September 5, 2025

Commissioner Michael Humphreys
Chairperson, Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners

Re: Exposure Draft of AI Systems Evaluation Tool

Dear Commissioner Humphreys,

On behalf of the American Academy of Actuaries[1] (Academy), we appreciate the opportunity to offer some high-level comments on the exposure draft of the NAIC's Big Data and Artificial Intelligence (H) Working Group's (Working Group) AI Systems Evaluation Tool (Evaluation Tool). We believe that the development of a consistent Evaluation Tool will provide meaningful benefits for both regulators and the broader insurance industry.

Specifically, the four exhibits provide a structured yet flexible foundation for regulators to evaluate financial risks posed by AI usage for both consumers and insurers. The Academy has been closely engaged with the NAIC and with individual states as we all consider the implications of increasing AI use in the insurance and financial services industries. As we have noted previously, both in respect to the NAIC's AI model bulletin and in our more recent comments to the Working Group in response to your request for information on proposing a NAIC model law, one of the Academy's top priorities is to provide an actuarial perspective on the responsible use of AI in insurance, while promoting the critical role actuaries play in ensuring that AI applications remain transparent, fair, and aligned with sound risk management practices.

We particularly appreciate the Working Group's inclusion of:

- **Regulatory Consistency That Prioritizes State Flexibility.** A consistent approach to AI oversight is valuable, as it helps to minimize fragmentation and compliance burdens while offering an understood standard for consumers and industry alike. As drafted, the Evaluation Tool offers a common framework that allows for regulatory consistency while offering state-specific flexibility.
- **Support for Principle-Based Flexibility.** The tool's optional exhibits and tailorable checklists enable regulators to adapt oversight to a specific AI footprint and risk profile. We encourage the Working Group to maintain this principle-based flexibility, as AI holds significant potential to improve underwriting, claims, and consumer experience. At the same time, AI potentially increases risks of opacity and systemic bias. The principle-based, adaptable structure of the Evaluation Tool enables regulators to balance responsible innovation with appropriate oversight.
- **Consumer Protection and Fairness.** The Evaluation Tool offers regulators support in addressing issues of bias, fairness, and equitable treatment of policyholders. As has been discussed by the

[1] The American Academy of Actuaries is a 20,000-member professional association whose mission is to serve the public and the U.S. actuarial profession. For 60 years, the Academy has assisted public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.

Working Group and others, finding methodologies that work to acknowledge and address the need to ensure that risk classification and ratemaking remain transparent and fair is a priority for regulators, industry, and actuaries.

The Academy also supports the Working Group's intention to deploy the Evaluation Tool on a pilot basis in 2026, with the intention of evaluating its use and making any appropriate changes and refinements based on the experience.

We appreciate the Working Group's continued collaborative efforts to consider a pragmatic and forward-looking approach to this evolving area. If you have any questions or would like to discuss these comments further, please contact Will Behnke, the Academy's Risk Management and Financial Reporting policy project manager (behnke@actuary.org).

Sincerely,

Jason Kehrberg, MAAA, FSA
Vice President, Life Practice Council

Annette James, MAAA, FSA
Vice President, Health Practice Council

Susan Kent, MAAA, FCAS
Vice President, Casualty Practice Council

Steve Malerich, MAAA, FSA
Vice President, Rick Management and Financial Reporting Council

Brian Bayerle
Chief Life Actuary
202-624-2169
BrianBayerle@acli.com

Colin Masterson
Sr. Policy Analyst
202-624-2463
ColinMasterson@acli.com

September 5, 2025

Commissioner Michael Humphreys
Chair, NAIC Big Data and Artificial Intelligence (AI) (H) Working Group

Re: AI Systems Evaluation Tool

Dear Commissioner Humphreys:

The American Council of Life Insurers (ACLI) appreciates the opportunity to provide feedback on the exposed AI Systems Evaluation Tool as put forward by the Big Data and AI (H) Working Group. We would also like to take this time to thank regulators and NAIC staff for the substantial amount of work that has gone into this proposal thus far as a part of the Working Group's broader AI strategy.

ACLI recognizes the need for regulators to have an informational tool to improve their understanding of company use of AI and their management of the risks associated with the technology as it impacts consumers. The *NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers* (AI Model Bulletin) was an important step towards understanding regulatory expectations around AI governance, and we continue to support its adoption to provide uniform guidance to insurers. While we appreciate the intent behind the proposed tool, we have concerns that the current approach would impose significant burdens without a commensurate benefit. We encourage a more targeted and streamlined framework that better aligns with the intended outcomes while minimizing unnecessary complexity or cost.

ACLI is providing these comments with the understanding that the tool as exposed is intended to be used as a data collection survey tool to provide regulators with a broad overview of how companies are using AI Systems in the business of insurance. If this tool is intended to be used in another manner, e.g., certain parts used to supplement market conduct exams, other parts intended to supplement financial examinations, we may wish to provide additional comments in the coming weeks. Additionally, should the intent of the data collection be to supplement examination activity, we would encourage the working group to consider routing the relevant pieces of the tool

The American Council of Life Insurers 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 93 percent of industry assets in the United States.
acli.com

through the appropriate NAIC committees (i.e. D and E Committees) to ensure the data collected is appropriate for various examination needs and memorialized in the appropriate instructions and documents sent to companies.

To make this tool more efficient and useful for both regulators and companies, we propose companies submit the response to their lead regulator under their current statutory authority which is crucial to the critical confidentiality protection appropriate for this effort. The domestic regulator would then be able to share the information with other regulators in a manner that maintains those protections. For insurers, it is critical to be able to provide this information once in a consistent manner. Without this approach, companies could be required to complete this request in every jurisdiction in which they are licensed with different variations on the questions and parts of the tool used.

In addition to suggesting this be submitted to a company's domestic regulator, we also suggest that the initial response be conducted via a data call with a narrower set of questions as set forth below to provide regulators information on company use of AI Systems. The domestic regulator may have follow-up inquiries as needed. Were this ultimately to become a more regular data collection, we would still recommend the domestic regulator be the one to do the collection to avoid a fragmented approach if many states are repeating the same request.

As a way to make this effort as valuable to regulators as possible, we have additional recommendations for consideration. A streamlined tool would be appropriate for uniformity of reporting between states. The scope of the tool should be limited to AI use that impacts consumers. Should the intent be to provide certain pieces of this tool to financial examiners, this effort would benefit from input from additional NAIC committee(s) and stakeholders. Additionally, the questions related to testing should be removed as they exceed the expectations outlined in the applicable laws and AI Model Bulletin. We suggest using the term "Adverse Consumer Outcome" rather than "Consumer Impact" to align with the AI Model Bulletin. The concept of "high-risk" should be defined by the company consistent with its risk management program. Lastly, given the significant effort to complete such requests, this should start as a one-time effort to gather information and let that information guide any additional questions for individual company follow up.

We have the following feedback regarding the exhibits for future dialogue of the tool. We note that we are recommending changing the order of Exhibits A and B with rationale below.

Definitions:

- The definition of "Consumer Impact" is far too broad and potentially brings in functions not contemplated. For example, an AI summary of a call between a policyholder and a company representative could be interpreted to be in scope. We suggest using the term "Adverse Consumer Outcome" as defined in the AI Model Bulletin to ensure that it follows existing statutory authority. This term would replace instances of "Consumer Impact" throughout the tool.

Exhibit B:

- The current requirements of Exhibit B are excessive, will be difficult for companies to complete, and create a burdensome amount of information for regulators.
- We suggest streamlining this entire section to be a written description of the company's risk management program around the use of AI rather than the specific criteria of the narrative and checklist. This description should include how a company assesses a model as "high-risk" within the risk management program.

- From a logical flow perspective, this overview of the risk management program makes sense as the first part of the report submission (i.e., swap Exhibit B with Exhibit A as listed in the exposure document while considering our proposals as contained within this letter).

Exhibit A:

- An indication of whether an AI System is used within a particular function provides greater value to regulators than the number of models used in customer facing functions.
- The assessment of high-risk should be consistent with the company's risk management program as is discussed in the prior exhibit.
- The exhibit should be limited to models that impact consumers. Functions such as Investment/Capital Management, Reserves/Valuations, Product Performance, Catastrophe Triage, and Strategic Operations (HR, Reinsurance, etc.) should be removed as rows.
- As previously indicated, financial questions should be beyond the scope of the tool.
- There is much development work on AI that may not lead to implementation, so we suggest removing the "Planned to be Implemented" column from the table.
- With those proposed changes, the columns of the table should be as follows:
    - Are AI System Model(s) with potential for an Adverse Consumer Outcome Currently in Use?
    - Are High-Risk AI System Model(s) with potential for an Adverse Consumer Outcome Currently in Use?
    - Were there Consumer Complaint(s) Resulting from AI Systems in the Past 12 Months by Program Area?
    - AI System Use Case(s)

Exhibit C:

- This should be focused on models that may have a consumer impact. Suggest changing header to "Exhibit C: High Risk AI Systems with Adverse Consumer Outcome Model Details." Corresponding changes would be made in the purpose.
- The two questions on testing should be removed given they are beyond the scope of the applicable laws and AI Model Bulletin.
- To capture the question on data currently posed in Exhibit D, include the following row: "Provide a description of data sources used in high-risk models."
- Remove "Discuss how the model affects the financial statements, risk assessment or controls." As previously discussed, financial issues should be addressed by a separate NAIC committee, and companies will discuss risk assessments and controls in their responses to the governance and risk management exhibit.
- Replace "Discuss how the model is reviewed for compliance with state and federal laws" with "Discuss how the model is reviewed for compliance with the unfair trade practices act and unfair claims settlement laws."

Exhibit D:

- We suggest removing Exhibit D. The focus of this tool is on AI, not on third-party data, which is being addressed elsewhere at the NAIC.

Thank you once again for the opportunity to provide our feedback on the AI Systems Evaluation Tool proposal and we look forward to additional discussion soon.

Sincerely,

Colin Masterson

cc: Scott Sobel, NAIC; Miguel Romero, NAIC

Guiding Greater Health

601 Pennsylvania Avenue, NW
South Building, Suite 500
Washington, D.C. 20004

T 202.778.3200
F 202.331.7487
ahip.org

September 5, 2025

Commissioner Michael Humphreys, 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 maromero@naic.org and Scott Sobel ssobel@naic.org

Re: NAIC Big Data and Artificial Intelligence (H) Working Group Artificial Intelligence (AI) Systems Evaluation Tool

Dear Commissioner Humphreys:

On behalf of AHIP, thank you for the opportunity to provide feedback on the proposed AI Systems Evaluation Tool ("Tool").

As outlined in AHIP's June 30 letter concerning the Artificial Intelligence ("AI") Model Law Request for Information distributed by the NAIC Big Data and Artificial Intelligence Working Group ("Working Group"), we continue to emphasize the importance of creating balanced policies that help realize the potential of AI and its innovation, while promoting patient safety and building trust among patients and providers. Health plans are using AI tools today to improve member service, support providers and reduce administrative friction and consumer costs. Health plans view these technologies as tools, not as a replacement for human decision-making. The concept of keeping a "human in the loop" is core to health plan use of AI.

To ensure AI policies improve care and protect Americans, AHIP remains committed to these guiding principles for consideration:

- **Take a Consistent, Uniform Approach.** A consistent national approach to AI oversight would provide protection for all American patients while minimizing additional administrative burdens and costs.
- **Define "AI."** AI definitions and other terms consistent with the National Institute for Science & Technology's AI Framework to build a national shared language.

- **Rely on Existing Laws.** New policy should not duplicate existing laws and regulations and instead should only fill gaps in existing health data and consumer protection laws and regulations.
- **Provide High-Level Oversight.** Guardrails that permit flexibility should be established and technologies or standards that may become outmoded should not be named in law.
- **Promote Risk-Based Approaches.** Policies should point to risk-based standards and confine any third-party evaluation requirements or government audits to "high-risk" uses.
- **Protect Intellectual Property.** Policies should require developers to provide sufficient transparency for deployers and explainability for consumers and should not put American companies at a competitive disadvantage by requiring disclosure of proprietary information.
- **Engage in Public-Private Partnerships.** The government should look to learnings from the private sector and collaborate with stakeholders to advance AI and inform policy.
- **Advance Industry Standards.** Laws and regulations should defer to industry standards where they exist and fund efforts to address ongoing gaps as technology evolves.
- **Guard Privacy.** Health plans are already subject to robust privacy regulations. While a national privacy law is needed to extend patient protections to all entities that hold health care data, establishing a new private right of action would chill innovation and raise costs.

AHIP's comments are provided below in three overarching categories.

- Overall Comments
- General Questions and Comments
- Specific Exhibit Comments

Overall Comments

We appreciate NAIC's effort to develop a common Tool that can be used consistently by all state regulators. A unified approach has the potential to reduce regulatory fragmentation and promote clarity for insurers operating across multiple jurisdictions. We, however, respectfully express concern regarding the current version of the Tool, particularly its breadth, operational burden and misalignment with the NAIC Model Bulletin on Artificial Intelligence ("Bulletin").

The volume and repetitive nature of the Tool, if adopted by regulators in its current form, would be difficult if not impossible to complete without significant staffing investments that could offset efficiency gains through new technology.

The Bulletin outlines a principles-based framework for AI System Governance, Risk Management, and Use Protocols, emphasizing insurer accountability and the demonstration of effective oversight through information and documents outlined in the Bulletin. Most importantly, the Bulletin repeatedly focuses on consumer protection, such as including definitions of "Adverse Consumer Outcome" and "Degree of Potential Harm to Consumers." In contrast however, the Tool appears to extend beyond the scope of the Bulletin (and its underlying model laws) by expanding the inquiries and evaluations into **all** uses of AI, without any limitations to those uses which have consumer impact, or are "consumer-facing", or are defined as "high risk" as some state legislation has done. We urge the NAIC to carefully review and compare the Tool and the Bulletin to ensure the Tool is appropriately aligned with the Bulletin's expectations and underlying laws.

Additionally, given the rapid pace of AI innovation, we are also concerned that a static Tool may quickly become outdated, whereas principles-based guidance could offer more flexibility and adaptability. We encourage the NAIC to consider refining the Tool to adopt a more risk-proportional and scalable structure and reinforce insurer-led accountability while maintaining consistency across states.

General Questions and Comments

In addition to the overall concerns expressed above, AHIP offers the following series of general questions and comments on the proposed Tool.

1. **Determine whether the Tool is a financial tool or a market conduct tool**. To avoid a patchwork of inconsistent state regulations, a determination must be made on whether this type of evaluation is a financial tool, or a market conduct tool, and whether it will be a required examination tool, or merely an option. AHIP recommends that the Tool be developed and improved in order to be part of a financial examination and be used to build additions to the NAIC Financial Examiners and Analysts Handbooks. This would be used by examiners to gain critical information in preparation for regular 5-year exams, as well as any called special examinations. This serves not only to minimize inconsistencies state to state but also brings clarity and sets expectations for insurers. It sets expectations for

regulators as well and allows for state-to-state exchanges and training between examiners and analysts on best practices, exam process efficiencies and other improvements. AHIP also recommends that the use of this Tool be coordinated by the "lead state" for multi-state insurers, to avoid wasteful duplicate requests by states. Similarly, insurers should have the option of responding on behalf of each legal entity.

If NAIC moves toward a pilot, AHIP respectfully requests a published pilot protocol, clear selection criteria, and an evaluation rubric to help guide state use.

2. **Compare with the current NAIC Market Conduct Examination Handbook.** To enhance this Tool's efficiencies, each exhibit should be methodically compared to existing questions and topics in the NAIC's Market Conduct Examination Handbook to ensure there are no overlapping materials. Additionally, it appears that several exhibits overlap with the Bulletin. A formal mapping exercise against both the Market Conduct Examination Handbook and the Bulletin before adoption would reduce redundancy and make the Tool easier to operationalize.

3. **Consider developing a materiality threshold.** A defined materiality threshold should be discussed and determined. This may take into consideration not only the size of the company, but the types and number of AI systems in use, including whether the AI systems are "high risk."

4. **Address concerns with the substantive breadth and detail of the data requests in this Tool.** The administrative burden of pulling together and maintaining all the information currently included would be substantial, especially across all the exhibits in the Tool. We recognize there is some value in certain inquiries around insurer enterprise governance practices, policies and control frameworks. However, the requests for details around **all** AI systems utilized, **every** use case being considered or deployed, **all** initiatives related to AI would be very burdensome given much if not most of the information may be obsolete in potentially short periods of time. We further note that these burdens are only magnified exponentially if updates to this information are expected annually, or even quarterly, as the Instructions appear to contemplate. We have noted this concern in some, but not all, locations in the comments below.

- A possible solution for the concerns above might be to limit the AI uses to which this Tool applies to "high-risk" uses. As currently proposed, the term "Consumer Impact" appears overly broad and could be interpreted to include an AI system used for things such as designing the layout for an advertising brochure or mailings. Regulators should consider defining terms such as this carefully and critically to provide clarity for regulators and insurers and avoid unintended application. The NAIC should consider anchoring a common "high-risk" definition. Otherwise, states could adopt their own interpretations, leading to inconsistent reporting and compliance obligations. To help inform that work, AHIP suggests drawing on NIST AI RMF or OECD criteria could help give regulators and insurers a consistent baseline for "high-risk" across jurisdictions and reviewing state legislation utilizing "high-risk" uses of AI. The use of such limiting language, or similar language, could make this Tool much more targeted and regulatorily efficient – the responses from insurers would be much more informative and useful for regulators.

5. **Consideration of a "range" approach.** In the event no limitations are applied to the Tool and the questions remain relative to inventories (e.g., the first question in Exhibit A), consideration should be given to the need for, and usefulness of, allowing answers to some questions involving large numbers to be answered in the form of ranges (e.g., between 500 and 1,000) as opposed to exact figures.

6. **Confidentiality and process issues necessitate attention.** If the Tool is used within examination or analysis processes, responses should clearly fall under the confidentiality protections that already apply under state law and the NAIC Handbooks. If, however, the NAIC envisions using the Tool outside those channels, regulators need to specify how responses will be safeguarded and confirm whether public records requests at the state level could expose confidential and proprietary information.

Specific Proposed Comments on Exhibits, Instructions, Definitions and Appendix

In addition, below are AHIP's more detailed comments relative to the Exhibits, Instructions, Definitions and Appendix.

Exhibit A: Quantify Regulated Entity's Use of AI systems

- The term "AI Systems Model" is used in four of the five column headings; however, this term is not defined or explained. AHIP recommends a definition or explanation of the term or that it be deleted.
- The column calling for all consumer complaints for the past 12 months, without limitation, would be extremely burdensome and, in some cases, not possible to report at that level of granularity.
- The last column asks for AI System Use Case(s). Given the potential breadth of AI uses, AHIP suggests this be limited to instances involving significant consumer impact, material financial impact, or "high-risk" cases and that terms be defined.
- In Consumer Complaints, Section 1, AHIP urges removal or significant revision of the proposed consumer complaint metric. As proposed, the metric is problematic for the following reasons:
    - Current health insurer systems track complaints by issue type (e.g., denial, access), not by technology type (e.g., whether AI was involved). Capturing this data would require new infrastructure and processes and divert resources.
    - Health insurance is already subject to extensive consumer protections under federal and state rules, including formal complaints, review and appeal pathways that apply regardless of the technology used.
    - The phrase "relied on AI system" is unclear and overly broad. If this question remains, we'd recommend it be changed to "solely used AI system(s) for critical decision" (or "denials"). The better solution is the removal of this metric. A more effective approach would provide support for and alignment with the NAIC AI Model Bulletin and focus on documentation related to internal governance and monitoring processes for AIS, rather than consumer-facing complaint metrics.

- As drafted Exhibit B requires governance information that would show regulators how complaints and consumer harms are tracked and mitigated. As a result, the information requested in Exhibit A appears to be duplicative.

- A definition of "AI Systems Model" should also be considered since this term is used extensively in Exhibit A but is not explained in the Tool.

Exhibit B: AI Systems Governance Risk Assessment Framework

- Narrative Question Two is overly broad and should be deleted here. It is more effectively divided up in Exhibit D.
- The information desired in Question Three is not clear. Please reword or clarify.
- Is Question Four intended to apply to all AI Systems? We submit this question as an example of many instances in which language should have limitations for clarity and to avoid unintended application.
- The answers sought by Question Five appear to be included in those in response to Question Four. Is there a difference?
- In Question Six, please define, clarify, or limit the term "Open-source AI." As currently proposed, this question, and Question Six would be highly burdensome.
- As currently proposed, Question Seven in the Tool is overly burdensome and beyond NAIC requirements and thus should be removed. This information is highly volatile, competitively sensitive.
- As proposed, Question Four (c) requests more detail than is required in the NAIC AI Model Bulletin. Because most insurers do not generally break this down to the subcategories as proposed in the Tool, AHIP suggests the question be revised to be more generalized.
- Question Five (b) is unclear. AHIP believes the proposed question is related to quality assurance; however, we would request clarification. We have similar concerns about Questions Five (c) and (g). Are these intended to determine whether the AI Systems were fit for purpose? Or if, once developed, an AI Systems was found to have more uses than originally intended? If so, we request clarification on how NAIC believes this question, and the information collected, will assist regulators?

Exhibit C: AI Systems High-Risk Model Details

- In the Purpose language, AHIP recommends it be revised as follows:

  Purpose: To obtain detailed information on high-risk AI system models, such as those solely making automated decisions, which could cause adverse consumer, financial, or financial reporting impact.

  We also recommend more specificity on these terms to avoid overbreadth and confusion. An example would be to clarify that Exhibit C applies only to "those solely making automated decisions *without human review*." This would clarify that the Exhibit is limited to fully automated AI, and that human review in these processes mitigates the "high-risk" designation. Additionally, Exhibit B already requests that plans disclose AI governance and risk frameworks which will include descriptions of human involvement & oversight. Another suggestion would be to clearly identify and define "high-risk" as previously suggested.

- In the existing Purpose language, is the last sentence intended to mean that insurers would decide which of their AI Systems uses are "high risk"?

Exhibit D: AI Systems Model Data Details

- The requested granularity around "Model Data Details" and "Type of Data Element Used" goes beyond what is required under the NAIC AI Model Bulletin and risks creating a second, duplicative compliance track. These should be revised to align with the Bulletin.

Definitions, Appendix, and Instructions

- To ensure the Tool serves as a complement to the Bulletin, again we recommend NAIC review the Evaluation Tool and the Bulletin for consistency across all provisions. This will help ensure state departments of insurance have clear, usable tools to support effective implementation. Additionally, alignment with NIST AI RMF terminology would reduce drift across frameworks and give both regulators and plans a shared vocabulary.

- To increase clarity, AHIP suggests the inclusion of a definition of "consumer." AHIP is concerned about the definition of "Internally Trained Models." It implies it is a model trained in internal data, regardless of where the model itself or underlying technology was developed. For instance, if an insurer sent its data to a vendor who then fine-tuned a model for the insurer, by this definition that would be considered internal. We submit there should be a distinction between whether the data is internal or, alternatively, the model's IP was developed internally. We see the same issues in the definition of "Externally Trained Models." From these two definitions it would be possible to create a model that is both internal and external, either or neither. Is the goal of these definitions to identify "who owns the IP" of a given model and therefore who can speak to how it was built, or whether the data is public (i.e., available to anyone on the internet), private (from a vendor), or private (owned by the company)? We believe these definitions need more discussion and revision to be consistent and useful in achieving regulators' information-gathering goals.
- The instructions suggest quarterly or annual updates, which would create an ongoing reporting cycle. We recommend linking the updates to regulator requests in connection with an exam or targeted review, so that reporting is purposeful and context driven.

Thank you for the opportunity to provide these comments. We look forward to discussing them and continuing to work with the Working Group in the future.

Sincerely,

Miranda Motter
AHIP Senior Vice President, State Affairs and Policy
mmotter@ahip.org
202-923-7346

AHIP 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.

September 5, 2025

Mr. Scott Sobel
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106-2197
Email: maromero@naic.org

Subject: AI Systems Evaluations - Optional Supplemental Exhibits for State Regulators

Mr. Sobel:

The American InsurTech Council (AITC) is an independent advocacy organization that is supported by a diverse group of sponsors and supporters that include legacy insurers, insurtech startups, technology developers, and other stakeholders in the insurance and technology sectors. These entities share a common interest in advancing ethical, technology-driven innovation within the insurance industry to improve business practices and consumer experiences.

I. Overview

AITC recognizes the significant effort by members of the Big Data and Artificial Intelligence (H) Working Group to develop the above referenced *AI Systems Evaluations - Optional Supplemental Exhibits for State Regulators* ("draft proposal"). We appreciate the opportunity to provide our comments and look forward to engaging with the Working Group as discussions on regulatory frameworks for insurer use of AI continue.

Our comments are intended as constructive contributions to the ongoing collaborative dialogue between regulators and stakeholders as we work together to establish an effective regulatory framework for insurer use of AI. Regulators, insurance carriers and developers share a common goal of ensuring robust consumer protections and regulatory oversight, while preserving the innovation environment that enables insurance companies to develop and deploy AI

technologies that ultimately benefit consumers through improved services, enhanced efficiency, and competitive market dynamics.

1. Support for the Optional Approach

We strongly support the NAIC making the exhibits optional rather than mandatory requirements. This approach provides flexibility to states, and to companies working to adopt AI governance practices appropriate to their scale and risk profile.

2. Level Playing Field Issues

Any future guidance or best practices should explicitly acknowledge the need for proportionate implementation based on carrier size, AI complexity, and risk exposure. A one-size-fits-all approach would disadvantage smaller carriers that lack the resources of large insurers. This is an important consideration for consumers, as smaller and mid-sized insurers are often better positioned to leverage AI due to fewer legacy systems and entrenched processes that can impede rapid technological adoption. We also recommend creation of safe harbors for preliminary research and planning activities.

3. The Draft Proposal Departs from the NAIC's Commitment to a Risk-Based Approach

Since adoption of its Principles for Artificial Intelligence in August, 2020, the NAIC has consistently maintained a firm commitment to a principles, risk-based approach to oversight of insurer use of AI and other emerging technologies over a highly-prescriptive, "one size fits all" approach.

The NAIC's choice of a risk-based framework acknowledges the reality of developing meaningful regulatory guidance in a rapidly changing environment. A regulatory framework for AI must work not in the brief time after its adoption, it must be both *durable and flexible* to remain relevant as the technology continues to evolve. A risk-based framework focuses on outcomes and potential harms – rather than specific technical implementations. This approach ensures the continued relevance of the framework as technology advances. Prescriptive frameworks, on the other hand, would be quickly rendered obsolete by technology advances, requiring constant updates to the regulatory framework that would always find itself running behind cycles of innovation.

Other important principles established in the Principles for Artificial Intelligence include:

- Outcome-focused standards (such as fairness, accuracy, and transparency) that apply regardless of the underlying AI technology used.
- Application of the principles all persons or entities facilitating the business of insurance that play an active role in the AI system life cycle, including third parties such as rating, data providers and advisory organizations, as well as insurance carriers.
- The importance of accountability, compliance, transparency, and safe, secure, fair and robust outputs while promoting innovation and accommodating technological change.

The NAIC Model AI Bulletin was a critical next step in the development of regulatory guidance. The Bulletin extended the Principles for Artificial Intelligence, establishing a risk-based regulatory approach that emphasizes:

- **Proportional oversight** based on system risk levels and consumer impact.
- **Focus on high-risk applications** that could materially affect consumers.
- **Graduated requirements** matching regulatory intensity to actual risk.
- **Preservation of innovation** for beneficial AI applications.
- **Efficient resource allocation** targeting genuine consumer protection needs.

The Principles for Artificial Intelligence and Model AI Bulletin served as the benchmark for our analysis of the draft proposal. Essentially, we asked ourselves: "is a given component of the draft proposal consistent (or inconsistent) with the risk-based approach that has already been established by existing NAIC guidance?"

In our view, the draft proposal departs in a number of material ways from the approach described in The Principles for Artificial Intelligence and Model AI Bulletin:

A. Departure From Risk Differentiation

Model AI Bulletin: Recognizes different risk levels and tailors oversight accordingly.

Draft Proposal: Applies uniform disclosure requirements to all AI systems regardless of:

- Consumer impact potential
- System complexity or autonomy
- Financial materiality
- Regulatory significance

B. Scope Expansion Beyond Consumer-Facing Systems

Model AI Bulletin: Focuses primarily on AI systems with direct consumer impact or material business consequences.

Draft Proposal: Requires comprehensive disclosure across all business operations including:

- Internal HR and administrative systems
- Back-office operational tools
- Investment and capital management
- Legal/compliance document processing

C. Elimination of Materiality Thresholds

Model AI Bulletin: Implicitly recognizes that not all AI usage poses material risk requiring regulatory attention.

Draft Proposal: Requires disclosure of all AI systems without any materiality filters for:

- Transaction volume
- Financial impact
- Number of consumers impacted
- System sophistication level

D. Process-Heavy vs. Outcome-Focused Regulation

Model AI Bulletin: Emphasizes risk management outcomes and consumer protection results.

Draft Proposal: Mandates detailed process data source mapping, documentation and system inventories regardless of effectiveness, materiality, risk to consumers, or risk mitigation.

E. Innovation Impact Inconsistency

Model AI Bulletin: Balances consumer protection with continued innovation and competitive markets.

Draft Proposal: Creates a highly prescriptive, costly administrative and compliance burden that will create barriers to entry and beneficial AI development, particularly the case for smaller companies.

4. The NIST AI RMF (1.0) Provides a Better Framework for Regulatory Oversight Purposes

The NIST AI Risk Management Framework (1.0) ("NIST AI RMF") is a practical, risk-based structure for AI governance that aligns well with regulatory priorities, goals and objectives. We respectfully urge the Working Group to consider utilizing the approach taken in the NIST AI RMF as a guide for the development of a regulatory framework for the following reasons:

- It is expressly risk-focused, avoiding reliance on the highly prescriptive components and requirements contained in the draft proposal.
- Provides a practical, risk-based structure for AI governance that aligns with the NAIC's stated policy determination favoring a risk-based approach:
    - **Risk-Based Prioritization.** Stresses proportionality: risks should be managed based on the likelihood and impact of harm. For example, complaint tracking

should focus on high-risk applications (e.g., underwriting, pricing, claims adjudication) rather than low-risk, routine tools (e.g., chatbots).
    - **Contextualization of Harms**: Encourages organizations to identify *how and where* AI may cause adverse effects. For complaints, this suggests distinguishing between consumer harm complaints (e.g., unfair claim denials) and service complaints (e.g., delays), rather than lumping them together.
    - **Documentation and Traceability**: Calls for documentation of AI lifecycle decisions in a way that supports accountability. This could replace blanket reporting by enabling insurers to document how they assess, categorize, and escalate complaints related to AI.
    - **Flexibility and Adaptability**: Designed to be adaptable across industries and organization sizes. This approach would help regulators avoid prescriptive, one-size-fits-all mandates and instead allow for tailored reporting obligations based
    - on insurer scale and risk profile.
- The NIST AI RMF has already been endorsed by the NAIC in the Model AI Bulletin for use by carriers in the design and implementation of their own AI governance and risk management frameworks: "The AIS Program may adopt, incorporate, or rely upon, in whole or in part, a framework or standards developed by an official third-party standard organization, such as the National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework, Version 1.0." (General Guidelines, 1.5).
- Many carriers are already using (or will be in the future) the NIST AI RMF (or a similar approach) in their own AI risk management program. A regulatory oversight framework that aligns with the NIST AI RMF would ensure consistency between carrier risk management and regulator expectations, streamline the regulatory oversight process, provide deeper insight into carriers' compliance with state laws regarding their use of AI, be easier to design and implement, while reducing costs and administrative burden.
- Because the NIST AI RMF will be updated on a continuous basis to keep pace with the rapidly changing environment involving use of AI, any changes needed to keep the regulatory framework current will be easier to implement.

II. Summary Comments On Exhibits A – D

Exhibit A: Quantify AI Systems Use

- **Overly broad AI definition**: captures basic automation and spreadsheet macros across 15 operational areas.[1]
- **No materiality thresholds** : treats simple tools same as complex algorithms.
- **Extends beyond consumer impact:** into HR, investment, and internal operations.
- **Vague key terms**: "Consumer Impact" and "Material Financial Impact" undefined.

[1] We continue to struggle with imprecise definitions of key terms that can create confusion and unintended consequences. Prescriptive frameworks accentuate this problem, while risk-based approaches afford some mitigation.

- **Future disclosure overreach**: requires 6-month strategic AI plans.
- **Complaint attribution without causation**: any AI-adjacent complaint counts.
- **Contradicts risk-based approach**: uniform requirements regardless of risk level.

Exhibit B: AI Governance Framework

- **More easily accomplished utilizing the NIST AI RFM Framework.**
- **Excessive scope beyond regulation**: demands internal IT and HR governance details.
- **Confidential strategic information exposure**: investment amounts and future initiatives**.**
- **Complete vendor relationship disclosure**: names, contracts, testing methods**.**
- **Proprietary governance framework details**: competitive intelligence exposure.
- **Board and organizational structure requirements**: beyond traditional regulatory scope**.**
- **Open-source software inventory demands**: potential security vulnerability exposure**.**
- **Employee policy and training disclosure**: internal HR confidentiality issues.

Exhibit C: High-Risk Model Details

- **Company self-defined risk criteria**: creates regulatory inconsistency and the potential for gaming.
- **Proprietary technical methodology exposure**: testing approaches and performance thresholds.
- **Vendor identification requirements**: may violate confidentiality agreements.

Exhibit D: AI Model Data Details

- **Disclosure of Intellectual Property**: Demands disclosure of IP and other highly confidential information that forms the core competitive advantage of insurance companies' AI systems, seeks protected trade secrets under both state and federal law, and raises significant concerns about proprietary business intelligence exposure.
- **Comprehensive data source mapping**: reveals entire AI development strategy.
- **Vendor relationship exposure**: requires disclosure of competitive intelligence on data partnerships.
- **Operational practice disclosure**: compels disclosure of detailed business process intelligence.
- **No risk differentiation**: applies the same requirements for high and low-risk applications.
- **Trade secret methodology exposure**: requires disclosure of data usage approaches that constitute IP.
- **Extensive vendor naming requirements**: disclosure of certain terms may breach confidentiality agreements.

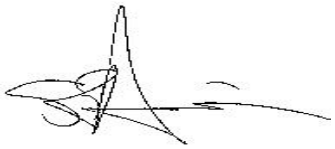
- **Internal capability revelation**: exposes competitive advantages and limitations.

Issues Applicable to All Exhibits

- **Abandons the risk-based approach established in NAIC Model AI Bulletin.**
- **Disproportionate small company burden**: favors large companies with significant compliance resources.
- **Innovation disincentive**: extensive disclosure may discourage beneficial AI development.
- **Competitive intelligence exposure**: provides roadmap for competitor replication.
- **Confidentiality protections:** trade secret safeguards need to be clearly defined.
- **Regulatory overreach**: extends into areas beyond traditional insurance regulation.
- **Implementation burden** - creates administrative costs without proportional consumer benefit.
- **Market consolidation risk** - compliance costs may force smaller players out of AI development.

Thank you again for the opportunity to address our comments. We look forward to further discussion of these issues.

Respectfully Submitted,



Scott R. Harrison
Co-Founder, American InsurTech Council
sharrison@americaninsurtech.com

cc: Michael Humphries, Chair, Big Data and Artificial Intelligence (H) Working Group

American Property Casualty Insurance Association
INSURING AMERICA apci.org

September 5, 2025

Commissioner Michael Humphreys
Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners

Re: Draft AI Systems Evaluation Tool Comments

Dear Commissioner Humphreys:

The American Property Casualty Insurance Association (APCIA) appreciates the opportunity to provide input in response to the Big Data and Artificial Intelligence (H) Working Group's draft AI Systems Evaluation Tool. APCIA promotes and protects the viability of private competition for the benefit of consumers and insurers, with a legacy dating back 150 years. APCIA members represent all sizes, structures, and regions—protecting families, communities, and businesses in the U.S. and across the globe.

General Comments/ Scope

APCIA is concerned about the extensive level of detail and amount of information that the proposal is seeking. The proposal includes dozens of questions seeking highly detailed responses without clarity regarding when companies will be asked for this information and how often companies will be required to update their responses. Adding this level of granularity and focus on a single topic would dramatically increase the workload associated with audits and examinations, as implementing the proposed exhibits and documentation standards would demand substantial time and resources from both insurers and regulators, which can potentially stifle innovation that would otherwise benefit consumers.

In contrast to the proposal, the NAIC's AI Model Bulletin embraces a risk-focused approach to AI regulation. For example, the Model Bulletin provides that an insurer's "AIS Program should be reflective of, and commensurate with, the Insurer's own assessment of the degree and nature of risk posed to consumers"; that "the AIS Program should be tailored to and proportionate with the Insurer's use and reliance on AI and AI Systems"; and that controls and procedures "should be focused on the mitigation of Adverse Consumer Outcomes" and "should reflect and align with the Degree of Potential Harm to Consumers with respect to that use case." We therefore recommend that the NAIC prioritize developing additional, uniform, and flexible guidelines for carrying out the regulatory expectations enumerated in the AI Model Bulletin.

If the Working Group proceeds with the proposed AI Systems Evaluation Tool, it should explicitly state that it is intended to work coextensively with the Model Bulletin. Additionally, we recommend implementing the proposal incrementally, beginning with an initial pilot phase to collect data on which questions are most effective. The initial pilot phase could then inform further development of the proposal so that it focuses on collecting only the information that is of greatest regulatory concern without creating an undue administrative burden.

As part of the pilot phase, the initial scope of the proposed AI Systems Evaluation Tool should be limited to AI systems 1) that are high-risk (as described below), 2) that directly make or recommend decisions, and 3) such decisions may result in an Adverse Consumer Outcome (using the same definition of "Adverse Consumer Outcome" as the AI Model Bulletin). We believe this would result in a more risk-focused approach that, compared to the current proposal, is significantly less likely to stifle innovation that could benefit the consumer and bring about business efficiency. After the pilot phase and any subsequent revisions to the proposal, companies should be allowed ample time to comply with the expectations laid out in the AI Systems Evaluation Tool.

We would also appreciate further clarity regarding whether the intention is to officially incorporate this proposal within the NAIC Market Regulation Handbook or the NAIC Financial Examination or Analysis Handbooks, either during the initial pilot phase or as a longer-term plan. Additionally, the proposal should clarify that the AI Systems Evaluation Tool should only be utilized by an insurer's lead state regulator. This would provide important context for both companies and regulators regarding the circumstances under which insurers will be asked to complete the AI Systems Evaluation Tool.

The proposal should also allow companies to complete the questions on a group-wide basis. As drafted, the exposure requires companies to complete the questions separately for each line of business, which would be extraordinarily burdensome and inefficient for companies writing multiple lines of business.

Confidentiality

The proposal requests highly sensitive and detailed information about insurers' models, data sources, and other proprietary information. Since the context for when insurers will be expected to complete the proposal is unclear, it is also unclear whether state confidentiality laws would apply to companies' responses to the proposed questions. Companies should only be asked to turn over confidential or proprietary information if the confidentiality of the answers and related documentation will remain subject to the highest level of confidentiality protections afforded under state law. The proposal can accomplish this by clearly stating that the questions should only be used under a state's existing examination authority.

Introduction

To reduce confusion, we recommend adding instructions to the Introduction section clarifying that the definitions from the AI Model Bulletin are used, as applicable, and inform the reader that additional definitions may be found in the definitions and appendix section. Additionally, the sentence including the phrase "the insurer's AI philosophy" should be reworded to clarify the intent.

Exhibit A

The first sentence of Exhibit A indicates that the purpose of the exhibit is to obtain information on "AI models that are new, updated, retired, etc.…" If retired models are included in this exhibit, a cutoff date should be specified. For example, the scope should be limited to models retired during the period being reviewed or in the previous 6- or 12-month period.

Use of AI System in Operations or Program Areas

APCIA is concerned about the extensive number of "operations or program areas" in the first column of Exhibit A on page four. We recommend that the number of operational or program areas be limited to those considered regulated insurance practices or activities. This would reduce the administrative burden on insurers while maintaining alignment with the scope of the AI Model Bulletin and the areas of greatest regulatory concern. Alternatively, if the number of operations or program areas is not streamlined, it would be helpful to provide additional clarity on what each operations or program area encompasses. For example, what is meant by "utilization management" and "product performance." Additionally, some of the categories like "producer services" and "customer service," and "claims/adjudication and "reserves/valuations," may be duplicative. Overlapping categories should be combined or the proposal should provide additional context about what should be included in the various categories.

In lieu of the extensive information requested by Exhibit A's "operations or program areas" chart, the Working Group should consider whether a more efficient approach would satisfy the same regulatory goals, such as allowing companies the option to submit an inventory of their high-risk AI Systems used in regulated insurance practices, instead of completing the chart.

Consumer Complaints

The first question in the "Consumer Complaints" chart broadly asks for total consumer complaints "resulting from a process that relied on AI systems." In order to balance regulatory impact with the administrative burden placed on insurers, this question should be rephrased to say, "What is the total number of consumer complaints directly resulting from decisions made or recommended by an AI system(s) in the past 12 months?" Importantly, it will take time for insurers to develop appropriate reporting capabilities, as it will not always be self-evident that a complaint was a direct result of an AI system. Consequently, we recommend removing this question for an initial pilot phase of the proposal, or at a minimum, providing for an extended adoption timeline for this question.

Likewise, the second question, which asks for the number of consumer complaints by program area, should be reworded to ask for the number of complaints that were the direct result of AI systems. We similarly recommend removing this question during the initial pilot phase. Additionally, we are concerned that the second question presumes that there is an express requirement for consumer disclosure of AI use in the AI Model Bulletin. The proposal should be aligned with the Model Bulletin, and if a question to this effect is included, it should request the insurer to explain whether it has consumer disclosure requirements, and if not, to explain its reasoning and risk assessment.

Finally, we note that the second and third questions refer to only "AI," rather than "AI systems." We recommend using the term "AI systems" or clarifying whether use of only "AI" was intentional.

Exhibit B

APCIA recommends eliminating the narrative option of Exhibit B and only moving forward with the checklist. The checklist is a more efficient and a less laborious method of collecting the necessary information. It encourages dialogue between the regulator and licensee if more information is necessary. Most importantly, it allows for consistency among jurisdictions, which is critical as companies continue to explore this rapidly evolving technology. As noted in the checklist, regulators are encouraged to coordinate; regulators being on the same page with reporting is foundational.

Narrative

If the narrative portion of Exhibit B is maintained, we offer several suggestions to improve this section. First, question 1(e) – which asks companies to discuss the divisional, operational, and cross functional responsibility for governance, consistency, and alignment – should be reworded to clarify the intent of the question.

Question 2 asks companies to report on AI systems that generate a consumer impact or financial transaction "indirectly." We recommend removing the term "indirectly," as these questions will generate responses that are overly broad and would likely provide little useful information to regulators. These questions should ask only for direct impacts, at least for an initial pilot phase of this proposal. Indirect impact is very subjective, may be unknowable, and could result in an undue burden on carriers who adopt an overly broad interpretation of "indirect."

Additionally, we would also appreciate clarification on the intent of question 3. Is this question asking whether, when developing, testing, and implementing AI Systems, companies deviate from established development, testing, and implementation practices for IT systems and data handling?

Question 4 requests discussion of the transparency and testing procedures performed. However, transparency and testing procedures are often model-specific, which makes it difficult for companies to provide an aggregate description for these procedures. Additionally, both questions 4 and 5 ask companies to provide their policy on various aspects of AI oversight, but we note that requesting a formal policy document may be premature for many carriers who are at the beginning of their implementation process.

Further, we recommend removing question 6 – which asks about open-source AI – because its relevance for regulatory policy is unclear. Finally, question 7 should also be removed because it is duplicative of other questions.

Checklist

Question 3(b), which asks about AI systems being used "ethically," should be removed because the term is subjective, and articulating how the applicable AI system was developed and is being used ethically will be difficult or impossible. By complying with the AI Model Bulletin and all other applicable laws and regulations, insurers will be functioning within those parameters and will thus be ethical.

Question 3(c) should be deleted or combined with question 3(a), as both questions are asking about ensuring compliance of AI systems. Additionally, question 3(d), which asks about "residual adverse consumer outcomes," should be clarified. And we believe question 3(g) should be deleted because it is unclear why companies must monitor and measure, and thus report, on the benefits of AI systems; if these systems are not beneficial, companies have no reason to keep them in production.

With regard to question 5, we would appreciate clarification regarding whether the uses of "data" and "internal data" are interchangeable. Further, the focus of question 5(b) seems to be on the representativeness of data, so we recommend that the phrase "and/or AI systems" be deleted from this question. Finally, questions 5(c) and 5(g) should be reworded to clarify their intent.

Exhibit C

This exhibit focuses on "high-risk" AI system, but "high risk" is not clearly defined. As mentioned above, we believe the proposal, as a whole, should be limited to AI systems that are high-risk. To determine which systems are considered "high risk," we would appreciate further consultation among regulators and industry on defining a common taxonomy. In the meantime, we believe relying on an insurer's own internally used definition of high-risk AI systems or high-risk models should be an acceptable option as a starting point.

Additionally, we are concerned that systems making "automated decision" are listed as the example of a "high-risk" model. Many systems that provide automated decisions are, in fact, low risk.

Exhibit D

The requirement to list all training data elements is overly burdensome. Insurers may train models on thousands of variables and only include a few in the final model. This exhibit should be limited to the variables included in the final model. Additionally, to the extent that companies are using third-party models, they may not have access to all training data elements since vendors do not typically disclose training data to their customers. So long as appropriate testing is performed, we do not believe companies should be required to list all training data elements, and we believe requiring companies to present this information would be inconsistent with the risk-based approach in the Model Bulletin.

Definitions

APCIA is concerned with the circular definition of "consumer impact" currently included in the proposal. We recommend defining the term "consumer" to include only individuals/natural persons, as opposed to entities. Similar to existing state privacy or consumer protection laws, the definition of "consumer" in the proposal should be defined as an individual who is a resident of the state. "Consumer" should not include an individual acting in a commercial or employment context or as an employee, owner, director, officer, or contractor of a company, partnership, sole proprietorship, nonprofit, or government agency.

We are also concerned that the term "bias" appears in multiple places without definition. Instead of using "bias," we recommend substituting "unfair discrimination" and/or "unfair trade practices" as appropriate.

Further, we appreciate that the definition of "generalized linear models" specifies they are not considered to be machine learning models for this evaluation. It would be helpful to further clarify that GLMs are similarly not considered to be AI models.

* * *

We appreciate your consideration of these views and welcome continued collaboration to ensure the proposal accommodates the dynamic nature of AI technologies while supporting innovation and competition.

Sincerely,

Matthew Vece
Sr. Director, Financial & Tax Counsel

Dave Snyder
Vice President, International & Counsel

BlueCross BlueShield Association
750 9th Street, NW
Washington, DC 20001
www.bcbs.com

September 5, 2025

Commissioner Michael Humphreys, Chair
Commissioner Mary Block, Co-Vice Chair
Commissioner Doug Ommen, Co-Vice Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
1101 K Street, NW, Suite 650
Washington, DC 20005

Submitted via email to Scott Sobel (ssobel@naic.org)

RE: Artificial Intelligence Systems Evaluation Tool

Dear Commissioners Humphreys, Block and Ommen:

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) Systems Evaluation Tool.

BCBSA is a national federation of 33 independent, community-based, and locally operated BCBS companies (Plans) that collectively cover, serve, and support 1 in 3 Americans in every ZIP code across all 50 states and Puerto Rico. BCBS Plans contract with 96% of hospitals and 95% of doctors across the country and serve those who are covered through Medicare, Medicaid, an employer, or purchase coverage on their own.

AI is transforming health care and responsible AI use has the potential to transform the efficiency of patient care, improve health outcomes, and lower health care costs. BCBS Plans are focused on deploying AI to improve outcomes, relieve administrative burdens, and offer data-driven insights so patients and their providers can make informed decisions on their health care. At the same time, BCBSA recognizes that AI must be deployed with oversight and accountability to reduce the risk of consumer harm. We share the Working Group's interest in finding pathways to drive responsible adoption of an AI Systems Evaluation Tool (AIS Tool). However, we have significant concerns with the Working Group's initial proposed draft AIS Tool.

We urge the NAIC not to move forward with the AIS Tool without taking additional time to consider stakeholder feedback and have more robust engagement. Stakeholder collaboration is critical to this process. Without the full perspectives of insurers, consumer advocates, technology experts, and state regulators, the NAIC risks developing a tool that is either misaligned with real-world practices or duplicative of existing laws and regulations. Given the speed at which AI technologies are evolving, there is also a significant risk of establishing an evaluation tool that quickly becomes outdated or fails to account for emerging use cases. Taking the time to incorporate stakeholders' perspectives through a more robust process will

help the NAIC identify the most material risks, ensure that reporting expectations are feasible, support innovation and safeguard consumers.

As to the subject AIS Tool itself, BCBSA has several foundational concerns as follows:

- **Alignment with NAIC's AI Model Bulletin and other existing guidance.** As drafted, the AIS Tool does not fully align with the NAIC Model Bulletin. This needs to be addressed to allow the tool to function as an extension of existing regulatory expectations, rather than as a parallel or conflicting framework. This consistency would enhance usability for state regulators, provide clarity for insurers, and ultimately strengthen regulatory oversight. By also ensuring the AIS Tool compliments existing and widely supported standards, such as the NIST AI Risk Management Framework—which provides a solid foundation for structured, risk-sensitive AI evaluation—regulators and regulated entities can avoid divergent requirements and thereby promote consistency.
  In addition, NAIC should consider how this tool can be aligned with existing NAIC resources, including the Financial and Market Conduct Handbook guidance. Without this assessment, NAIC runs the risk of creating duplicative or conflicting reporting, reducing the value of the collective oversight tools.

- **Scope.** The AIS Tool requests extensive details on all AI systems, every use case, and all AI initiatives. The scope is so broad that insurers would be compelled to generate large volumes of technical and operational material, much of which may have little bearing on consumer protection or financial oversight. Compiling and maintaining this information would be extremely burdensome to the point of having a negative impact on innovations that can meaningfully improve consumer experiences and reduce overall administrative burdens. Moreover, beyond the burdens, there is a question of the feasibility of this level of reporting. AI usage evolves rapidly and information may become obsolete in a short period of time. This approach risks overwhelming regulators with data difficult to act upon, ultimately obscuring rather than clarifying risk.

  These risks are amplified if there is ambiguity about the frequency of reporting and/or the organizational reporting level (*e.g.*, by business segment, by parent company, or subsidiary). It is further complicated by not appropriately differentiating between predictive AI and AI systems or by level of risk. To create an effective evaluation tool, we recommend that the NAIC, with stakeholder input, examine whether the scope as drafted is both reasonable and meaningful.

- **Intent.** Given the scope of the documentation being requested, it is not clear what risks the NAIC is most interested in managing against. While we appreciate and share the NAIC's concerns about responsible adoption and the risks that these technologies can pose for consumers, the possible use cases are incredibly varied as are the associated risks.

  In developing an evaluation tool, we urge the NAIC to first assess – with additional time to consider stakeholder input – what risks are of greatest concern so it can clearly and directly focus the tool on those priorities. This is needed to create consistent expectations between regulators and insurers, minimize duplication with existing laws and regulations, and prevent overregulation of this rapidly evolving technology. A risk-focused approach is essential to ensure the AIS Tool provides meaningful insights and actionable value.

- **Use Case.** Based on the proposed voluntary adoption structure of the AIS Tool, the following are concerns that need to be addressed before a tool can be finalized to ensure there is effective coordination and consistent application of the tool among states:
    - Since the AIS Tool is not being proposed as an addition to an existing NAIC Market Regulation or Financial handbook, it is not immediately clear whether it would be used for a market conduct or a financial examination. It is essential to identify which regulatory function is primarily responsible for the AIS Tool for effective coordination. This should be done at the front-end of the tool's development to support consistency with other tools supported by the designated handbook.
    - The AIS Tool should not be utilized as a regular reporting requirement and instead should be an optional tool that examiners may elect to utilize as part of a routine or targeted examination. There is a significant amount of reporting in the tool as currently drafted. Insurers should only be expected to report using this tool if the regulators have identified a need for the information and intend to meaningfully use it to promote appropriate consumer protections.
    - The AIS Tool should be administered by the insurer's lead or domestic state to foster coordination and collaboration among states and avoid duplicate requests by states which would place additional burdens on insurers. An enterprise response should also be allowed to avoid insurers having to provide responses for each legal entity. Without articulating these parameters in the development tool, there is a substantial risk of the reporting burden increasing exponentially with marginal increased benefit and transparency to regulators.
    - Clarity is needed on how examinations using the AIS Tool would be conducted. Without a consistent examination approach, there is significant risk of over-examination and variation in regulatory findings.
- **Confidentiality**. Not all use cases would provide the same level of confidentiality of the proprietary information embedded in the AIS Tool as what is inherent in a market conduct or financial examination. This would pose significant risk that sensitive and proprietary information could be made public through these evaluations. Inclusion in an NAIC handbook affords the necessary confidentiality protections and should be identified by NAIC as the appropriate pathway.

Thank you for the opportunity to provide feedback on the development of an AI System Evaluation Tool. Based on our member Plans' work in this area, we have additional detailed and specific feedback on the tool that we look forward to sharing with the NAIC in the future as the process and implementation is further refined. In turn, as the NAIC continues to evaluate the appropriate next steps for the AIS Tool, we encourage ongoing engagement with a broad range of stakeholders.

Continued collaboration will be essential to developing a thoughtful and balanced evaluation tool that not only protects consumers but also promotes innovation. Without industry engagement, the AI Tool risks duplicating existing oversight obligations or introducing inconsistencies that make compliance fragmented and inefficient. Such duplication can dilute, rather than enhance,

the effectiveness of regulatory oversight. If you have any questions, please contact Randi Chapman at Randi.Chapman@bcbsa.com.

Sincerely,

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

EVERSHEDS SUTHERLAND

THE Committee OF Annuity Insurers
www.annuity-insurers.org

September 5, 2025

Chair Michael Humphreys (PA)
Co-Vice Chair Doug Ommen (IA)
Co-Vice Chair Mary Block (VT)
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: AI Systems Evaluations Tool

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

The Committee of Annuity Insurers (CAI or Committee)[1] appreciates the opportunity to submit comments to the NAIC Big Data and Artificial Intelligence (H) Working Group (BDAI WG) on the draft AI Systems Evaluations – Optional Supplemental Exhibits for State Regulations ("AI Evaluation Tool").

The Committee recognizes the value of a tool to assist regulators when reviewing how insurers use AI systems. However, the Committee believes that the current draft AI Evaluation Tool is overbroad and should be pared down and refined, as discussed below.

INTRODUCTION/BACKGROUND

On July 7, 2025, the BDAI WG exposed for public comment its draft AI Evaluation Tool that is intended to enable regulators to "identify and assess AI systems' related risks on an ongoing basis." Those risks include financial and consumer risks because "the use of AI systems may lead to unintended consumer harms or compromise of the financial soundness of an insurance company."

The draft AI Evaluation Tool consists of four exhibits that request extensive information from insurers. Exhibit A is a chart intended to quantify and describe a company's use of AI models in 18 areas, including areas that are not core to the business of insurance, such as HR and legal and compliance. Exhibit B is intended to elicit information on a company's governance of AI risks, tracking many of the aspects of the NAIC Model AI Bulletin,[2] but expanding its scope to include uses of AI systems for financial transactions and financial reporting, and open-source AI systems, which are not mentioned in the Model AI Bulletin. Exhibit B also asks questions on how the company assesses and addresses risks in 18 areas, including testing for bias and addressing AI risks in ORSA and ERM functions. Exhibit C asks detailed questions about testing, uses and outcomes of "high-risk AI models," which are defined as AI system models that could cause adverse consumer, financial or financial reporting impacts. Exhibit

[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 state regulatory, securities and tax issues affecting annuities. The CAI's current 32 member companies represent approximately 80% of the annuity business in the United States. More information is available at https://www.annuity-insurers.org/.

[2] The Model Bulletin on the Use of Artificial Intelligence (AI) Systems by Insurers (Model AI Bulletin), NAIC, December 2023.

D focuses on identifying the kinds of data and sources for the data used in both internal and external AI systems, listing 24 categories, such as occupation, on-line social media and voice analytics.

Prior to the proposed AI Evaluation Tool, the NAIC issued its groundbreaking Model AI Bulletin that sets forth regulatory expectations on the use of AI systems (AIS or AI Systems) that play a role in making decisions impacting consumers. The Model AI Bulletin states that insurers should mitigate the risk that the use of AI Systems will result in adverse consumer decisions that are arbitrary or capricious, unfairly discriminatory, or otherwise violate unfair trade practices or other relevant laws. The controls and processes adopted should be reflective of the insurer's own assessment of the degree and nature of the risk posed to consumers by the AIS. The Model AI Bulletin also advises insurers about the types of information and documentation that an insurance regulator may request during an investigation or examination regarding an insurer's use of AI Systems, with a focus on documentation evidencing the insurer's AIS Program and use protocols for consumer-facing transactions.

Below are the Committee's comments based on our review of the exhibits making up the AI Evaluation Tool, as well as the BDAI WG's framing discussion regarding the purposes that the AI Evaluation Tool will serve, and how regulators could implement this resource in their oversight of the insurance industry.

COMMENTS

1. **The scope of the proposed AI Evaluation Tool should be aligned with the Model AI Bulletin.**

The NAIC Model AI Bulletin is a flexible, principles-based framework for regulating AI Systems that sets forth its scope in its opening sentence: "This bulletin is issued to remind all insurers… that **decisions or actions impacting consumers** that are made or supported by advanced analytical and computational technologies, including Artificial Intelligence Systems, must comply with all applicable insurance laws and regulations." (emphasis added). Also, in Section 3 titled "Regulatory Guidance and Expectations," the Model AI Bulletin states that "[it] is important that insurers adopt and implement controls specifically related to their use of AI that are designed to mitigate the **risk of Adverse Consumer Outcomes**." (emphasis added). The Model AI Bulletin outlines the legislative authorities that form the basis for the Model AI Bulletin, all of which focus either on consumer outcomes or corporate governance. The Model AI Bulletin does not express regulatory expectations with regard to the use of AI Systems in financial transactions and financial reporting or the use of open-source AI.

The proposed AI Evaluation Tool, by contrast, is intended to assist regulators in probing and evaluating governance and risk management measures that insurers have in place when using AI systems throughout the entire organization, not only when using AI in consumer transactions but also in financial transactions and financial reporting and when using open-source AI Systems. This represents a dramatic expansion in the scope of regulatory expectations for the evaluation of AI Systems during examinations without a clear articulation of the statutory basis for this expanded scope.

Today, insurers' understanding of the scope of regulatory expectations for AI Systems is the AI Systems Program Guidelines spelled out in the Model AI Bulletin. For instance, Section 1.4 of the Model AI Bulletin states that "[t]he AIS Program should be tailored to and proportionate with the insurer's use and reliance on AI and AI Systems. Controls and procedures **should be focused on the mitigation of Adverse Consumer Outcomes…** and the scope of controls and procedures applicable to a given AI System use case should reflect and align with the **Degree of Potential Harm to Consumers** with respect to that use case." (emphasis added). By contrast, Exhibit B of the AI Evaluation Tool would require companies to explain the AI System(s) used to generate a financial transaction or to generate information reported in financial statements within the context of their governance framework. The AI Evaluation Tool also requests information on the use of open-source AI Systems, their uses and prohibitions on use.

The Committee requests that the AI Evaluation Tool be revised to align with the current Model AI Bulletin and its focus on AI Systems that could have adverse consumer outcomes. The broad

expansion of regulatory focus reflected in the draft AI Evaluation Tool is premature. Insurers should be provided with guidelines and an understanding of the legal foundations for the Tool's scope and of regulatory expectations. It is only fair that insurers have such understanding before regulators use the AI Evaluation Tool to assess other uses of AI.

2. **The AI Evaluation Tool should be revised to focus on AI Systems used in core insurance functions that are or could impact significant decisions about consumers.**

The AI Evaluation Tool has a very broad reach, requesting information and data about the AI Systems used in 15 project areas of an insurer's operations. That broad reach goes far beyond those core insurance areas where an AI System could be used to make or contribute to making significant decisions about consumers, which would include areas such as underwriting, pricing, claims processing and fraud detection. The expansive reach of the AI Evaluation Tool to include areas such as legal/compliance, producer services, HR and product performance that are not core insurance functions that could adversely impact consumers will be costly for insurers to respond to and could potentially open up new areas for evaluation and examination that may not have a clear basis in law.

More specifically, requesting information about the number of AI systems used in the 15 project areas and how many were implemented in the past 12 months or are planned to be implemented in the next 6 months would seem to have little inherent value relative to the burden these requests would place on insurers to obtain this information. Other requests in the AI Evaluation Tool are so broadly framed, such as questions that ask about AI uses that may be "indirect," that they could be used to assess programs that go far beyond core insurance functions.

In short, the AI Evaluation Tool as currently structured would be extremely burdensome. One large CAI member estimates that each exam contemplated by the AI Evaluation Tool would require the hiring of consultants to complete the extensive list of questions and would easily cost hundreds of thousands of dollars to finalize, given the scope of the responses and detail expected. Others have stated that preparations for a market conduct exam based on the AI Evaluation Tool would require them to put together a large team who would have to be pulled from their normal jobs for several weeks, if not longer – not a desirable outcome.

The Committee requests that the scope of the AI Evaluation tool be narrowed and tailored to focus on those AI Systems used in core insurance functions that could harm consumers.

3. **The AI Evaluation Tool should be integrated into market conduct exams, and possibly into financial exams at a later date.**

Committee members are confused about whether the BDAI Working Group intends to publish a standalone AI Evaluation Tool, and if so, whether it will be NAIC-endorsed. Or does the Working Group intend for the AI Evaluation Tool to become part of the NAIC's market conduct and/or financial examination handbook?

Related to this is the question of confidentiality. The information requested by the AI Evaluation Tool is extensive and, in most cases, highly proprietary. The Committee strongly urges that the Working Group assure that the information obtained by regulators when using the AI Evaluation Tool is protected as confidential to the maximum extent possible. This can best be accomplished if the AI Evaluation Tool is only used in the context of a formal examination.

As currently drafted, the AI Evaluation Tool combines the review of consumer and financial risks. It is unclear how the various questions would be allocated and used in the financial and market conduct examination and enforcement contexts. For instance, it is unclear whether similar data on AI Systems would be produced in a financial examination as would be produced in a market conduct examination. In order to bring the Tool under the umbrella of a market conduct or a financial examination, the Tool should be refined so that it is clear which questions would be used in each type of exam.

The Committee requests that the Working Group provide further information on how the AI Evaluation Tool will align with the market conduct and financial examinations processes and urges the

Working Group to ensure the confidentiality of the data produced by insurers when responding to questions in the AI Evaluation Tool.

4. The AI Evaluation Tool should be accompanied by a best practice guide to TEVV (testing, evaluation, verification and validation) practices to provide a baseline for the evaluation of insurers' practices.

The AI Evaluation Tool is meant to assist regulators during regulatory exams in assessing and evaluating whether insurers have adequate oversight mechanisms in place for their AI Systems and whether the mechanisms are functioning effectively. However, without objective standards that can be used to measure the performance, adequacy and effectiveness of an AI System and the company's oversight mechanisms for such AI Systems, a regulatory exercise can become an exercise in subjective judgment that varies by evaluator. Subjective judgments can be arbitrary and capricious unless there is objective guidance that is uniformly applied.

Efforts to set standards for AI performance[3] are underway at a global scale by organizations such as NIST (the National Institute for Standards and Technology at the Department of Commerce (DOC))[4] and by the ISO (International Organization for Standardization) and the IEC (International Electrotechnical Commission). NIST's global engagement plan, which lays out required areas for standards development for AI systems according to tiers of need and readiness, classifies TEVV activities (testing, evaluation, validation and verification) as "urgently needed and ready for standardization."[5] This would include standards related to data integrity, data quality assessment, change management, and model performance, among others. Similarly, actuarial and accounting boards have not yet settled on the standards or metrics for measuring the reliability and accuracy of AI models.

The AI Evaluation Tool asks insurers to articulate their TEVV activities and to be judged on them, even before NIST has issued TEVV standards for AI performance. Such evaluations will require specialized regulator expertise to process and evaluate the responses produced in response to the AI Evaluation Tool. Until objective agreed-upon standards for the testing, evaluation, validation and verification of the performance of AI models are published by recognized standard-setting organizations, examiners, regulators, developers and deployers of AI Systems will be hard pressed to speak definitively and in the same language about risk assessment, mitigation and management of the complex tech stacks of AI systems.

In the interim, the Committee respectfully submits that it would be desirable that the BDAI Working Group work collaboratively with insurers to develop and publish guidance on best-in-class TEVV methodologies that can be used to inform examiners, regulators and insurers of regulatory expectations.

In sum, we urge the BDAI WG to narrow the scope of the current AI Evaluation Tool and align it with the current Model AI Bulletin. We also urge that before AIS exams and the AI Evaluation Tool are expanded to include other areas, such as financial transactions and financial reporting, the BDAI Working Group and/or other appropriate NAIC bodies articulate regulatory expectations in the form of guidelines and provide the legal foundation for examinations of AI Systems used in such other areas.

Thank you for the opportunity to comment on the draft AI Evaluation Tool. We hope that you find these comments helpful. Please do not hesitate to contact us if you have any questions and/or would like to discuss our comments further.

---

[4] NIST's website includes a section on AI Standards and notes that technical standards for AI data, performance and governance are essential for promoting innovation and public trust. https://www.nist.gov/artificial-intelligence/ai-standards.

[5] A Plan for Global Engagement on AI Standards, NIST AI 100-5 at 2.

Sincerely,

For The Committee of Annuity Insurers

Eversheds Sutherland (US) LLP

Mary Jane Wilson-Bilik
Stephen E. Roth
Alexander F. L. Sand
Eversheds Sutherland (US) LLP

GREATER NEW YORK HOSPITAL ASSOCIATION
PRESIDENT, KENNETH E. RASKE • 555 WEST 57TH STREET, NEW YORK, NY 10019 • T (212) 246-7100 • F (212) 262-6350 • WWW.GNYHA.ORG

September
Five
2 0 2 5

Scott Sobel
Big Data and AI (H) Working Group
National Association of Insurance Commissioners

Re: AI Systems Evaluation Tool

Dear Mr. Sobel:

On behalf of the more than 220 voluntary and public hospitals and health systems that make up the acute care membership of Greater New York Hospital Association (GNYHA), thank you for the opportunity to comment on the National Association of Insurance Commissioners' (NAIC) Artificial Intelligence (AI) Systems Evaluation Tool. We appreciate the NAIC Big Data and AI (H) Working Group's (H Working Group) continued focus on insurer use of AI. As emphasized in our June response to the H Working Group's request for information (RFI) on considerations for a model law regulating the use of AI in the insurance industry, our hospital members are particularly concerned with health insurers' use of AI in the utilization review (UR) process. Our comments today focus on ensuring a Systems Evaluation Tool collects meaningful data on this issue—from understanding the scope of AI use in UR to tracking the consumer impacts.

As reported in our previous comments, insurers deny our member hospitals' claims at alarming rates. New York State commercial plans reported denying approximately 23% of all inpatient hospital claims in 2022 and 25% in 2023.[1] While we understand that payers will continue to adopt technology in pursuit of administrative efficiencies, AI must not exacerbate or fortify this already untenable denial dynamic.

We identified several core principles for permissible and prohibited use of AI in the claims review process in our June comments. We now encourage the NAIC to develop tools that facilitate robust and meaningful regulatory oversight of insurers to ensure they are using AI consistent with these core principles. Overall, our assessment is that the proposed Systems Evaluation Tool does largely align with this oversight goal. However, we propose that the NAIC enhance some questions and specifically focus them on AI in UR. Doing so will help ensure the tool serves its intended purpose, allowing regulators to determine whether additional analysis focused on consumer risk is needed. We discuss our recommendations below within the context of our previously proposed core principles.

[1] Based on New York State Insurance Law Section 345 Health Care Claims Reports.

Denials must not be automated. Recent media coverage and legislative reports flag concerns with the correlation between automation and increased denials.[2] It is imperative that each denial of a health care service be made by an appropriately qualified physician licensed in the same or similar specialty as the treating provider who has carefully reviewed the individual patient's unique medical history and circumstances, including the treating provider's recommendation.

We note that the proposed Exhibit B: (Checklist) AI Systems Governance Risk Assessment Framework includes a few questions related to the concept of there being a human-in-the-loop, with a human actively involved and participating in review or operations.[3] Specifically, the checklist includes questions: 1) verifying that a human-in-the-loop is consistently and meaningfully contributing to decision-making and 2) evaluating the effectiveness of using a human-in-the-loop. However, this language does not appear in the narrative version of the Framework (Exhibit B: (Narrative) AI Systems Governance Risk Assessment Framework). *We urge that the Framework explicitly incorporate human-in-the-loop considerations in both the checklist and narrative formats, particularly as they relate to UR*. Ensuring that this appears in both formats would support a consistent evaluation approach regardless of which is used. This addition will also be useful in monitoring implementation of the recent pledge by several health plans that all prior authorization denials based on medical necessity for clinical factors will be reviewed by a licensed and qualified clinician.[4] As previously indicated, we appreciate this commitment and urge that the standard not be limited to prior authorization reviews but should apply to all clinical reviews across the UR continuum.

Additionally, as discussed below, we strongly support that the proposed Systems Evaluation Tool will collect data on consumer impact. As an overall comment, we flag that *with respect to UR, it is important to collect data on how not just patients but also providers are impacted*. Collecting data on complaints by both patients and the providers delivering or proposing to deliver services would improve the tool's utility and ensure appropriate interpretation of its data collection elements.

Insurer use of AI must be transparent. *We strongly support the questions in Exhibit A: Quantify Regulated Entity's Use of AI Systems that are intended to obtain information pertaining to the number of AI models that are new, updated, retired, etc.* These questions will help facilitate risk assessment, while also quantifying the entity's use of AI systems. As we note prior, there should also be explicit questions regarding AI in UR.

Regarding UR, we also *strongly support the inclusion of questions and collection of data about consumer complaints*. However, our member hospitals report that they are often unaware when AI is being used in the claim review process. The proposed tool asks insurers to report the number of consumer complaints and to describe the processes for tracking these complaints in Exhibit A. But tracking is only helpful if stakeholders are aware that AI was used in a denial and avail themselves of complaint processes.

---

[2] US Senate Permanent Subcommittee on Investigations, "Refusal of Recovery: How Medicare Advantage Insurers Have Denied Patients Access to Post-Acute Care," Majority Staff Report, October 17, 2024. https://www.propublica.org/article/cigna-pxdx-medical-health-insurance-rejection-claims

[3] IBM. What Is Human In The Loop (HITL)? https://www.ibm.com/think/topics/human-in-the-loop

[4] Health Plan Commitment to Improve Prior Authorization, June 2025, AHIP.org

We thus urge that Exhibit A of the *proposed Systems Evaluation Tool include a question asking insurers to discuss their policies for consumer disclosure and notice of AI use specifically in UR*. As noted in our RFI response, insurers should be required to disclose to regulators, insureds, and providers how they use AI in the claims review process. This should include notice when a new AI process is adopted and annually thereafter. Insurer websites should also clearly state when AI is used for UR. Additionally, to foster transparency, all denial letters should include a statement explaining if and how AI was used in the review process and where the insured and provider can find more information about the specific AI employed.

Further, to facilitate an understanding of the applicable review standards and expected outcomes, *detailed information about the algorithm, including its criteria and training data sets, must be available to providers upon request*. We note that making the content of the Systems Evaluation Tools available to stakeholders would support such transparency. For example, access to the information in Exhibits C: AI Systems High-Risk Model Details and Exhibit D: AI Systems Model Data Details would assist providers in evaluating model specifications and understanding the data sets used. Training data is foundational to the performance of AI models. The quality and diversity of data sources influence model reliability and may lead to bias. Biased training data can compromise the integrity of the model's output, which is particularly harmful in UR if providers are unaware of the issues. If providers or other consumers are denied access to training data information, it impairs their ability to critically assess the validity of model outputs and ultimately would undermine transparency.

Exhibit C also raises the *need to define risks and risk-criteria*. Exhibit C states that it is intended to collect detailed information on high-risk AI models that support automated decision-making and may result in adverse consumer, financial, or financial reporting outcomes. However, the tool does not define high-risk AI systems. Instead, each insurer is responsible for making that determination. Without clear criteria for defining high-risk AI, insurers may hesitate to classify systems as high-risk or fail to recognize certain systems as such. This could weaken oversight and limit the level of detail collected in situations where additional scrutiny is warranted. Further, allowing each insurer to set its own definitions introduces the potential for inconsistent interpretations of high and low risk and undermines transparency goals. Responsibility for defining AI system risk criteria should not rest solely with insurers. That responsibility should be assigned to regulators or another independent body to ensure consistent classification and effective oversight. Providing clear examples of high-risk models may also help define risk categories and criteria.

Third-party vendors must be held to the same standards. The NAIC queried in the RFI whether third-party vendors should be included in an AI model law. We strongly advocated for applying the model law to both insurers and any vendors with which they contract and/or to which they delegate UR functions. These entities may include, but are not limited to, UR agents, behavioral health management companies, and pharmacy benefit managers. The core principles for permissible AI use are no less salient because, for example, behavioral health services are carved out to a vendor rather than administered by the insurer directly. It is imperative that third-party vendors comply with all regulatory requirements governing the use of AI in UR. For these same reasons, *it is critical that any AI Systems Evaluation Tool unequivocally apply to and collect the same data regarding third-party vendors.*

Finally, we take this opportunity to reiterate that **AI processes must comply with all UR laws and requirements.** AI should be used to support compliance with UR guardrails and not as a tool to circumvent requirements. Regardless of whether AI is part of the UR process, UR must meet all procedural and substantive requirements. For example, the Centers for Medicare & Medicaid Services has explicitly cautioned Medicare Advantage (MA) plans that algorithms and AI must comply with all requirements applicable to the coverage determination process. MA regulations require medical necessity determinations to be based on an individual patient's circumstances, so algorithms that base determinations on larger data sets are therefore impermissible. Additionally, AI may not be used to shift coverage criteria over time—coverage criteria must be static and publicly accessible.[5]

State insurance law should hold all health plans to similar standards. Insurers should be required to submit AI algorithms and training data sets to the insurance regulators and certify that they do not rely on information or operate in ways that do not comply with UR laws, regulations, or guidance. Such submissions should occur prior to implementation, annually thereafter, and whenever algorithms or training sets are modified. Regulators must also have clear authority to audit AI used in UR, including outcomes.

Thank you for the opportunity to provide these comments. Please contact me at eleish@gnyha.org with any questions on GNYHA's recommendations.

Sincerely,

Emily Leish
Senior Vice President, Health Finance and Managed Care

[5] Frequently Asked Questions on Coverage Criteria and Utilization Management Requirements in CMS Final Rule (CMS-4201-F), February 6, 2024.

InsurTech COALITION

Dear Big Data and AI Working Group Members:

The Insurtech Coalition ("IC") welcomes this opportunity to provide comments to the NAIC's Big Data and AI Working Group (BDAIWG) on the draft AI Systems Evaluation Tool. IC commends the BDAIWG for its proactive and thoughtful approach to the rapid and transformative adoption of artificial intelligence and machine learning (AI systems) within the insurance sector. IC shares the BDAIWG's goal of fostering a dynamic and innovative market while ensuring robust, effective, and proportional consumer protection. The development of a standardized framework for the evaluation of AI systems is a crucial step toward achieving regulatory clarity and consistency nationwide.

While the draft tool represents a valuable first step, IC has identified several structural and operational elements that could inadvertently create significant administrative and financial burdens for insurers and a less effective examination process for regulators. The current design of the tool, with its commingled scope and reliance on extensive open-ended questions, risks becoming a costly and inefficient exercise that does not align with its stated purpose as an optional, supplemental guide.

IC respectfully submits four primary recommendations to enhance the tool's effectiveness, reduce compliance and exam costs, and promote a more targeted and efficient regulatory approach:

1. **Set Standards**: Current market conduct and financial examinations are tied to specific, well articulated and objective criteria. In many instances in which the tool requires insurers to detail how their policies or procedures "ensure" an outcome or "manage" another. Before the tool could have real utility, standards for "ensuring" or "managing" are necessary.
2. **Bifurcation**: The single, multipurpose tool should be separated into two distinct, purpose-built exhibits: one for financial examinations and one for market conduct examinations.
3. **Streamlined Data Collection**: The numerous open-ended questions requiring narrative "discussions" and "explanations" should be replaced with structured, checklist-based prompts or specific requests for documentation, which would provide greater consistency and efficiency.
4. **Proportional Application**: The tool should be formally designated as an optional, supplemental exhibit to be included in the appropriate NAIC examination handbooks, clarifying its role as a guide for examiners rather than a mandatory, standalone data request. We believe it should be clear that it is for domestic states to use in their exams of their domestic companies.

The following sections provide a detailed analysis of the draft tool and the rationale supporting these recommendations, which IC believes will result in a more practical and effective framework

for both the insurance industry and the consuming public.

II. Background and Overall Perspective: Fostering Responsible Innovation

The Insurtech Coalition represents a broad cross-section of companies dedicated to leveraging technology and data to improve the insurance value chain. These companies are at the forefront of developing and implementing technology, which includes AI systems that can offer substantial benefits to both insurers and consumers. These benefits include the facilitation of innovative products, improved customer interfaces, and the enhancement of efficiency and accuracy throughout underwriting, ratemaking, and claims processes. IC's mission is to promote the responsible use of these technologies and advocate for a regulatory environment that supports innovation while effectively managing risks.

We do not believe that AI and machine learning systems are inherently more prone to consumer harm or financial risk than other technologies used by insurers. The insurance industry is already one of the most highly regulated sectors, with a robust framework designed to protect consumers and ensure financial solvency. As such, we believe that existing regulations which focus on outcomes or set definite criteria (e.g., investment laws) are sufficient to oversee the use of AI systems. However, we appreciate the work of the Big Data and AI Working Group in creating a resource to aid regulators. We believe this document would be effective as an exhibit to the financial and market conduct examiners' handbooks, providing guidance on how domestic states can evaluate AI systems to ensure compliance with the regulations they are already subject to.

III. Analysis of the Draft AI Systems Evaluation Tool: Identifying Structural and Operational Misalignments

While the goal of comprehensive risk assessment is commendable, the current design creates a functional disparity between the tool’s stated purpose and its practical application. The tool, in its current form, would impose a significant administrative and financial burden on insurers, especially smaller new carriers, and introduce inefficiencies into the regulatory examination process.

A. The Unwarranted Administrative and Financial Burden

The draft AI Systems Evaluation Tool is presented as an optional, supplemental exhibit, but its comprehensive and detailed nature suggests it would operate as a de facto mandatory data request. The nature of many questions indicate that the data requests have few limits and no guidelines for regulators or companies. The document requires "detailed information across multiple exhibits," which would be a significant undertaking for any insurer, but specifically for smaller companies without teams dedicated just to examinations. The time and resources required to compile the requested information would be substantial, leading to increased costs that could ultimately be passed on to consumers or disincentivize the responsible adoption of new technologies.

The exhibits, as drafted, require a level of detail that may impose significant burden without

commensurate regulatory benefit. Both the narrative and checklist formats call for "comprehensive information." The narrative version requests broad qualitative discussions, while the checklist, though more structured, still includes detailed criteria and numerous open-ended questions. While the underlying focus on AI governance is appropriate, the current drafting risks producing expansive, inconsistent responses that could be difficult for examiners to analyze efficiently.

B. The Compounding Inefficiency of Commingled Financial and Market Conduct Scrutiny

The current tool's structure conflates financial and market conduct risks, a design choice that would create significant inefficiency and jurisdictional confusion. An analysis of the exhibits shows that they simultaneously assess a company's financial stability and its compliance with consumer protection laws.

- **Exhibit A** tracks both the number of AI models with a "Material Financial Impact" and the number of models with "Consumer Impact," as well as consumer complaints.
- **Exhibit B** similarly combines these two areas. The narrative version asks about the integration of AI into a company's "Own Risk and Solvency Assessment (ORSA) and Enterprise Risk Management (ERM) assessments" (financial risk) while also inquiring about the use of AI systems that generate "consumer impact directly or indirectly" (market conduct risk). The checklist version follows this pattern, with questions addressing ERM and ORSA (items 3h, 3i) alongside questions on adverse consumer outcomes and disclosures (items 3d, 3p).
- **Exhibits C and D** also combine these risk categories. Exhibit C focuses on models that could cause "adverse consumer, financial, or financial reporting impact" , and Exhibit D's purpose is to identify risk of "consumer adverse impact, unfair trade practices, financial, or financial reporting impact".

This commingling of financial and market conduct questions within a single tool is problematic. It would force a financial examiner, whose primary expertise lies in solvency and financial stability, to analyze data on consumer complaints and unfair trade practices, which are the purview of a market conduct examiner. Conversely, a market conduct examiner would be compelled to review a company's ORSA and ERM processes. This not only burdens the insurer by forcing a single, multi-purpose response that must satisfy two different regulatory perspectives, but also strains regulatory resources by requiring examiners to work outside of their specialized areas of expertise. A bifurcated approach would allow for more focused, efficient, and relevant examinations, benefiting both the regulator and the regulated entity by directing the right questions to the right experts. This structural issue provides the core justification for the recommendation to separate the tool.

IV. Specific Recommendations for Improvement: A Path Toward a More Effective Tool

The following recommendations are designed to address the structural and operational misalignments identified in the analysis above, providing a path toward a more practical, efficient, and effective regulatory framework.

A. Recommendation: Separate Tools for Financial and Market Conduct Exams

IC strongly recommends that the draft tool be separated into two distinct and purpose-built exhibits: an "AI Financial Examination Exhibit" and an "AI Market Conduct Examination Exhibit." As demonstrated, the current commingled structure creates inefficiencies and jurisdictional confusion. A bifurcated approach would allow each regulatory body to conduct a more targeted and effective review, reducing the burden on insurers to prepare a single document for two different, and often unrelated, purposes. .This proposed structure ensures that each regulator receives only the information relevant to their area of expertise, allowing for a more focused and effective review. The separation also reduces the burden on insurers by allowing them to respond to a more targeted set of questions for each specific examination type.

B. Recommendation: Prioritize Checklists and Specific Prompts Over Open-Ended Narrative

The draft tool's reliance on numerous open-ended "Discuss" and "Explain" questions, particularly in the Exhibit B narrative, is a significant point of concern. These prompts, which cover broad topics from governance frameworks to vendor oversight, are inefficient and introduce unnecessary ambiguity.

The primary issues with these open-ended inquiries are twofold:

- **Subjectivity:** They lead to inconsistent and highly subjective responses from different insurers. One company's "discussion" of its governance structure may be a high-level overview, while another may provide a detailed, multi-page treatise. This lack of standardization undermines the tool's purpose as a consistent evaluation instrument, making it difficult for regulators to compare practices across the industry.
- **Time and Cost:** Drafting comprehensive, well-reasoned narrative responses is an extremely time-consuming and expensive process for an insurer. It requires significant cross-functional input from legal, compliance, actuarial, IT, and business units. This directly contributes to the high "costs of exams" that are a primary concern for the Coaltion.

The reliance on ambiguous, open-ended questions leads to a broader, more significant issue: a lack of regulatory predictability. The causal link is clear: open-ended questions lead to inconsistent responses and subjective examiner interpretations, which in turn create regulatory unpredictability. This increased risk aversion can ultimately dampen innovation, hindering the very benefits that AI systems promise.

To mitigate these risks, IC recommends replacing the open-ended questions with more structured, binary (Yes/No), or quantitative prompts. A checklist-based format that directs the insurer to reference specific sections of their internal governance policies would be far more effective. For example, instead of asking, "Discuss the governance structure," the tool could ask, "Does the company have a documented AI Governance Policy? If yes, please provide the

section number or page reference for the policy." This approach provides regulators with the necessary information while standardizing the response and reducing the administrative burden on insurers.

C. Recommendation: Clarify the Tool's Purpose as an Exhibit to the Handbook

The draft tool's most significant strength is its stated purpose as an "Optional Supplemental Exhibit." The document notes that it is intended to "supplement existing... review procedures". IC strongly supports this concept but argues that the tool's demanding nature is incongruous with this intent. To ensure its purpose is clearly understood and its application is proportional, the tool should be officially designated as a supplemental exhibit to the NAIC's Financial Condition Examiners Handbook or the Market Conduct Examiner Handbook.

This formal designation would clarify the tool's role and reinforce its position as a guide to aid examiners in their review procedures, rather than a standalone, mandatory data call. This approach would also serve to acknowledge the reality that regulators may not have deep expertise in AI governance, model validation, and data science, as noted in the user query. The current draft, with its reliance on technical terms and open-ended questions, places an undue burden on examiners who may lack this specialized skill set. A simpler, structured, and handbook-integrated exhibit would be more practical and effective, providing a standardized framework that is accessible to a wider range of examiners. By providing a clear, simplified, and well-structured guide, the tool would empower examiners to conduct more focused and effective reviews, thereby benefiting both regulators and the insurance industry.

V. Summary of Recommendations and Concluding Remarks

The Insurtech Coalition reiterates its support for the BDAIWG's efforts to create a robust and practical framework for evaluating AI systems in the insurance sector. The draft tool is a crucial step forward, and IC believes the following adjustments will transform it into an even more effective instrument:

- **Bifurcate the tool** into distinct financial and market conduct exhibits to promote targeted, efficient, and specialized examinations.
- **Replace open-ended questions** with structured, checklist-based prompts to ensure consistent responses, reduce administrative burden, and increase regulatory predictability.
- **Formally integrate the tool** as a supplemental exhibit within the NAIC's examination handbooks to clarify its purpose and ensure proportional application.

IC is committed to continued partnership with the BDAIWG and other regulatory bodies to refine this tool and future frameworks. The recommendations presented herein are intended to foster a regulatory environment that supports responsible innovation and efficiency, ultimately resulting in a more dynamic insurance market that better serves the consuming public. IC looks forward to a collaborative effort to develop a final tool that is both practical for the industry and effective for

regulators.

Sincerely,

The Insurtech Coaltion

INSURED RETIREMENT INSTITUTE

Electronically Submitted to ssobel@naic.org

September 5, 2025

To: NAIC Big Data and AI Working Group (the "Working Group")

Re: AI Systems Evaluation Tool

On behalf of our members, the Insured Retirement Institute (IRI)[1] appreciates the opportunity to provide input to the Working Group on the Artificial Intelligence Systems Evaluations – Optional Supplemental Exhibits for State Regulations ("Draft Tool"). Overall, our members do not necessarily oppose a tool or other guidelines that provide structure to a state regulator's evaluation or examination of an insurer's use of AI. Any new tool, however, should be 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 ("Model Bulletin"). Our members also have several open questions/concerns about the draft tool and its exhibits, and we would urge the Working Group to continue to work with stakeholders to ensure that any tool or guidelines provides appropriate clarity for regulators and insurers. Our general feedback, along with more specific comments/questions, are as follows:

The current Draft Tool is overbroad and goes beyond a focus on consumer harm and the requirements outlined in the Model Bulletin. A tool used for examinations should appropriately reflect the existing laws, regulations, and guidance.

As we mentioned in our letter dated June 30, 2025, in response to the AI Request for Information, a tool should be flexible enough to remain in place as the industry and technology evolve over time. We recommend that the Draft Tool better align with and reflect the guidelines as outlined in the Model Bulletin, so that it can be effectively utilized by those states that have adopted the Model Bulletin thus far. Our members strongly support the continued adoption of the Model Bulletin, and we continue to believe that any tool should be reflective of the Model Bulletin's

[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.

principles, which are grounded in existing laws. This will ensure that any tool or guidance is appropriately focused on governance, risk, and supervision.

Any tool needs to provide sufficient clarity as to the purpose, goals, and usage of the tool so that companies have an expectation and understanding of what information may be gathered and reviewed during an exam or other evaluation. Unfortunately, this current draft lacks that clarity, has a much broader focus than just consumer harm, and goes beyond the guiding principles in the Model Bulletin.

We understand that the Working Group may be intending for states to pilot the Draft Tool in the coming months, but if the Draft Tool is also intended to serve as a checklist for insurers, we think there is more work that needs to be done. We also understand there may be a need for an immediate resource, but given the existing laws and regulations in place that govern unfair trade practices and unfair discrimination, it's unclear that there needs to be something so specific to AI, versus relying on the existing market conduct examination guidelines (or simply adding to and ensuring that those guidelines reflect the principles outlined in the Model Bulletin). It may also make sense at this time to narrowly tailor a tool to specific lines of business or to areas where there are specific concerns. Such a broad tool will likely encompass much more than a typical exam would entail and may end up not being that useful for regulators or for companies.

Having a standalone tool outside of the guidelines that are intended to provide regulatory guidance on market conduct and financial examinations creates confusion. For example, this Draft Tool combines reviews of both financial and consumer risk, and it is unclear when or how the various questions/exhibits would be utilized. Additionally, on page 1, the "Intent" section references use of the Draft Tool on an "on-going basis." This seems to imply that this Draft Tool would be used more for reporting, or for other uses outside of exams. As such, we strongly recommend that the Working Group simply rely on or update existing guidelines to reflect the Model Bulletin.

It is important to ensure that any tool being used for examinations and enforcement action reflects and ties back to existing legal requirements. This ensures that any tool or guidance is used consistently by regulators, and companies understand how it will be utilized. Because the Draft Tool is overbroad and sweeping in a way that goes above and beyond existing requirements, it fails to provide insurers with sufficient certainty as to how and on what they may be examined.

To address the above-mentioned concerns and to better align with the Model Bulletin, the Draft Tool should be more streamlined and consolidated and utilize clear terminology and definitions.

A tool better aligned with the existing requirements of the Model Bulletin will ensure that the tool is not duplicative and focuses on relevant and necessary information for examination

purposes. The Draft Tool currently asks for the same or similar data several times in different ways, and it requests information well beyond what would be expected for an examination (i.e., the number of AI system licenses). The Draft Tool should focus specifically on the AI systems and use cases currently being used and the risks, governance, and supervision associated with those systems and use cases, while also recognizing that an insurer's governance framework should be commensurate with the type of use cases.

Additionally, terminology should be consistent with the Model Bulletin. It is important that language in the tool be tailored in such a way to ensure companies understand the regulatory expectations. This includes the use of certain undefined terms (or those that may be used interchangeably with defined terms) and terms that may require additional clarity. Some examples include:

1) Clarification that "company" does not include non-insurance affiliates.
2) While "AI System" is defined, other terms like "AI models" and "AI systems models" are not defined and appear to be used interchangeably.
3) "AI Governance Framework" is not defined and does not align with the Model Bulletin, which uses the term "AIS Program."
4) "Material financial impact", "consumer impact", "open-source AI", "model risk classification", "adverse consumer, financial, or financial reporting impact", and "residual AI systems risk" all need to be defined or further clarified.

There are several specific comments on the exhibits where information could be better consolidated and streamlined, as well as a number of questions and comments about terminology. We have outlined these below for each Exhibit:

Exhibit A:

1) The list of "Program Areas" in the chart is overly broad and does not align with the Model Bulletin. The inclusion of internal operations such as Legal and HR go beyond the typical insurance regulatory examination, and it is unclear what information would be requested here.
2) Consistent with our comments above, the information gathered here should be geared towards the type of regulator review involved. Blurring the lines between the types of regulator exams creates confusion about what information should be provided and when.
3) It is unclear how companies would be expected to track consumer complaints related to AI. A complaint could result from a decision that may have involved AI in some way, but the use of AI may not have contributed to the specific complaint. An expectation within the tool to track this type of information, which would be a new requirement for companies, necessitates continued discussion and consideration among stakeholders before moving forward.

4) It could be difficult to gather the exact number of AI systems implemented in the last few months and any planning to be implemented as this could vary depending on what factors are being considered, and it is unclear how this information would be relevant in the context of an examination. Additionally, there is potential competitive and proprietary risk associated with disclosing future AI development initiatives. This kind of disclosure could inadvertently expose sensitive strategic information without adding any real benefit to an exam. The focus should be on governance and processes for AI systems, and less so on the timing of adoption or implementation. Any tool should align with the Model Bulletin and focus on the risk of those AI systems currently in use, and the governance processes and guardrails in place.
5) Further clarity and/or parameters are needed to help companies understand how to answer any question that contains "resulting from a process that relied on AI."

Exhibit B:

1) There is confusion about the two different versions of Exhibit B, one of which is a narrative, and the other is a checklist. There is not complete overlap between the two versions, which creates confusion about the expectations when different information is being requested depending on what version is being used. It is also unclear how the narrative questions are intended to be used by the regulators relative to the actual use cases.
2) Some of the questions in the narrative version require discussion of AI systems' uses that may be "indirect." The scope of this could become quite broad and goes beyond what is contemplated by the Model Bulletin.
3) Transparency and testing procedures can be specific to an AI system, which could make it difficult to aggregate a description to each line of business.
4) The checklist requires the insurer to provide details about how they monitor and measure the benefits of AI systems. While AI benefits are important in balancing risk, it's unclear how this information would be used in the context of an exam, and what the requirement would be for companies in regard to monitoring the benefits.
5) It is unclear why the number of AI licenses for a specific AI system is relevant information for an examination and is recommended that this be removed.

Exhibit C:

1) This Exhibit has several terms ("residual", "adverse consumer, financial, or financial reporting impact", "model risk classification" e.g.) that are unclear and may be defined in different ways by different companies. As such, our members have raised concerns about being able to effectively and accurately respond to some of these questions.
2) It is unclear as to how, or for what purpose, the regulators will use the information regarding "high risk" systems. There may be more value to focus on questions about how

such systems are reviewed, especially those involving a consumer complaint or an identified regulatory deficiency, versus asking companies to provide a comprehensive list of what systems they have deemed to be high risk. This will keep the focus on areas where there is consumer harm or other identified issues involving consumers.

Exhibit D:

1) Consistent with our comments above, a wide variety of uses (consumer impact, financial reporting) are contemplated in the same grid. We recommend that the information gathered appropriately correspond to the underlying regulatory obligations, and ultimately this would be better suited to the corresponding guidelines already in place.
2) Column 3 in this Exhibit is quite broad and should be limited to use of this data within an AI system.
3) Members have concerns about having to provide names of each vendor. Additionally, the level of detail that could be required here may be difficult to obtain from third-party vendors. It is recommended that this be limited to categories or types of vendors, or potentially just to internally developed models or systems where the insurer is the actual developer.

As you will see, there are several outstanding questions that we believe need to be worked through and addressed to ensure clarity and consistency when it comes to examinations. In conclusion, it is important that any Draft Tool align with the Model Bulletin, and we strongly recommend that the Working Group consider supplementing any existing examination guidelines, in order to maintain consistency and clarity for regulators and insurers when it comes to examinations. We encourage the Working Group to continue working with stakeholders and soliciting feedback on this important issue.

On behalf of IRI and our members, thank you again for the opportunity to provide these comments. We would be happy to discuss further with you and look forward to continued collaboration and partnership with the Working Group.

Sincerely,

Sarah E. Wood

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

From: Lederer, Julie
Sent: Tuesday, August 26, 2025 7:00 PM
To: Sobel, Scott <SSobel@naic.org>
Subject: RE: Extension for Public Exposure of AI Systems Evaluation Tool to 9/5

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

The Missouri Department of Commerce and Insurance offers the following comments on the AI Systems Evaluation Tool:

General comments:

1. The exhibits seem to assume that models can be bifurcated into "AI" and "non-AI," but the use isn't this clear-cut. For example, AI might be used to review qualitative claims handling entries, which could then be used in a large claim reserving model. Is that reserving model then considered an "AI model" for the purposes of these exhibits?

2. Some of the information requests (for example, on the sources of data and the governance framework) are pertinent to all types of models, not just ones that incorporate AI. Regulators using these exhibits may learn much more about an insurer's AI systems than its other tools. Could the working group instead consider a broader risk-focused framework that incorporates all types of models? This initiative would benefit from input from the actuarial task forces.

3. Has the working group asked the actuarial task forces for input?

4. These exhibits are extensive and would require significant time for insurers to complete.

Exhibit A:

5. It would be helpful to know how an insurer is using AI, but collecting information on the number of AI models does not seem risk-focused. It's not clear what regulators would do with count data.

Exhibit B (checklist):

6. The checklist format seems to presuppose that the NAIC has provided written guidance on what should be in an AI governance framework.

Definitions:

7. The definition of "AI System" should incorporate the idea that the system has some degree of independence and continues to learn after deployment. The definition as

written would apply to a variety of traditional models, such as generalized linear models, which is not the intention.

8. Some of the examples of GLMs in the definition of “Generalized Linear Models (GLMs)” are not GLMs. E.g., GAMs are extensions of GLMs but not GLMs, elastic net is a regularization technique that can be applied to GLMs, etc.

9. The working group may want to align the definition of “Third Party” with the definition under development by the Third-Party Data and Models (H) Working Group.

Julie Lederer, FCAS, MAAA
Property and Casualty Actuary
Missouri Department of Commerce & Insurance

MONITAUR

September 5, 2025

NAIC - Big Data and Artificial Intelligence (H) Working Group
Attn: Commissioner Humphreys(chair) and members of the working roup
RE: AI System Evaluation Tool Exposure Draft

Commissioner and members of the Big Data and Artificial Intelligence Working Group,

Monitaur supports the NAIC's continued efforts to find the balance of regulating Artificial Intelligence(AI) in a manner that prioritizes consumer protection while supporting important opportunities for innovation and betterment. Like you, we believe that AI, as defined in the NAIC Model Bulletin(Bulletin), can "*simplify and automate processes, and promote efficiency and accuracy*"[1] as well as "*...contribute to safe and stable insurance markets.*"[2]. We believe that thoughtful AI Governance not only mitigates risk and undesirable adverse outcomes on consumers, but also improves AI System(AIS) performance, reliability, and ensures it is 'fit for purpose'.

A tool that enables regulators to evaluate AI and an AIS Program is important, and we offer this feedback to help the NAIC and regulators achieve an approach that is more focused and effective than this first exposure draft.

Credentials
For background, my professional career began as an auditor and developed into an AI auditor after years building models and earning a Masters degree in Data Science. I've been the individual responsible for examining and auditing complex AI systems for concepts like fairness, compliance, transparency, robustness, accuracy, and security. I was one of the first contributors to ISACA's ML Assurance guidance and, as co-founder of Monitaur, my goal is to enable companies to use AI systems that businesses, regulators, and consumers can trust.

Our company has the great privilege of being the preferred AI Governance solution provider and partner to some of our country's leading insurance brands, data providers, and technologists. Since 2020, we have appreciated opportunities you have afforded our company to offer insights about AI risk and governance, and we hope to continue to be helpful and thoughtful.

We believe you will receive feedback from a wide variety of stakeholders with skills and expertise in a wide range of domains. The focus of this response is from my professional background as a machine learning auditor and data scientist, as well as from our company's focus on AI Governance for the insurance industry. Our feedback aims to help the working group and interested parties consider strategies for reasonable and

[1] NAIC MODEL BULLETIN: USE OF ARTIFICIAL INTELLIGENCE SYSTEMS BY INSURERS. *NAIC*, https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-ai-model-bulletin.pdf.pdf
[2] Ibid

effective methods of evaluating AI Systems and enforcing the Bulletin. Our feedback does not speak for any of our customers.

Comments

AI Governance, by the related name of "Model Risk Management", has a long history in financial services, with detailed requirements enacted in earnest in response to the 2010 Dodd-Frank Act; can trace its roots in the United States back further to 2000's OCC Bulletin 2000-16[3]. Given this corpus of knowledge and experience from over two decades of experience in operational risk management of models, which fundamentally, AI Systems are models, we first would like to advocate for at least referencing OCC's learnings for what battle-tested, AIS examinations could look like. Within the OCC's latest revision of the *"Comptroller's Handbook: Model Risk Management"*[4], there is an Examination Procedures section titled: *"Quality of Model Risk Management"*, which outlines a list of evaluation criteria, and conversely, questions to ask the regulatee. Although the depth and breadth of examinations performed by OCC may not be necessary, given the lesser commensurate aggregated AIS and model risk of an Insurer, the structure and clarity of purpose could be an inspiration for the NAIC's development of regulator guidance, such as the AI Systems Evaluation Tool.

The Bulletin provides a valuable roadmap, or guide, for what companies should aim to establish for their own AI Governance program and, in our opinion, the logical complementary guide for regulators would be something more akin to the OCC Examination Procedures section mentioned above. The reasons are as follows:

- Examinations would first evaluate the policy, procedure, and controls for explicit components expected by the Bulletin
- Examinations would then sample AI Systems and their documentation for implementation-level validation of effective adherence to the policy (which as mentioned in the previous step, has already been evaluated for alignment with the Bulletin)
- AI System examinations would focus on more material or consequential modeling systems(high-risk), which helps to ensure the scope and effort by both regulators and industry are across the most critical AI Systems

As with any examination or audit, problematic findings or issues with the above would then enable a deeper evaluation and collection of information, but an OCC-like examination pointed specifically at evaluating policies, procedures, and controls and then testing those on individual examples is a proven methodology to audit.

---

[3] OCC.Gov . "Risk Modeling: Model Validation" May 30, 2000
https://www.occ.gov/static/rescinded-bulletins/bulletin-2000-16.pdf

[4] OCC.Gov. "Comptroller's Handbook: Model Risk Management." August 18, 2021.
https://www.occ.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/index-model-risk-management.html.

Exhibit B

Many of the components of the Exhibit B checklist align with the above methodology; however, as high-level feedback, here are a couple of observations for consideration:

- Sections 1-3 in particular align well with evaluating the policy, but the reference components or questions could be much more literally and directly aligned with language from the bulletin. As an example: *does the insurance company policy include language that defines an internal AI System accountability structure?* This question would help to directly evaluate conformity with section 2.3 of the Bulletin.
- Sections 4-6 are less consistent in offering company-level policy evaluation versus starting to dive into AIS-level examination questions. As an example, it is not productive or overly valuable to get a generalized response to Ref 5.c "*The extent to which the data and/or AI systems are representative of the population the model is being applied to*"; however, this is exactly the sort of evaluation that might be performed at an AIS-level where the training data and target population should be clearly defined and verifiable.

There are also certain questions and the responses in the Narrative section of Exhibit B that are better in the checklist or that may not yield much meaningful insights to specifically evaluate how a company is confirming(or not) with the explicit language and expectations of the Bulletin. For example, the approach to the open-source question is not established in the Bulletin and it would not help in the evaluation of an AI Governance program or an AI System.

Exhibit C

If you were to preserve the general structure of this evaluation tool, we suggest eliminating the Narrative, focusing Exhibit B on policy evaluation, and focusing Exhibit C on an evaluation of sampled high-risk AI Systems for evidence of adherence to policy.

To illustrate an example of a potential process:

- An exam begins with a carrier enabling an examiner to verify that the content of the policy, procedures, and controls satisfy the guidance and requirements of the Bulletin. The review can quite simply leverage the bulletin as the checklist for language that speaks to required elements. Gaps can be identified and discussed for materiality or improved focus.
    - Does the policy and/or controls include expectations for monitoring a model for drift?
    - Does the policy provide guidance of how AIS are evaluated and tiered for risk?
- An exam is supported by a generalized summary of the inventory and how that inventory is segmented across key lines of business or risk-levels.
    - Does the inventory clearly demonstrate a segmentation of risk-levels and is there evidence of the application of policy guidance for risk tiering?
- The exam continues by gathering carrier responses and/or some form of documentation package that enables an examiner to validate the policy is being performed across multiple high-prisk AI Systems
    - Is there evidence of the methods and frequency by which drift monitoring was performed?

Exhibit D
We will defer to others to offer broader reflections about Exhibit D, but as a single point of feedback, if the objective of this Evaluation Tool is for the sole purpose of enforcing the Bulletin and helping conduct an AIS examination, Exhibit D does not appear particularly useful or objective-aligned. While we understand and have been a part of many conversations across the industry about the relationship between AI and new data available and used across insurance, and while the Bulletin did really well to focus on the AIS or model, this section drifts away from that singular focus.

Definition of AI System Models–Suggestion to align with AIS definition
One final comment regarding terminology - this document introduces “AI System model(s)”, “AI System Use Cases”, and sometimes standalone “model” as a question or comment reference point. We strongly suggest doing everything in your power to avoid navigating the nuance of these and other terms - the AIS definition is purposeful and we are comfortable that industry generally has solved how to identify a unique AIS for the purpose of inventory and Bulletin-conforming governance.

In our experience, a singular focus on mastering the fundamentals and starting simply, often simpler than feels ‘correct’, is a critical first step, and a common stumbling block. We recommend the NAIC focuses its efforts singularly on ensuring that the main building blocks are in place, as thoughtfully outlined in the NAIC AI Model Bulletin; and can be referenced in-depth in the OCC Comptroller’s handbook. To quote Albert Einstein: “Make everything as simple as possible, but not simpler.”

In conclusion, we believe that the NAIC AI Model Bulletin is a well-produced, guiding document, and definitions, and approaches of subsequent documents, such as the AI Systems Evaluation Tool, should be directly traced to its sections. Focusing this initial Evaluation document on ensuring the fundamentals of a comprehensive AIS program have been enacted, leaving more specific details, as needed for commensurate risk evaluation, for consideration once programs are more mature.

As always, we are at your disposal for ways we can be of assistance in the NAIC’s efforts.

Sincerely,
Andrew Clark, PhD
CTO and co-founder
Monitaur

Date: September 5, 2025

From: NAIC Consumer Representatives

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

CC: Miguel Romero and Scott Sobel

Re: Comments to AI Systems Evaluation Tool

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 Systems Evaluation Tool are critical to assisting regulators to operationalize the broad principles of consumer protection and regulatory oversight over AI articulated in the AI Model Bulletin.

Please see our below comments on the AI Systems Evaluation Tool.

Instructions

We appreciate the detail included in the instructions section on the purpose of each exhibit. While we are in favor of including consumer complaints as a key way to identify and assess risk of use of AI models, we are confused about the inclusion of identifying reputational risk simultaneously. If reputational risk is meant to capture risk to a consumer, then that should be specified and we would recommend using different language. If reputational risk was meant to refer to potential risk to the regulated entity, then we recommend this be treated as a separate goal.

We also suggest adding language in the instructions reminding regulated entities that agreements with vendors do not prevent companies from disclosing proprietary information for the purpose of the evaluation.

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

We very much appreciate the inclusion of a separate section on consumer complaints. However, we think these questions could benefit from more specificity. Question (1) asks regulated entities to provide the total number of complaints "resulting from a process that relied on AI system(s) in past 12 months." We assume this language captures consumer complaints beyond complaints that specifically mention use of an AI system in the adverse action, but this point should be clarified. The Working Group should consider adding language to assist regulated identities to identify a complaint that results from a process relying on an AI system. For instance, the Working Group should consider whether AI augmented an adverse decision that led to a complaint (e.g., an AI tool synthesized large datasets that were ultimately reviewed by claims adjudicators or AI tool that makes a recommendation about a claim or coverage denial

should both be captured by the evaluation tool). Use of AI for an administrative function only that is not related to the subject of the complaint, however, would not be included. We think a specific definition here is necessary and support inclusion of a wide range of consumer complaints.

We appreciate the inclusion of question (2) because consumers may be unaware of how or even whether AI was used in an adverse insurance decision. This question should also include whether the regulated entity provided information about the data and factors used in a decision and recourse to consumers to correct erroneous data aggregated by AI systems.

Exhibit B: (Narrative) AI Systems Governance Risk Assessment Framework
While we appreciate the goal of ensuring a broad capture of consumer indirect and direct harm, we are concerned that question (2), which asks regulated entities to discuss the uses of AI systems, may be too broad. For instance, we are not sure how regulated entities would consistently answer how an AI system is used in decisions that "generates consumer impact directly or indirectly" because it's not clear what is meant by "consumer impact."

In question (2), we support the goal of ensuring that the incorporation of AI systems into insurer workflows does not adversely affect consumers regardless of whether the consumer interacts directly with the AI system or its output.

We believe that end-to-end testing of all the systems involved servicing consumer workflows is essential. When the insurer integrates an AI system into a workflow for servicing a consumer need, both aberrant behavior of the AI system itself as well as problems in the handoffs between the AI system and the other systems and processes used in that workflow can result in adverse impacts to consumers.

Of particular concern are cases when an insurer replaces human evaluation processes with a third party AI system, as these AI-based systems are likely to work in some cases and fail unexpectedly in others. In such failure scenarios, successfully serving the consumer often requires a human to take control back from the AI system and override the AI system. Successfully dealing with these scenarios requires detailed procedures and software setups for overrides, along with testing, training, tracking, and oversight to ensure that the consumer does not end up falling through the cracks.

Handling failure scenarios is often a lower priority in automation projects and therefore deserves careful oversight to avoid consumer harms.

We also think there should be an additional question that asks the extent to which the regulated entity uses an external entity to conduct audits of AI tools as opposed to conducting internal audits.

Exhibit B: (Checklist) AI Systems Governance Risk Assessment Framework

We appreciate the inclusion of a checklist to accompany the narrative and suggest that the question about whether audits are conducted by an external entity be added here as well.

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 include risks of bias and discrimination against protected classes. This language should also be used for questions 4d, 5d, and 6b.

Exhibit C: AI Systems High-Risk Model Details
We are concerned that the evaluation tool allows regulated entities to define which models are "high risk" and which are not. We believe this will leave too much discretion to regulated entities and could end up with some high risk models not subject to the additional scrutiny required by the evaluation too. We encourage the Working Group to look to state definitions of high-risk models (e.g., Colorado) to guide its definition.

We also believe that any testing/review of an AI model should be done with a representative consumer sample that includes demographics and communities that would be subject to the tool.

Exhibit D: AI Systems Model Data Details
We are concerned that the data details included may not be representative of data sources used in the health insurance line. Health specific data details could include electronic health records, genetic information, pre-existing conditions, data from health-related personal applications, and diagnostic data.

We look forward to continuing to partner with you to provide the consumer perspective in this process. If you have questions please contact Amy Killelea (amyk@killeleaconsulting.com).

Consumer Representatives,

Theresa Alban
Chuck Bell
Bonnie Burns
Jalisa Clark
Brenda Cude
Lucy Culp
Deborah Darcy
Shamus Durac
Eric Ellsworth
Joe Feldman
Lauren Finke
Adam Fox

Anna Hyde
Amy Killelea
Kenneth Klein
Jennifer Snow
Deborah Steinberg
Wayne Turner
Richard M. Weber
Silvia Yee

NATIONAL ALLIANCE OF LIFE COMPANIES

An Association of Life and Health Insurance Companies

September 5, 2025

Mr. Scott Sobel
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106-2197
Email: maromero@naic.org

Subject: AI Systems Evaluations - Optional Supplemental Exhibits for State Regulators

Mr. Sobel:

This letter is submitted on behalf of the National Alliance of Life Companies (NALC). The 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.

The NALC appreciates the opportunity to provide comments on the Big Date and AI (H) Working Group's ("Working Group") exposure draft of the AI Systems Evaluations - Optional Supplemental Exhibits for State Regulators ("proposal" or "exhibits"). We commend the Working Group for making these exhibits optional rather than mandatory requirements. This approach would allow NALC members and similarly situated companies to adopt AI governance practices appropriate to their scale and risk profile while avoiding disproportionate compliance burdens that would stifle innovation or create barriers to entry.

As summarized below, however, we do have a number of concerns with the proposal.

1. Inconsistency with NAIC Model AI Bulletin Risk-Based Framework

The NALC strongly supported development of the NAIC Model AI Bulletin: Use of Artificial Intelligence Systems By Insurers ("Model AI Bulletin"). We believe that its description of a risk - based approach is the clearest statement to date of NAIC policy and regulatory expectations concerning insurance carrers' responsible use of AI, and the steps that insurance carriers should take to ensure good governance, ethical use of AI, and compliance with the various existing laws and regulations mentioned therein. Development of any additional regulatory guidance concerning insurer use of AI should be consistent with, and build upon, the approach defined in the Model AI Bulletin.

The proposed exhibits and components are highly prescriuptive and conflict with the risk-based approach Model AI Bulletin:

- Uniform Treatment: The exhibits require identical disclosure for all AI systems regardless of risk level, consumer impact, or regulatory significance
- Scope Mismatch: Required inventories include low-risk internal systems that the Model Bulletin suggests should receive minimal oversight

- Regulatory Efficiency: The comprehensive disclosure approach conflicts with the Model Bulletin's emphasis on focusing regulatory resources on high-risk applications

This inconsistency creates regulatory confusion and undermines the carefully considered risk-based framework that has already been adopted by the NAIC.

2. Disproportionate Impact on Smaller and Mid-Sized Carriers

The proposed exhibits impose uniform compliance requirements regardless of company size or AI sophistication. This creates several problems:

- Resource Disparity: Smaller carriers lack the compliance infrastructure and dedicated AI teams that large insurers maintain, making these detailed inventory requirements prohibitively expensive.
- Competitive Disadvantage: Complex disclosure requirements may force smaller carriers to abandon beneficial AI applications, reducing their ability to compete with larger companies.
- Market Consolidation Risk: Excessive compliance costs could accelerate industry consolidation by making it economically unfeasible for smaller players to maintain AI capabilities.

3. Overly Prescriptive Requirements

The approach and exhibits employ a highly detailed, one-size-fits-all inventory methodology that:

- Lacks Flexibility: Rigid categories and disclosure requirements don't accommodate the diverse ways companies implement AI across different business models and operational contexts.
- Creates Administrative Burden: Extensive data collection requirements divert resources from actual risk management and consumer protection activities.
- Generates Information Overload: Comprehensive inventories may overwhelm regulators with excessive data of varying relevance and quality.

For example, Exbihit A would require carriers to inventory all AI Systems across more than fifteen operational areas. These include a number that have: (i) no direct consumer impact, (ii) and fall outside of the typical scope of regulatory review, e.g.:

- Internal HR systems
- Investment management tools
- Legal/compliance processes
- Strategic operations
- Product performance analytics

We believe that the scope of Exhibit A should be limited to consumer-facing systems that:

- Directly impact consumer outcomes
- Influence pricing, underwriting, or claims decisions
- Generate consumer communications or recommendations

A riak-based approach would allow companies to demonstrate AI governance appropriate to their specific circumstances while achieving regulatory objectives.

The NALC respectfully suggests the following modifications to address these concerns.

Implement Graduated Requirements

- Create disclosure tiers based on company size, AI system risk level, or consumer impact.
- Allow smaller carriers simplified reporting focused on consumer-facing systems.
- Establish materiality thresholds to exclude low-risk applications.

Flexible, Risk-Based Standards

- Replace rigid inventory categories with outcome-focused requirements.
- Allow companies to demonstrate appropriate AI governance through various means.
- Focus on risk management processes rather than exhaustive system cataloging.

Align with Risk-Based Framework

- Limit detailed disclosure to high-risk AI systems as defined in the Model AI Bulletin.
- Exclude internal operational systems with minimal consumer impact.
- Ensure consistency between evaluation tools and established regulatory principles.
- Create safe harbors for preliminary research and planning activities.
- Implement risk-based tiers with different requirements based on system impact levels.
- Ensure confidentiality protections for competitively sensitive information.
- Limit complaint reporting to matters tied directly to consumer outcomes (e.g., claim payments, policy issuance, pricing decisions) rather than routine service dissatisfaction or unrelated issues.

Conclusion

We urge the NAIC to revise the proposed exhibits to:

- Protect smaller carriers from disproportionate compliance burdens.
- Adopt more flexible, risk-appropriate disclosure requirements.
- Maintain consistency with the established risk-based regulatory framework.

These modifications would achieve legitimate regulatory oversight objectives while preserving market competition and innovation incentives across all segments of the insurance industry.

We appreciate the opportunity to comment and look forward to working with the NAIC to develop balanced AI oversight that protects consumers while supporting a competitive insurance marketplace.

Respectfully,

Scott R. Harrison
CEO, National Alliance of Life Companies

cc: Michael Humphries, Chair, Big Data and Artificial Intelligence (H) Working Group

NAMIC
NATIONAL ASSOCIATION OF MUTUAL INSURANCE COMPANIES

317.875.5250 | [F] 317.879.8408
3601 Vincennes Road, Indianapolis, Indiana 46268

202.628.1558 | [F] 202.628.1601
20 F Street N.W., Suite 510 | Washington, D.C. 20001

September 5, 2025

Commissioner Humphreys (PA), Chair
NAIC Big Data and Artificial Intelligence (H) Working Group
c/o Miguel Romero, Director, P&C Regulatory Services
Via email maromero@naic.org

Re: NAMIC Comments on the Big Data and AI (H) Working Group AI Systems Evaluation Tool Exposure

Dear Commissioner Humphreys and Members of the Working Group:

On behalf of the National Association of Mutual Insurance Companies (NAMIC)[1], we would like to thank the NAIC Big Data and Artificial Intelligence (H) Working Group for requesting and accepting comments on its recent AI Systems Evaluation Tool exposure. NAMIC appreciates the Working Group's continued attention to the topic of artificial intelligence, as the use of AI in the insurance industry promises numerous benefits for consumers and carriers alike. We continue to encourage, and believe it critical, that the NAIC provide transparent, robust, and iterative opportunities to work collaboratively with all stakeholders on these matters to avoid inadvertently creating overly burdensome frameworks that prevent the benefits of AI from being realized by consumers and the industry.

In this vein, NAMIC provides the following substantive comments and areas of encouraged focus for the Working Group. We have also included a technical edits section, which provides our initial edit considerations for the exposed tool draft.

SUBSTANTIVE COMMENTS

[1] Start from existing financial and market conduct exam standards, then identify gaps
NAMIC encourages the Working Group to avoid creating new and separate standards for AI systems without first identifying what exists in financial and market conduct exam standards today, and from there, identifying key risks and any potential gaps that may need addressing.

To create evaluation standards without first examining and recognizing what already exists in a regulator's toolbox will only result in a confusing, conflicting exam framework that will increase compliance burden and cost (for departments, insurers, and ultimately consumers) without demonstrable consumer benefit, and

[1] The National Association of Mutual Insurance Companies consists of over 1,300 member companies, including six of the top 10 property/casualty insurers in the United States. The association supports local and regional mutual insurance companies on main streets across America as well as many of the country's largest national insurers. NAMIC member companies write $383 billion in annual premiums and represent 61 percent of homeowners, 48 percent of automobile, and 25 percent of the business insurance markets. Through its advocacy programs NAMIC promotes public policy solutions that benefit member companies and the policyholders they serve and fosters greater understanding and recognition of the unique alignment of interests between management and policyholders of mutual companies.

stifle innovation of a technology that has the potential to benefit consumers and industry alike. It is for this reason that the Working Group was charged with first identifying "existing tools, resources, materials, and training that will assist and guide regulators in their review of AI systems."[2] NAMIC respectfully suggests that that Working Group has not yet completed this important charge, and that it would be inappropriate for the Working Group to utilize an evaluation tool as an exploratory exercise to complete this charge.

By first documenting all the ways in which AI is currently regulated and examined under existing statutes and regulations, the Working Group will be better equipped to complete its secondary charge of creating new regulatory tools – if necessary. It is worthwhile to note that the Working Group has completed numerous surveys in recent years that provide insight into how carriers are using AI, which the Working Group may find useful in identifying existing statutes and regulations that cover the AI uses by carriers. Completing this first charge allows any new regulatory tools to be more targeted, risk-based, and effective.

Overlap with Existing Market Conduct, Financial, and ORSA Exams

NAMIC notes for the Working Group that there are areas of overlap between the proposed evaluation tool and existing market conduct, financial, and ORSA exams. By way of example, there are places throughout the evaluation tool that address AI use relative to underwriting or ratemaking. Any model used in ratemaking would be filed under existing law and regulations, and existing market conduct exam guidelines and financial exam standards thoroughly address consumer impact and compliance with existing law, regardless of whether AI is being used in those processes. As another example, insurers can expect in existing financial exams to produce documentation on governance, and notably, IT exam portions of the financial exam cover governance, process, policy and IT protocols. Taking this further, non-financial risks in general would be covered by the Inherent Risk Identification or Risk Strategies aspects of a financial exam. As such, we reiterate our suggestion that the Working Group leverage its survey results to identify areas in which insurers are using AI, identify existing regulatory tools that are applicable, and identify how different jurisdictions may already be covering those uses. From there, the Working Group can then determine whether there are questions it feels aren't being asked as part of existing processes, and turn its focus toward those questions or gaps, if they exist.

The Working Group Should Support Innovation and Advancement of AI

NAMIC continues to believe that a goal of this working group should be supporting the innovation and advancement of AI in insurance, which will benefit consumers. As the NAIC noted in its recent letter to Senate leadership, "State insurance regulators understand that AI is a transformative technology that can be leveraged to benefit insurance policyholders by, among other things, creating new product offerings, improving the efficiency of the insurance business, and transforming the consumer experience."[3] Avoiding dual, overlapping, and conflicting standards furthers that goal and helps consumers realize the benefits of AI in the insurance industry.

---

[2] *2025 Charges*, NAIC Big Data and Artificial Intelligence Working Group, https://content.naic.org/committees/h/big-data-artificial-intelligence-wg.

[3] NAIC Letter to The Honorable John Thune, The Honorable Chuck Shumer, The Honorable Ted Cruz, and The Honorable Maria Cantwell, https://www.insurereinsure.com/wp-content/uploads/sites/919/2025/06/government-affairs-letter-ai-moratorium.pdf

[2] Focus evaluation standards on outcomes rather than the tool used

AI is not a new technology. AI development and use date to the 1950s, and AI models like machine learning models have been used in insurance for quite some time, without the need for separate, distinct standards aimed specifically at the technology. Importantly, there is no bright line which constitutes whether something is or is not "AI." Technologies that "perform functions that are normally associated with human intelligence" (NAIC AI definition) have been available since the earliest computers. For example, machine learning models such as Generalized Linear Models (GLMs) have been used by actuaries and carriers for pricing and managing risk for many years, and the use of these tools has been governed by existing insurance law on pricing and risk analysis. Notably, NAMIC is not aware of any identified consumer harm or regulatory gaps derived from the longstanding use of GLMs. This is because existing insurance law focuses on the insurance activity, and regardless of whether AI models are incorporated into a process or not, existing statutory frameworks apply equally. Without clarifying the scope of the tool, or focusing on outcomes, it is likely to be either redundant with other regulatory oversight or irrelevant to any novel risks presented by newer technologies. NAMIC suggests the evaluation framework focus on specific activities which may need further oversight, not on the tool being used for those activities.

AI is simply an advanced data analytics tool used at the direction of human workers in furtherance of traditional insurance activities already governed by existing law and exam processes.[4] Yet, the exposed AI Systems Evaluation Tool appears premised on the unproven assumption that AI is inherently riskier than humans doing the same work without AI, as it holds AI to a different standard than non-AI tools when used for those same processes and outcomes. We encourage the Working Group to have a comprehensive, thoughtful discussion around the consequences of creating separate standards for a specific data analytics tool when existing law and exam processes already cover the regulated activities for which AI would be used.[5]

Because the current draft of the tool is not tied to verified risks of AI nor to gaps in existing regulation, the evaluation tool appears to be creating a solution disjointed from needs of consumer protection and of the market. The tool's introduction suggests the exhibits could be used to supplement "market conduct, product review, form filing, financial analysis, and financial examination review procedures," which encompasses nearly all regulatory engagement. NAMIC encourages the Working Group to be far more precise and targeted in the proposed questions, the intended uses, and the problem the Working Group is attempting to solve. As exposed, the tool would generate stifling compliance costs without clear consumer benefits.

---

[4] The NAIC itself has recognized this through the Model Bulletin on the Use of Artificial Intelligence Systems by Insurers, wherein it states that "decisions or actions impacting consumers that are made or supported by advanced analytical and computational technologies, including Artificial Intelligence (AI) Systems, must comply with all applicable insurance laws and regulations." Through this verbiage, the NAIC's Model Bulletin correctly reinforces that there is already a robust statutory framework existing in insurance laws and regulations that apply to insurers' use of AI.

[5] *See*, *The $38 Billion Mistake: Why AI Regulation Could Crush Florida's Economy* (finding that the potential costs to the state's economy from stricter AI rules would cost $38 billion annually over the next ten years, and "significantly harm Florida's economic growth.") https://jamesmadison.org/wp-content/uploads/PolicyBrief_AI-Jun2025-v03.pdf.

[3] Evaluation tools must be tied to existing law and the requirements therein
Evaluation tools are meant to evaluate compliance with existing law, and the market regulation and financial examination handbooks are both rooted in NAIC model law at a minimum. Any evaluation tool that addresses AI models must therefore be tied to existing law and the requirements therein, rather than create new requirements. In the current draft, there are questions and standards aimed at aspects like bias testing requirements, human in the loop requirements, adverse consumer outcomes, and third-party data and model vendor requirements, all of which establish standards untethered to existing insurance statutes. As NAMIC noted in its comments throughout the *Model Bulletin on Use of Artificial Intelligence Systems by Insurers* drafting process, the bulletin's use of the term "bias" is in a manner which we believe cannot be implemented by insurers consistent with the controlling legal framework pertaining to fair and unfair discrimination in risk classification, and "adverse consumer outcomes" is not a standard that exists within the insurance code. With respect to using a "human in the loop," NCOIL is currently considering a draft model law requiring a human in the loop for claim denials. Notwithstanding the fundamental flaws in the NCOIL draft, the fact that a model law would be necessary for that requirement indicates it is not a requirement in existing law today. Finally, the NAIC Third Party Data and Model Vendor Working Group's very purpose is to examine the limits of regulatory authority over such third-party vendors, and determine whether a model law or regulatory framework may be necessary in that area.

NAMIC also suggests the Working Group consider use of the evaluation tool by, and submission of responses to, an insurer's domestic regulator within the scope of that regulator's authority. Those responses to the domestic regulator could then be shared with other departments in accordance with appropriate confidentiality protections. A divergent, state-by-state approach to deploying the tool may be suboptimal in the way of consistency and effort for departments as well as for carriers.

INITIAL TECHNICAL EDITS

Notwithstanding NAMIC's foundational concerns highlighted in the preceding sections of this letter, we provide initial technical edit suggestions and explanations as follows.

Exhibit A – ***Narrow the scope and focus on areas of use rather than quantity***
At the outset, the information sought in Exhibit A appears to closely resemble the recent surveys conducted by the Big Data and Artificial Intelligence Working Group. NAMIC therefore encourages the Working Group to identify the need for including this information in an evaluation tool when the Working Group has obtained this information through its surveys. If the Working Group isn't accomplishing anything distinct with Exhibit A that isn't already accomplished through the surveys, Exhibit A should be removed. If the Working Group identifies a need for Exhibit A, NAMIC then encourages the Working Group to narrow the scope. First, the definition of AI Systems is broad and NAMIC encourages the Working Group to articulate which types of AI models it is most concerned with. Second, NAMIC encourages the Working Group to align

the scope of the evaluation tool itself and the exhibits with the scope of the Third-Party Data and Models Working Group. The Third-Party Working Group stated that its focus is on insurance practices with direct consumer impact like rating, underwriting, claims, and fraud detection.

Separately, NAMIC suggests that the focus of Exhibit A be on the areas AI Systems are being used rather than focusing on number of models. Focus on areas of use will provide insight into whether AI is being used in specific business practices and is a less burdensome ask of companies than asking insurers to detail the number of models for each category. NAMIC notes that a focus on the areas of use is consistent with how global regulatory inquiries are handled, which would bring NAIC in alignment from an international perspective.

Finally, NAMIC suggests removal of retired models from scope of evaluation and review. As an alternative option, NAMIC suggests a timeframe limitation on evaluation of retired models with this tool.

In line with these thematic suggestions, we propose initial technical edits for Exhibit A below.

Exhibit A Technical Edits

- Suggested title change:
    - Exhibit A. How are Regulated Entities Using AI Systems
- Company Instructions:
    - We suggest that the company instructions read: "Provide the areas in which AI systems are being used as requested. See definitions below." Simply referring to the definitions section rather than attempting redefinition here avoids ambiguity in application and scope.
- Proposed Removals:
    - (P.3) Period Defining the Last 12 Months
    - (P. 3) Period Defining the Next 6 Months
    - (P. 4) Number of AI System Model(s) Implemented in Past 12 Months
    - (P. 4) Number of AI system Model(s) Planned to be Implemented Within Next 6 Months
    - (P. 4) "Number of" removed from all columns except for Consumer Complaints Column
    - (P. 4) Legal/Compliance
    - (P. 4) Investment/Capital Management
    - (P.4) Product Performance
    - (P.5) Strategic Operations (HR, Reinsurance, etc.)
    - (P. 5) Consumer Complaints. 1. What is the total number of consumer complaints resulting from a process that relied on AI system(s) in the past 12 months? (information is already captured in the chart portion).

Exhibit B – ***Delete the narrative portion and follow the checklist***

The substance of the narrative and checklist portions of Exhibit B differ, with the checklist more closely following what is detailed in the *NAIC Model Bulletin on Use of Artificial Intelligence Systems by Insurers*.

NAMIC therefore encourages the Working Group to delete the narrative portion of Exhibit B, and instead leverage the checklist. Following our substantive comments that the evaluation tool must be tied to existing law and the requirements therein, NAMIC proposes initial technical edits for Exhibit B below.

Exhibit B Checklist Technical Edits:

- Change "Page #" column to "Yes/No"
    - Each item may not be in the AI Systems Program documentation itself, but may be an element of the company's framework (e.g., training).

- Evaluation Metrics Lacking Existing Legal Standard – Proposed Removal
    - **Removal of "bias" as an evaluation metric.** "Bias" is not an existing standard in insurance law, and because the evaluation tool metrics must be tied to existing law and the requirements therein, NAMIC requests removal of all references to "bias" in the Exhibit B checklist. In our review, this would include removal of 4b. in its entirety, and removal of the term "bias" from 4d. and 5d.
    - **Removal of Third Party or External Data/AI system evaluation metrics.** NAMIC suggests removal of Ref. 6 (External Data and AI System Practices) in its entirety. Third-party vendor and external data or AI system practices are currently under review by the NAIC Third Party Data and Models (H) Working Group. Because proposed standards or frameworks have not been reached yet, it would be premature to include related evaluation standards or metrics in the evaluation tool.
    - **Removal of "Human in the Loop" evaluation metrics.** Necessity for a "human in the loop" will vary based on use case and is not a current standard or requirement in insurance law. NAMIC encourages further discussion on this topic and the Working Group to evidence what problem it is looking to solve with "human in the loop" standards ahead of including it in the proposed evaluation tool. NAMIC therefore suggests removal of 4h. and 4i. in their entirety.

- Redundant Evaluation Metrics – Proposed Removal
    - NAMIC noticed areas within the Exhibit B checklist that ask for information that is already covered by Exhibit A, or is duplicative within Exhibit B itself. **We noted this redundancy in the following Ref. #s, and therefore suggest removal:**
        - (3d.) Other boxes in Ref. 3 cover consumer harms in the insurance code
        - (3o.) The information in 3o. is already covered by Exhibit A
        - (4f.) This information is covered by other questions in the Exhibit B checklist
        - (5a.) This information appears covered by 4a.

- Ambiguous or Overly Broad Evaluation Metrics – Proposed Removal or Additional Clarity
    - In NAMIC's review, we noted multiple instances in the Exhibit B checklist where the evaluation standard was either ambiguous or overly broad, and **suggest removal or clarity as follows.**
    - 3b. Ensures AI Systems are used ethically
        - This metric is ambiguous, and NAMIC therefore suggests its removal.

- 3g. Monitors and measures the benefits of AI systems
    - Monitoring or measuring benefits of an AI System is up to the insurer to identify, and NAMIC is not certain that this information is directly tied to identifying risks with respect to use of AI in insurance operations. We therefore suggest its removal.
- 3m. Quantifies AI system risk levels
    - This practice may vary by company. It may be the case that companies tier AI use by importance and risk, but it is unclear or may vary how risk is "quantified." We therefore suggest its removal.
- 3p. Ensures consumer awareness in use of AI systems through disclosures, policies, and procedures for consumer notification.
    - NAMIC recommends deleting or considerably narrowing the scope in 3p., as insurers might use AI models to automate backend processes and in other areas that are not consumer-facing nor have consumer impact.
- 5b. The extent to which the data and/or AI systems are representative of the population the model is being applied to
    - It is unclear what the Working Group is asking for in 5b., NAMIC encourages rewording or removal.
- 5c. Additional purposes the model outputs or inputs from other models are used for
    - It is unclear what the Working Group is asking for in 5c., NAMIC encourages rewording or removal.

Exhibit C – ***Delete due to redundancy***
Exhibit C appears duplicative of information that would be gathered in Exhibits A and B, and we therefore suggest its removal from the evaluation tool.

Exhibit D – ***Delete due to redundancy and lack of utility***
NAMIC suggests removing Exhibit D in its entirety. These models would already be subject to reviews conducted in market conduct examinations when assessing compliance with rating laws and unfair trade practices. Further, its inclusion of new standards focused on data and third-party use is premature, as the Third-Party Data and Models Working Group is currently undertaking an effort focused on third parties.

IN SUMMARY

We close by again thanking the Working Group for allowing NAMIC to submit comments to engage in this extremely important discussion regarding artificial intelligence and insurance, and we urge you to continue offering additional iterative opportunities for robust, transparent conversations on the matter. NAMIC endeavors through these comments to highlight areas that the Working Group should especially direct its

focus. Finally, we would be remiss if we did not encourage open collaboration with other workstreams at the NAIC that are similarly looking at (or have already developed) model review standards.[6]

NAMIC looks forward to continuing our work with the Working Group to arrive at solutions that protect and stabilize the insurance marketplace while fostering growth and innovation that benefit all stakeholders.

Sincerely,

Lindsey Klarkowski
Policy Vice President – Data Science, Artificial Intelligence, and Cybersecurity
NAMIC

[6] As of the date of this letter's drafting, NAMIC is aware of the following workstreams addressing model evaluation to some extent: GO/ES, CASTF, Financial Condition Committee, Third Party Data and Models WG.

REINSURANCE ASSOCIATION OF AMERICA

September 5, 2025

Michael Humphreys, Chair
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
c/o Scott Sobel
AI Policy Advisor
Via email ssobel@naic.org

Re: RAA Comments Regarding the AI Systems Evaluation Tool

Dear Commissioner Humphreys:

The Reinsurance Association of America (RAA) appreciates the opportunity to submit comments to the Big Data and Artificial Intelligence (H) Working Group regarding the AI System Evaluation Tool, and also for the Working Group's extension of the comment deadline. The Reinsurance Association of America (RAA) is a national trade association representing reinsurance companies doing business in the United States. RAA membership is diverse, including reinsurance underwriters and intermediaries licensed in the U.S. and those that conduct business on a cross-border basis. The RAA also has life reinsurance affiliates and insurance-linked securities (ILS) fund managers and market participants that are engaged in the assumption of property/casualty risks. The RAA represents its members before state, federal and international bodies.

The RAA appreciates the Working Group's continued thoughtful engagement on this important issue. The RAA's main concern with the AI System Evaluation Tool is the lack of differing treatment of reinsurers. Many of the concerns, objectives, and proposed uses of the AI System Evaluation Tool seem targeted towards concerns with the use of AI by primary insurers, which would make less sense if applied to reinsurers. The RAA urges the Working Group to consider whether reinsurers should be treated differently under the AI System Evaluation Tool.

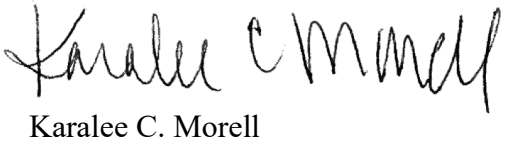
The RAA appreciates the goal of the Working Group is to establish a useful, standardized tool for evaluating AI systems in use in the insurance industry, particularly with regard to how the use of AI impacts consumers and the consumer experience. Given that lens, it is imperative that any promulgated regulation relating to AI: (1) accounts for the contractual separation of reinsurance from consumers; and (2) recognizes that the use of AI by reinsurers represents a drastically reduced, if not eliminated, risk of adverse consumer outcomes. For example, Exhibit A contains a section on consumer complaints; the reinsurance industry does not have a direct relationship with consumers and would be unable to evaluate their use of AI on this standard. Accordingly, the RAA respectfully requests that the tool include instructions making clear which portions of the AI Systems Tool are relevant to reinsurers specifically, as opposed to other industry participants (insurers, brokers, etc.). This will not only make responding more efficient; it will also help enable

the Tool to effectively evaluate the use of AI as it impacts consumers. Including some recognition that reinsurers are not expected to respond to all aspects of the AI System Evaluation Tool and the reinsurance industry's use of AI will be different from other responders as it impacts consumers would enable reinsurers to fill out the AI System Evaluation Tool in a way that is most useful to regulators.

While the RAA is most concerned with the differing treatment of reinsurers under the AI System Evaluation Tool, we also share the concerns that the primary insurance trades have highlighted in their comment letters. In particular, the RAA is concerned about the clarity of the existing definitions and lack of definitions for other terms throughout the Tool. The RAA is also concerned that the definitions may be overly inclusive and cause routine analytics tools to be considered AI Systems. The RAA also is concerned about how the AI System Evaluation Tool will balance the understandable need for transparency of AI models and the proprietary nature of the models. Some level of confidentiality is needed to ensure proprietary information is not exposed. Lastly, it also remains unclear how the AI System Evaluation Tool will currently fit into the NAIC regulatory framework and some clarity on how the Tool will be used going forward would be helpful.

The RAA appreciates the opportunity to work with you on this important project and specifically to address our reinsurance-specific concerns. We would be happy to meet with members of the Big Data and Artificial Intelligence (H) Working Group and NAIC staff to discuss how reinsurers should be treated differently under the AI Systems Evaluation Tool. We look forward to further engagement on these issues.

Sincerely,



Karalee C. Morell
SVP and General Counsel
Reinsurance Association of America

MEMO

COMPANY NAME**: REGULATORY COMPLIANCE & EXAMINATIONS CONSULTANTS LLC**

To: Scott Sobel AI Systems FCAS, MAAA, CSPA, MSPA

From: Earnest Collins CISSP, CISM, CISA, CGAI, CFE, AES

Date: September 2, 2025

Re: AI System Evaluation Tool Draft for exposure 07 07 2025

COMMENTS:

AI Systems Evaluations Optional Supplemental Exhibits for State Regulator

---

Exhibit A: Quantify the Regulated Entity's Use of AI systems.

- Should we have a column with **Model Type,** i.e., AI Generative, Predictive, Deterministic, Prescription, Anomaly Detection, Natural Language Processing, etc.?
- The **model type** will impact risk identification and assessment, as well as the AI system's key risk governance frameworks, which will help define high-risk models.

Exhibit B: (Narrative) AI Systems Governance Risk Assessment Framework

Should we include an AI system used to address the following questions?

- How does the Insurance Company Assess **autonomy**, **reversibility,** and **reporting impact** risk of AI systems

Exhibit C: AI Systems High–Risk Model Details NO CHANGES

Exhibit D: AI Systems Model Data Details

- Should there be a column for Data Architecture Paradigm and Integration Spaces,
- i.e., Data Lakehouse, Open Data Architecture & Intelligent Pipelines Orchestration?
- **Risk** Inconsistent metadata and lineage tracking across cloud, on-prem, and third-party systems.
- **Impact**: Enables lateral movement, data exfiltration, and unauthorized access to sensitive datasets.
- Expand the attack surface for Supply Chain Vulnerabilities

Texas Department of Insurance
PO Box 12030 | Austin, TX 78711 | 800-578-4677 | tdi.texas.gov

September 3, 2025

NAIC Big Data and Artificial Intelligence (H) Working Group — VIA EMAIL ONLY:
Attn: Scott Sobel, NAIC AI Policy Advisor and — ssobel@naic.org
Miguel Romero, Director, P&C Regulatory Services — maromero@naic.org
Michael Humphreys, Chair — mhumphreys@pa.gov
Mary Block, Co-Vice Chair — mary.block@vermont.gov
Doug Ommen, Co-Vice Chair — doug.ommen@iid.iowa.gov

RE: Draft AI Systems Evaluation Tool

Dear Commissioner Humphreys and working group members:

Thank you for this opportunity to comment on the draft AI Systems Evaluation Tool (the Tool), which we believe has a nexus to the group's recent public request for information (RFI) about: (1) proposing an AI model law, and (2) which template should be used for the Tool.

TDI submits the following comments for the working group's consideration:

Coordination with Other Committees

The NAIC has a history of ensuring coordination among committees and groups that have overlapping interests in a topic. This coordination can occur in many ways, such as joint meetings, referrals between groups, and joint exposures. The coordination is also actively discussed at meetings. With the Innovation, Cybersecurity, and Technology (H) Committee being newly formed, this coordination effort must be actively considered by groups reporting to it. The subject matter experts for the functions where the Tool will be most beneficial to insurance department processes may not be closely monitoring this working group. Therefore, we are concerned that the working group has not discussed a plan to seek input on the Tool from other committees, particularly the Market Regulation and Consumer Affairs (D) and Financial Condition (E) Committees. Likewise, other committees are not necessarily following the events of the H Committee and may have a stake and interest in usage of the Tool, even if the Exhibits are held out to be for "optional" use.

We believe that such a Tool would ordinarily be discussed and considered in the Market Conduct Examination Guidelines (D) Working Group and the Examination Oversight (E) Task Force and its

appropriate working groups for any usage in exams or analysis functions. The stated purpose of developing the Tool is to assist insurance departments and insurers in evaluating this evolving area of insurer operations. Without coordination and discussion with the established groups responsible for oversight of these functions, potential improvements to the Tool may not be identified and the full benefit of the Tool may not be realized.

The expectation of coordination is noted in the Working Group's charges because charge 4(b)

states:

"*Facilitate discussion* related to AI systems evaluation, including: . . .

(b) Develop new regulatory tools or regulatory guidance to assist regulators in their review of AI systems used by licensees, including an insurer's AI program."

We believe that facilitation ensures that regulators in the appropriate working groups and committees are informed, consulted, and have an opportunity to provide input before tools and guidance are launched and made available for use.

Optional or Immediate Use of the Tool May Lead to Compliance Concerns

Exposure of this Tool - even for optional, immediate use – may create compliance concerns for two reasons. First, the Tool contains definitions not in an adopted model law and may innocently lead to the failure of state regulators to align the Tool with their own state-specific laws. Second, in the May 2025 RFI and in the Working Group's discussions, there have been statements that the Exhibits can be used by insurers for "self-assessment." If the Tool is used in any manner for insurer "self-assessment," it can and may lead insurers to an erroneous belief or misunderstanding that they are in compliance.

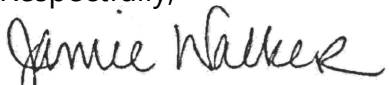
Concern if Use of the Tool is Required in the Future

In addition, while currently optional, any future required use of the Tool may cause concerns while state regulatory approaches to artificial intelligence remain in flux. Lastly, Texas would have significant concerns about any future consideration of requiring states to use the Tool in the future, either through inclusion in a handbook or another accreditation standard related to AI, given the inherent rapidly changing environment of use and regulation of artificial intelligence.

TDIappreciates the working group's efforts. We hope our comments are productive in reaching clarification on issues that are significant to TDI and all states' regulatory efforts.

Thank you for your time, attention, and consideration.

Respectfully,



Jamie Walker
Deputy Commissioner

COMMONWEALTH OF VIRGINIA

SCOTT A. WHITE
COMMISSIONER OF INSURANCE
STATE CORPORATION COMMISSION
BUREAU OF INSURANCE

P.O. BOX 1157
RICHMOND, VIRGINIA 23218

1300 E. MAIN STREET
RICHMOND, VIRGINIA 23219

TELEPHONE: (804) 371-9741
www.scc.virginia.gov/boi

September 5, 2025

The Honorable Michael Humphreys
Pennsylvania Insurance Department
1326 Strawberry Street
Harrisburg, PA 17120

RE: Comments on the AI Systems Evaluation Tool

Dear Commissioner Humphreys,

The Virginia Bureau of Insurance ("Bureau") appreciates the ability to offer its comments on the Request for Comments on the AI Systems Evaluation Tool issued by the Big Data and Artificial Intelligence (H) Working Group.

The Bureau had provided for circulation a document "Virginia AI Examination Framework Draft," which was included in the Notice from the Big Data and AI Working Group to membership dated July 3, 2025. A copy of that document is included in these comments for reference.

It is the Bureau's belief that the AI Systems Evaluation Tool ("AI Tool") and the Virginia AI Examination Framework ("VA Framework") are highly complementary in nature and, if blended, could provide a unified regulatory oversight tool to aid insurance regulators in their efforts related to AI risk management and consumer protection.

The AI Tool presented as optional exhibits for state regulators is designed to quantify AI system usage and facilitate risk assessment on an ongoing basis. It focuses on collecting detailed information directly from insurers through four main exhibits:

> Exhibit A - quantifies the use of AI systems, including current and planned AI model counts, impacts on consumers, and consumer complaint information.
>
> Exhibit B - provides checklist/narrative options for AI Systems Governance Risk Assessment information collection on financial reporting, IT systems and data, and the insurer's Risk Assessment Framework. It collects detailed information on bias assessment, compliance, data privacy, employee training, and vendor standards.
>
> Exhibit C - seeks specific information on high-risk AI models that could cause adverse consumer, financial, or financial reporting impact, including their risk classification, testing, and use cases.
>
> Exhibit D - requires detailed information on the sources and types of data used in AI models (internal vs. third-party) to identify potential risks of unfair trade practices or adverse consumer impact.

The AI Tool exhibits strength in comprehensive data collection, detailed governance checklists, granular data element analysis, and direct consumer impact focus. The AI Tool may not be as strong due to the use of optional supplemental exhibits, which has potential for inconsistent adoption and application across the NAIC membership. The AI Tool is also less prescriptive on

examination methodology. In other words, the AI Tool collects vital data, but does not provide direct, step-by-step guidance on how to conduct the examination or how to evaluate the submitted information.

In blending the AI Tool with the VA Framework, the Bureau thought to enhance the AI Tool by providing draft examination procedures for assessment of insurers' AI/ML governance and risk management practices. The VA Framework builds on the principles-based examination framework already in use by financial regulators, prioritizing governance effectiveness and focusing attention on AI systems with significant consumer impact.

The VA Framework suggests an examination structure with three main components:

> AI Governance Assessment - determines if the insurer has established appropriate governance structures and policies for overseeing AI/ML systems, including Board oversight, risk assessment methodology, and Enterprise Risk Management integration.
>
> AI Systems Inventory and Risk Classification - assesses whether the insurer properly identifies, classifies, and documents AI systems based on risk level and potential consumer impact, including evaluation of high-risk systems and human oversight.
>
> Data Governance Assessment - evaluates if the insurer has adequate controls for data quality, representativeness, and compliance, including due diligence for third-party data.

The VA Framework exhibits strengths in examination-oriented guidance, a principles-based approach, strong emphasis on governance structure, prioritization of significant consumer impacts, and specifically defines a "High-Risk AI System." The VA Framework is not as strong as the AI Tool in that it collects less quantitative detail, has less granular data element review, and is less detailed on specific model testing frequencies.

It is the Bureau's recommendation that the AI Tool and VA Framework be blended where the VA Framework provides examination procedures and best practices for conducting AI examinations and that the AI Tool's exhibits be mandated as the documentation request within the examination framework. It is also our recommendation that a definition of "High-Risk AI System" be included in any examination framework produced.

In summation, by integrating these concepts, regulators would have actionable examination steps with comprehensive, quantitative data collection, leading to a more robust AI examination process.

The Bureau offers these comments for the Working Group's consideration and looks forward to providing a draft of the blended approach at the Working Group's request.

Sincerely,
/s/ Eric C. Lowe
Chief, Data Analytics and Cyber
Innovative Solutions & Strategies Division

Virginia AI Examination Framework Draft for Discussion, as circulated on July 3, 2025

AI/ML Examination Best Practices for Insurance Regulators

Purpose

This document provides insurance regulators with examination procedures to assess insurers' AI/ML governance and risk management practices. It offers a principles-based examination framework that prioritizes governance effectiveness over quantifying AI usage, focusing regulatory attention on systems with significant consumer impact.

Examination Approach

Initial Documentation Request

Request these core governance documents at the beginning of the examination AI/ML Governance Policy/Framework:

- Board and committee materials discussing AI oversight Enterprise risk management
- Documentation addressing AI Inventory of AI systems with risk classifications
- Data governance policies
- Vendor management policies for AI providers

Examination Procedures

Component 1: AI Governance Assessment

Objective: Determine whether the insurer has established appropriate governance structures and policies for overseeing AI/ML systems.

Examination Steps

1.1 Governance Framework Review

Obtain and review AI Governance Policy

Best Practice Indicators:

- Formal, board-approved AI governance policy exists
- Policy addresses key risk areas including consumer protection, bias prevention, and regulatory compliance
- Policy integrates with enterprise risk management framework

Red Flags:

- No formal governance policy
- Does not account for key AI risks

1.2 Board and Senior Management Oversight Assessment

Review board/committee minutes for AI discussions

Best Practice Indicators:
- Regular board/committee discussions of AI risks
- Senior management demonstrates understanding of AI risks

Red Flags:
- No evidence of board-level discussions about AI

1.3 Risk Assessment Methodology Evaluation
Review methodology for classifying AI systems by risk level

Best Practice Indicators:
- Documented methodology for risk classification
- Risk levels consider potential consumer impact
- Higher-risk systems subject to enhanced controls
- Regular reassessment of risk classifications

Red Flags:
- No risk classification methodology
- Consumer impact not considered in risk assessment
- Uniform controls regardless of risk level
- No process for risk reassessment

1.4 Enterprise Risk Management Integration
Verify AI risks are included in ERM framework

Best Practice Indicators:
- AI risks formally integrated into ERM framework
- AI considerations included in ORSA

Red Flags:
- AI risks excluded from ERM framework
- ORSA does not address AI risks

Component 2: AI Systems Inventory and Risk Classification

Objective: Assess whether the insurer properly identifies, classifies, and documents AI systems based on risk level and potential consumer impact.

Examination Steps

2.1 Systems Inventory Assessment
Review comprehensive inventory of AI systems including both internally developed and third-party systems

Best Practice Indicators:
- Comprehensive inventory of all AI systems
- Both internal and vendor systems included
- Clear documentation of system purposes

Red Flags:
- Incomplete or missing inventory
- Third-party systems not included

2.2 Risk Classification Review

Evaluate criteria used for risk classification

Best Practice Indicators:

- Documented risk classification criteria
- Control requirements scaled by risk level

Red Flags:

- Undefined or subjective classification criteria
- Inconsistent application of criteria

2.3 High-Risk Systems Identification

Review identification of high-risk systems including the identification of systems making automated decisions

Best Practice Indicators:

- Clear identification of high-risk systems
- Appropriate human oversight for high-risk systems (Human-in-the-Loop)

Red Flags:

- No distinction for high-risk systems
- Automated decision systems not identified

Component 3: Data Governance Assessment

Objective: Evaluate whether the insurer has implemented adequate controls to ensure data quality, representativeness, and compliance with regulations.

Examination Steps

3.1 Data Governance Framework Review

Assess data governance framework for AI systems

Best Practice Indicators:

- Comprehensive data governance framework
- Robust data quality management

Red Flags:

- Inadequate or missing data governance
- Poor data quality management

3.2 Third-Party Data Assessment

Review due diligence for third-party data providers

Best Practice Indicators:

- Thorough due diligence for data providers
- Comprehensive validation of data quality

Red Flags:

- Limited due diligence for data providers
- Inadequate validation of data quality

3.3 Sensitive Data Review

Identify use of sensitive or protected characteristics

Best Practice Indicators:

- Clear identification of sensitive data

Red Flags:

- Unclear identification of sensitive data

Examination Report Documentation

When documenting examination findings related to AI/ML systems, consider these guidelines:

Governance Assessment Summary

Document the insurer's AI governance maturity level (advanced, developing, or inadequate) based on:

- Governance framework completeness
- Board and senior management oversight
- Integration with enterprise risk management

Risk Classification Assessment

Summarize findings regarding the insurer's approach to:

- Identifying and classifying AI systems
- Risk-based controls and oversight
- High-risk system identification

Data Governance Assessment

Summarize findings regarding:

- Third-party data controls
- Sensitive data handling

Definitions and Appendix

For the purposes of this examination, the following terms are defined:

"Adverse Consumer Outcome" refers to a decision by an Insurer that is subject to insurance regulatory standards enforced by the Department that adversely impacts the consumer in a manner that violates those standards.

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

"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.

"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.

"Consequential Decision" means a decision that has a material legal or similarly significant effect on the provision or denial to any consumer or the cost or terms of insurance.

"Degree of Potential Harm to Consumers" refers to the severity of adverse economic impact that a consumer might experience as a result of an Adverse Consumer Outcome.

"Externally Trained Models" Transferred learnings from pre-trained models developed by a third party on external reference datasets.

"Generalized Linear Models (GLMs)" Including Ordinary least squares (OLS), Elastic Net/LASSO/Ridge regression, Logistic Regression, and Generalized Additive Models (GAMs) are not considered to be machine learning models for this examination.

"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.

"High Risk Artificial Intelligence System" means any artificial intelligence system that is specifically intended to autonomously make, or be a controlling factor in making, a consequential decision. A system or service is not a "high-risk artificial intelligence system" if it is intended to (i) perform a narrow procedural task, (ii) improve the result of a previously completed human activity, (iii) detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review, or (iv) perform a preparatory task to an assessment relevant to a consequential decision.

"Internally Trained Models" Models developed from data internally obtained by the insurer.

"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.

"Machine Learning Models" Include but not limited to: Clustering, such as K-means and Principle Component Analysis (PCA), Decision Trees (DTs), Tree-Based Ensembles (Random Forest, Gradient Boosting), Support Vector Machines (SVM/SVR), K-Nearest Neighbors (kNN), and Genetic Algorithms (GAs).

"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 Models" Include 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.

"Third Party" for purposes of this bulletin means an organization other than the Insurer that provides services, data, or other resources related to AI.

"Validation Method" The source of the reference data used for validation, whether Internal, External, or Both.