution.” In insurance operations, “automation” already names the deterministic, non-AI path: business rules engines, robotic process automation, and straight-through processing. A compliance officer answering Exhibit C item 10 is asked to use a word that the company’s operations and IT staff use for the opposite of an AI System. The result is good-faith misclassification at Exhibit C and under-reporting at Exhibit A, varying by filer. The Supplement already has the right word. Exhibit B (Narrative) 1.g asks how the company assesses “autonomy, reversibility, and reporting impact risk.” The rename aligns the definitions with that question.

The current definition also has three structural problems. It is defined by absence (“does not involve human intervention”), so it says what is missing but not what the system does. It has no point in time: a system approved once at deployment and a system reviewed on every transaction both satisfy it. And it asks the wrong question. The question for a regulator is not whether a human was present but whose decision took effect.

I propose the following text for the Definitions and Appendix section:

“Responsible Party” refers to a natural person, or an identified role occupied by a natural person, accountable within the company for a given decision or act. A Responsible Party is expected to exercise that accountability with the degree of diligence reasonable in light of the scope of the decision or act and its potential for Adverse Consumer Outcomes or material financial impact.

“Support” refers to an AI System that produces output informing a decision or act but that neither recommends nor carries out a particular decision or act.

“Augmentation” refers to an AI System that recommends a decision or act to a Responsible Party, where the decision or act of that Responsible Party is what takes effect and the company retains a record identifying the Responsible Party and the decision or act. For classification purposes, an AI System that would otherwise constitute Autonomous Execution but meets either supervision-and-review exception in that definition shall be classified as Augmentation.

“Autonomous Execution” refers to the operation of an AI System whose decisions or acts occur absent ongoing supervision and review by a Responsible Party. An AI System is not autonomously executing where: (1) a Responsible Party maintains ongoing supervision and review of its decisions

or acts; or (2) a defined escalation path results in those decisions or acts being reviewed by a Responsible Party, such review occurring before the decision or act takes effect where its effect is not reversible, and otherwise within a reasonable time of escalation being triggered. In each case, records evidencing the supervision or review shall be retained in accordance with applicable record retention requirements.

And, for Exhibit C item 10, which requires a single selection:

Where an AI System operates in more than one mode, it shall be classified according to the most autonomous mode in which it operates, and the company may describe the conditions under which each mode applies.

These place the three categories on one axis defined by two questions: does the system recommend a decision or act, and whose decision takes effect? Support recommends nothing. Under Augmentation, the Responsible Party's decision takes effect. Under Autonomous Execution, the system's does. A system that acts on its own but whose acts are supervised, or reviewed through a defined escalation path, is classified as Augmentation, because a Responsible Party remains accountable for what takes effect. "Supervision" is a standard insurers already work under, and it does not require a person to review every output. "Responsible Party" requires an accountable person rather than just a present one. Review timing depends on reversibility, which is the pairing in 1.g, because review after an irreversible act is not supervision. Retention follows existing state requirements rather than setting a new period. The diligence expectation scales with the Supplement's own terms, Adverse Consumer Outcome and Material Financial Impact. It is stated as a second sentence so that the classification at item 10 does not require a compliance officer to judge a colleague's diligence. The rename would also carry through to the Direct Consumer Impact definition, the Exhibit A instruction ("augment or automate"), and the item 10 labels. The Agentic AI definition is well drafted and should stay as it is.

Count generative and agentic models separately in Exhibit A. Columns (5) and (8) combine "Generative/Agentic AI Models" into one count. Generative systems produce content; agentic systems take actions. Agentic operation is independent of the axis above, though agentic systems rarely operate under Augmentation. With the two combined, a regulator cannot tell from Exhibit A how many systems in Claims or Prior Authorization can act, and Exhibit A exists to decide whether to escalate to Exhibits B through D. Splitting each column in two is a one-line change.

2. Evidence of oversight of actions

The Supplement's own definition says agentic AI is distinguished by taking actions and actively executing a process. The exhibits then evaluate the system and its models: inventory, design, testing, data sources, and governance policy. No question asks for evidence about an executed action.

Exhibit B (Narrative) 2.b shows the gap. It asks whether there is "a human in the loop" and, in the same sentence, whether there are "high volume transactions." The second question matters because volume can empty out the first. But nothing asks whether the company could show that a human in fact reviewed a given transaction. The first is a policy attribute; the second is an evidentiary fact, and without it the Supplement cannot tell a policy of human review from human review.

That gap is the reason for both edits above. The proposed definitions do not add a records requirement. A company may classify a system as Autonomous Execution and keep no records. Records are the cost of claiming the less autonomous classification, so the burden falls on the party asserting supervision, and no regulator has to prove autonomy. Nor do the definitions prohibit high-volume approval. If a company classifies a system as Augmentation and its records show one adjuster approving four thousand recommendations an hour, the regulator can draw its own conclusion. The aim is visibility. Once agentic

systems are counted separately and autonomy is defined by evidenced supervision, the evidence question has a place in a future version.

Two terms are left open on purpose. “Reversible,” used here and in 1.g, is left to its ordinary meaning; for an insurance act it is concrete — a payment can be recovered, a denial re-adjudicated, a policy reinstated. “Reasonable time” is left for regulators to develop.

I would be glad to discuss these comments on the October 8 call.

Respectfully submitted,

Jason Crittenden

Founder, Shadow Dynamic Systems LLC

Chicago, Illinois

jason@shadowdynamicsystems.com

From Nicholas Templeman
Director, CSOAI Ltd

The Big Data and Artificial Intelligence (H) Working Group is charged with facilitating discussion on AI systems evaluation. I would like to offer one narrow input from an independent measurement body, for whatever use the group finds in it.

We publish signed measurement records for AI models answering frozen, published question banks. The per-item rows sit beside each score, so a regulator or reviewer can recompute the number instead of relying on the tester's assertion. The banks do not change between runs, which is what makes results comparable over time.

The part we think is most relevant to an evaluation discussion is how failure is handled. We withdrew 44 of our own signed results because reading the rows showed no model could have answered those questions incorrectly; the instrument could not fail, so a perfect score carried no information. The withdrawals are published with reasons. Separately, a signed result that is wrong is superseded in public rather than edited, with the original left online beside its replacement.

We do not certify, we issue no marks, and we take nothing of value from a developer whose system we measure. Verification is free and the data is CC-BY-4.0. If a written comment would be useful to the group, I am glad to provide one.

To: Big Data and Artificial Intelligence (H) Working Group, National Association of Insurance Commissioners, by email to Scott Sobel and Miguel Romero

CSOAI Ltd (England and Wales, Companies House 16939677; 3rd Floor, 86-90 Paul Street, London EC2A 4NE) publishes measurements of AI models at <https://councilof.ai>. We run models against frozen, published question banks, grade the answers by fixed rules, sign each result with Ed25519, and publish the files so that anyone can check them free of charge. Our results are measurements, not approvals. CSOAI is not an insurer and offers no underwriting, risk rating or advice. We are not proposing that insurers or regulators use our service. An AI assistant helped draft this letter. A person at CSOAI reviewed it, and every figure in it was read from the public file named beside it.

Our comments concern the questions an insurer answers about testing third-party AI models. These are the Exhibit B (Narrative) question "Discuss the validation and testing procedures performed on third-party vendor-supplied AI Systems", Checklist item 3p, and Exhibit C, Refs 1, 11 and 12. As drafted, these ask the company to discuss testing. They do not ask for the facts a regulator would need to check a reported test result later. The words "hash", "confidence", "interval", "sample" and "record" do not appear in the draft. We suggest five additions. Each is illustrated by an error or practice in our own published record.

1. Identify the test set exactly. Ref 1 asks for the model name and version, but nothing identifies the data the model was tested on. Proposed addition to Ref 11: "Identify each test or validation dataset by name, version and date and, where available, a cryptographic hash (for example SHA-256) of the file used." For example, one of our signed results names the SHA-

256 of its question bank, and the bank file published beside it hashes to that value. Without that link, two results citing the same dataset name may rest on different questions.

2. Report the count before the metric. Proposed addition to Ref 11: "For each metric, state the number of test cases scored and the number excluded, with the reason for each exclusion." In our withdrawal ledger, three results read accuracy 1 on 4 cases and 26 read accuracy 1 on 30 cases. Those are not the same finding.
3. State uncertainty, and whether differences were tested. Proposed addition to Ref 11: "State the uncertainty of each metric, such as a confidence interval, and whether any reported difference between models or versions was tested statistically, and by what method. A difference that was not tested should be described as untested." On our public board (<https://councilof.ai/api/gspc>), 14 axes compare AI models. Two have a statistical separation determination, and both are ties; the other 12 are marked UNTESTED. In one tie, the paired test of the leading model against a comparison model gave McNemar $p = 0.69$. A table of point scores alone would have read as a ranking.
4. Keep item-level records that a reviewer can re-grade. Proposed addition to the Exhibit B (Narrative) vendor question: "State whether the company holds or can obtain per-test-case results (input, expected output, observed output and grading outcome) sufficient to recompute each reported metric, or holds aggregate results alone." Our note at <https://councilof.ai/notes/item-evidence-regrade/> links a 30-row item file whose hash matches the value in the signed result. Counting the correct rows gives 4 of 30, the published 0.1333. Our correction C-2026-0914-01 shows why this matters. On 14 September 2026 we withdrew 44 signed results, 26 of which read accuracy 1 at 30 cases. Their signature, content identifier and per-item evidence checks all passed. The problem surfaced when we read the item rows: each prompt offered a single permitted answer, so no model could answer it wrongly. Proposed addition to Ref 11: "Describe how the company confirmed that the test could produce a failing result."
5. Correct by supersession, and do not treat a missing result as zero. Ref 12 asks for nothing beyond the last date of testing, and the Instructions let regulators accept a prior submission "if the prior response is still current and applicable." Proposed revision to Ref 12: "Last date of model testing, the model version tested, and any earlier reported result that has since been corrected, superseded or withdrawn, with the date and reason." Proposed new definition: "A test that was not performed, or whose result was withdrawn, should be reported as not tested, not as a zero, a pass or a fail." Our note at <https://councilof.ai/notes/supersede-never-edit/> shows a corrected result published as a new file. The original stays online, with a ledger row naming its replacement and the reason. The 44 withdrawn results remain listed at <https://councilof.ai/interop/mill-cards-signed/WITHDRAWN.jsonl>, each citing the correction. None was replaced by a zero, and those model-and-bank pairs now have no result until a sound test produces one.

These are drafting suggestions. None requires a particular tool or provider, and an insurer or its vendor could answer each from its own records. Our public corrections record, <https://councilof.ai/api/corrections>, listed 52 entries as of 15 September 2026. If any figure above is wrong, we will correct it there.

Thank you for the opportunity to comment.

Nicholas Templeman
Director, CSOAI Ltd
nicholas@csoai.org
<https://councilof.ai>