1. Consider Adoption of its June 1 Meeting Minutes

Attachment 1

—Commissioner Nathan Houdek (WI)

Draft: 6/10/26

Big Data and Artificial Intelligence (H) Working Group
Virtual Meeting
June 1, 2026

The Big Data and Artificial Intelligence (H) Working Group of the Innovation, Cybersecurity, and Technology (H) Committee met June 1, 2026. The following Working Group members participated: Nathan Houdek, Chair, Timothy Cornelius, and Coral Manning (WI); Doug Ommen, Co-Vice Chair (IA); Mary Block, Co-Vice Chair (VT); Alex Romero (AK); Mark Fowler (AL); Lori Drever Munn (AZ); Ken Allen (CA); Jason Lapham (CO); Wanchin Chou and George Bradner (CT); Karima M. Woods (DC); Ron Waye (FL); Shannon Hohl (ID); Ryan Gillespie and Ricardo Mancilla (IL); Shaun Orme (KY); Caleb Huntington (MA); Raymond Guzman (MD); Sandra Darby (ME); Kate Stojsih (MI); Phil Vigliaturo (MN); Brad Gerling (MO); Colton Schulz (ND); Connie Van Slyke (NE); Christian Citarella (NH); Randall Currier (NJ); Gennady Stolyarov II (NV); Nishtha Ram (NY); Matt Walsh (OH); Nicole Nash (OK); David Dahl (OR); Michael Humphreys and Joe Schwartz (PA); Elizabeth Kelleher Dwyer (RI); Andreea Savu (SC); Travis Jordan (SD); Michael Schulz (TN); Nicole Elliott (TX); Eric Lowe (VA); Curt Kwak (WA); and Lela D. Ladd (WY).

1. Adopted its Spring National Meeting Minutes

Stolyarov noted two corrections to the Spring National Meeting minutes.

Colton Schulz made a motion, seconded by Stolyarov, to adopt the Working Group's March 24 minutes (*see NAIC Proceedings – Spring 2026, Innovation, Cybersecurity, and Technology (H) Committee, Attachment One*) with the two corrections proposed by Stolyarov. The motion passed unanimously.

2. Received an Update on the AI Systems Evaluation Tool Pilot Process

Commissioner Houdek noted that the Working Group would provide interim updates between the Spring National Meeting and Summer National Meeting to keep stakeholders apprised of progress.

Manning stated that the pilot officially started in March 2026 and involves 12 participating states. All states have sent, or are in the process of sending, inquiries to their respective companies. The group of pilot state regulators continues to hold weekly calls to share insights on anticipated responses and has received multiple training sessions on data science, compliance, and governance concepts related to the tool. Training will continue throughout the summer. Pilot state regulators have largely identified the companies that will receive inquiries and, in many cases, have already notified those companies. When a company is part of a larger group, pilot states coordinate to avoid duplicate requests.

Manning said that participating states are moving at varying paces. Some have already sent Exhibit A and/or Exhibit B and are reviewing responses, and at least one state has progressed to Exhibit C. Other states have not yet sent inquiries, as they are synchronizing the exhibits with examination timing. Implementation methods include formal examinations, separate surveys or data calls, or a combination of both. A survey has been developed and will soon be deployed to all participating carriers, with responses open through Sept. 30. Two more draft version of the tool will be exposed for public comment prior to the Fall National Meeting.

Manning noted that key areas identified for revision include tightening definitions, including clarifying the distinction between "AI program" and "governance program" in Exhibit A to ensure consistency with the *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers*; tightening materiality language; and clarifying

© 2026 National Association of Insurance Commissioners

Exhibit A information requests. The tool has been effective in allowing states to tailor their use to individual companies. She outlined the timeline for the remainder of the pilot process through the Fall National Meeting. The pilot will continue through September, with official pilot-state feedback closing after that. Company surveys will be open through Sept. 30. Tool version 5.0 will be drafted using feedback received through Aug. 1 and introduced at an end-of-August Working Group meeting, then exposed for a 30-day public comment period throughout September. In October, the Working Group will review comments on version 5.0, pull company survey responses through September 30, and draft version 6.0, which will be exposed for 14 days. In early November, quick revisions will be made, and a Working Group meeting will be held the week before the Fall National Meeting to address comments received on version 6.0. Version 7.0 will be presented at the Fall National Meeting for adoption consideration. An additional Working Group meeting with actuarial panelists is scheduled for July 22.

Lindsey Stephani (National Association of Mutual Insurance Companies—NAMIC) commented that industry is receiving similar feedback that definitions could use additional clarity, the scope of the tool should be tightened, and that it is encouraging that materiality is being considered. She asked whether the Working Group is examining the definition of AI and whether the potential inclusion of predictive or rules-based models is within scope. Manning responded that the pilot group is currently in the collection phase, gathering comments from individual states before conducting broader internal discussions. She indicated that specific definitional feedback submitted in writing would be considered in the next draft.

3. Heard a Panel Discussion on AI Governance Trends

Commissioner Houdek welcomed panelists Michael David Ruiz (Deloitte), Fred Karlinsky (Greenberg Traurig), and Mary Jane Wilson-Bilik (Eversheds Sutherland) and moderated a discussion on AI governance trends.

In response to a question about AI governance practices that are working well, Wilson-Bilik stated that companies are making meaningful progress in building governance and risk management programs that incorporate AI. She identified the following effective practices: establishing board- and senior-management-level AI accountability structures; developing strong data governance policies to ensure data is accurate, representative, and free from bias; monitoring outcomes for bias and hallucinations; and preparing AI inventories. She is seeing significant success in developing cross-functional governance teams and integrating AI into compliance and vendor governance programs. She noted that addressing human-in-the-loop requirements, particularly in the context of agentic AI, remains a difficult and emerging challenge.

Ruiz echoed the importance of cross-functional governance teams, noting that effective governance brings together stakeholders from three lines of defense, where operations, IT, technology, and data are brought to the table at the same time as the second line of defense of risk and compliance and legal, and then internal audit as the third line of defense. He stated that a good starting point is to establish a clear AI vision and philosophy from the top of an organization and translate them into policies aligned with the NAIC AI principles of transparency, explainability, security, and responsible use.

Karlinsky stated that lessons from the cyber governance space, including board-level ownership, cross-functional engagement, and treating AI as a transformational business issue, are being successfully applied in the AI context. These are transformational issues for companies, and they need to ensure they are addressed at the highest level and involve a broad cross-section of people working in the space.

He noted that the NAIC AI model bulletin, adopted by 25 states, provides a strong foundation for governance. He identified risk-based guidance, transparency, and explainability requirements, human-in-the-loop requirements, and proactive bias testing with annual attestations (as implemented in Colorado and New York) as emerging best

© 2026 National Association of Insurance Commissioners

practices. He stated that the most effective governance programs create guardrails without gridlock, enabling innovation while protecting consumers.

Commissioner Houdek asked about areas where the language in the model bulletin or the AI Systems Evaluation Tool may be unclear or difficult to apply. Karlinsky stated that the NAIC's definition of AI has drawn comment from some in the industry, with some preferring the National Institute of Standards and Technology (NIST) definition. He also pointed out some confusion about the tool's context and use in a market conduct examination. He noted that while the tool is a solid starting point, technology regularly outpaces regulation and legislation, and that continued iteration, developed through regulator and industry collaboration, is necessary.

Wilson-Bilik stated that she has heard from companies that some questions included in the tool for financial examinations appear more oriented toward market conduct or consumer protection, raising the question of whether a single tool can serve both purposes. She identified materiality and the sometimes interchangeable use of the terms "model" and "system" as areas requiring definitional clarity. She also stated that the growing use of agentic AI raises significant new regulatory concerns, including AI agents exceeding their authority, cascading errors in downstream systems, real-time data governance challenges, and new cybersecurity risks. She suggested that both the model bulletin and the tool may need to specifically address agentic AI, noting that the bulletin provides a solid foundation but does not yet account for the systemic risks posed by agents that can operate beyond the boundaries of a single underwriting or marketing tool.

Ruiz agreed, stating that a comprehensive AI inventory should distinguish among generative AI, agentic AI, and simpler rules-based tools, and that a risk-based tiering approach applied consistently across the full AI lifecycle is the appropriate framework.

Karlinsky agreed that this would be a good first step, but companies may not be able to identify all the AI models. However, the industry is collectively gaining experience over time. The important part is that the industry is working more closely with the regulatory community to benefit consumers.

Commissioner Houdek asked what factors make a model or AI system high risk. Wilson-Bilik identified the following: the intended use of the model; the sensitivity and accuracy of the data on which it is trained and which it analyzes; the model's ability to interpret context (particularly in chatbot applications); the model's potential adverse consumer impacts; and potential downstream and cascading systemic effects. She mentioned the recent Mythos model developed by Anthropic as an example of the power and danger of applications of AI.

Karlinsky cited vendor risk, model drift, proxy discrimination, black-box models, data governance, and velocity mismatch as key risk-focus considerations, noting that many are analogous to issues previously encountered in credit scoring and cyber regulation.

Ruiz stated that insurance companies are implementing risk-tiering models across the full AI lifecycle, from use-case identification through deployment and ongoing monitoring, treating the initial risk rating as the "building permit" for a model's development and continued operation.

Eric Ellsworth (NAIC Consumer Representative) raised the challenge from a consumer perspective of fragmented accountability across business units and vendors, i.e., the "nobody home" issue, where consumers cannot easily reach a person, such that accountability is fragmented across business units and vendors, with no clear person having enough knowledge of what the system is doing to solve their problem. He asked the panel to discuss maintaining organizational knowledge of process steps and gears, evaluation criteria, and availability of an empowered human to solve problems, as models and agents are adopted.

© 2026 National Association of Insurance Commissioners

Wilson-Bilik referenced the Colorado AI Act (SB 26-189), which requires companies to designate a trained individual with the authority to override consequential automated decisions, and noted that this type of "meaningful human review" standard (defined as someone who is trained in the workings of the of the actual automatic decision making technology [ADMT]) could serve as a model for addressing consumer accountability.

Karlinsky acknowledged that the ultimate goal is for models to perform well enough that continuous human intervention is not required, while cautioning that it is premature to assume that point has been reached.

Ruiz stated that tracking the right key performance indicators as model capabilities evolve and reporting those metrics to governance stakeholders is essential.

Stolyarov asked whether insurance organizations have implemented sandbox or test environments for AI agents that can operate but cannot permanently or irreversibly alter any system, and then, when the AI agent provides its recommendation, it is escalated to a human reviewer for action. The consequences of the AI agent in this environment would not flow through to the organization, the customers, or the general public unless a human being says this is an acceptable result.

Ruiz replied that he has not yet seen a seamlessly deployed test-to-production human-in-the-loop interaction in practice, but noted that the concept is actively developing and that the proliferation of agentic AI will demand its further formalization within governance frameworks.

Karlinsky commented that he has seen a lot of testing to get the right answers using AI, but believes it is still in its early stages and that there will be more testing. Wilson-Bilik stated that rigorous testing prior to production deployment is essential and represents sound governance practice, but that it is still too early to be able to know how this will play out.

Miguel Romero (NAIC) asked the panelists to confirm his understanding that the way a model is used and the data underpinning it are the foremost drivers of model risk. Wilson-Bilik agreed, adding that potential consumer outcomes and financial impacts must also be incorporated into the risk assessment, including the model's training. Ruiz agreed, adding that the criticality of the use case and the breadth of affected stakeholders, particularly where model outcomes directly impact consumers in areas such as underwriting and claims decisions, are equally important dimensions of model risk. Karlinsky agreed that consumer impact is key, but model drift and how models are applied are equally important.

Ellsworth commented that the most important thing for consumers is being able to accomplish the tasks they are interacting with the company to accomplish, and often that involves crossing many system boundaries. He asked whether the panelists had heard of any emphasis on testing types for accountability across systems, because the natural breakdown is for someone to own each one and someone to think about each one. Karlinsky replied that a cross-disciplinary approach is something he has been seeing for a while. Ruiz added that comprehensive model inventories should document system dependencies, interconnectedness, and cross-stream and downstream impacts, with designated model owners and stakeholders responsible for those dependencies.

4. Discussed Other Matters

Commissioner Houdek noted that overall engagement from pilot state regulators, companies, and industry throughout the pilot process has been very positive. Romero confirmed that the next Working Group meeting is scheduled for Wednesday, July 22, 2026, at 11:00 a.m. CT and is posted on the NAIC website. The July 22 meeting

© 2026 National Association of Insurance Commissioners

will include an update on the AI systems evaluation tool pilot process and a panel discussion with actuarial panelists.

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

Member Meetings/H Cmte/2026_Summer/WG-BDAI/ 2026 0601Interim-Meeting/Minutes-BDAIWG060126-Final.docx

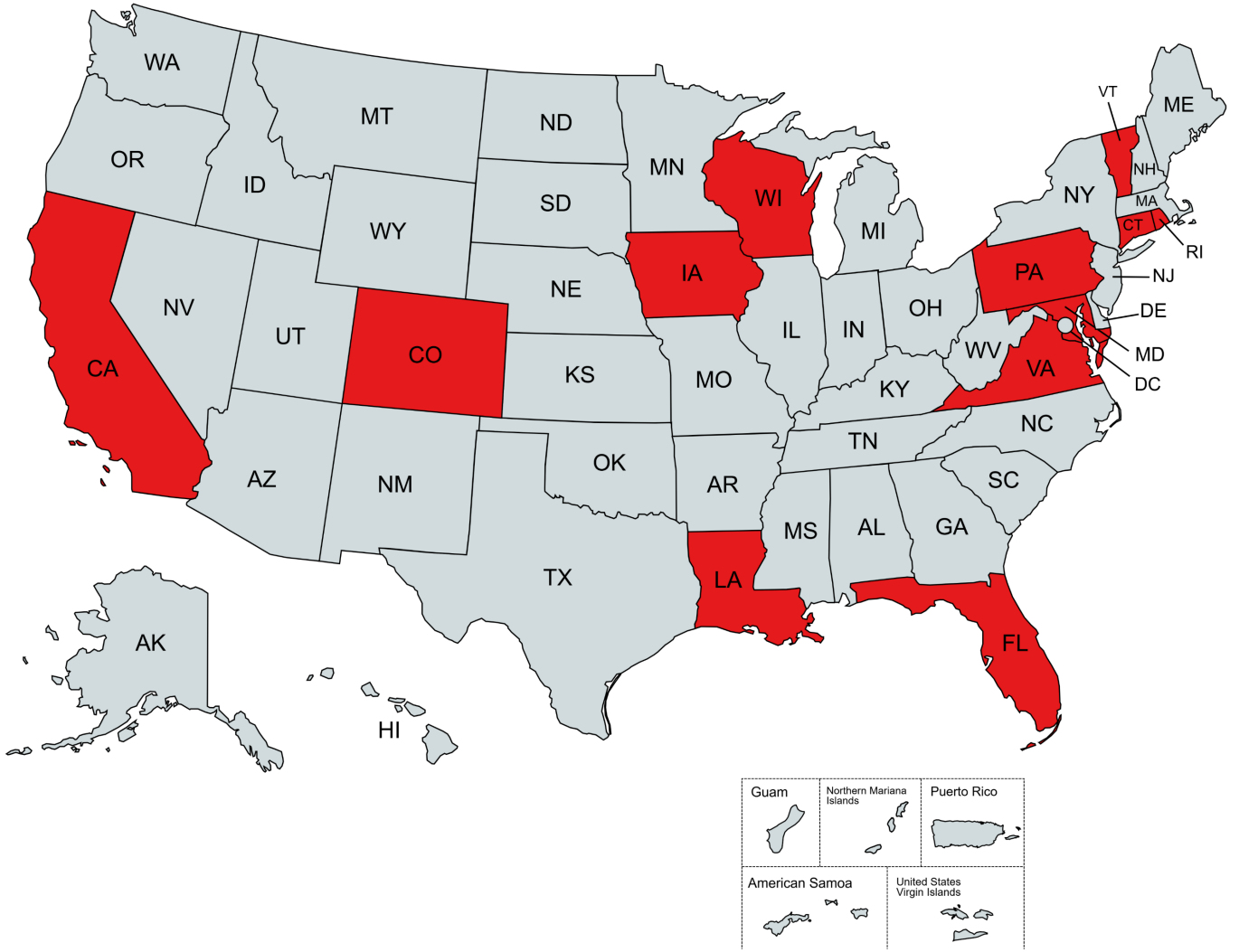
© 2026 National Association of Insurance Commissioners

2. Provide an Update on the AI Systems Evaluation Tool Pilot Process

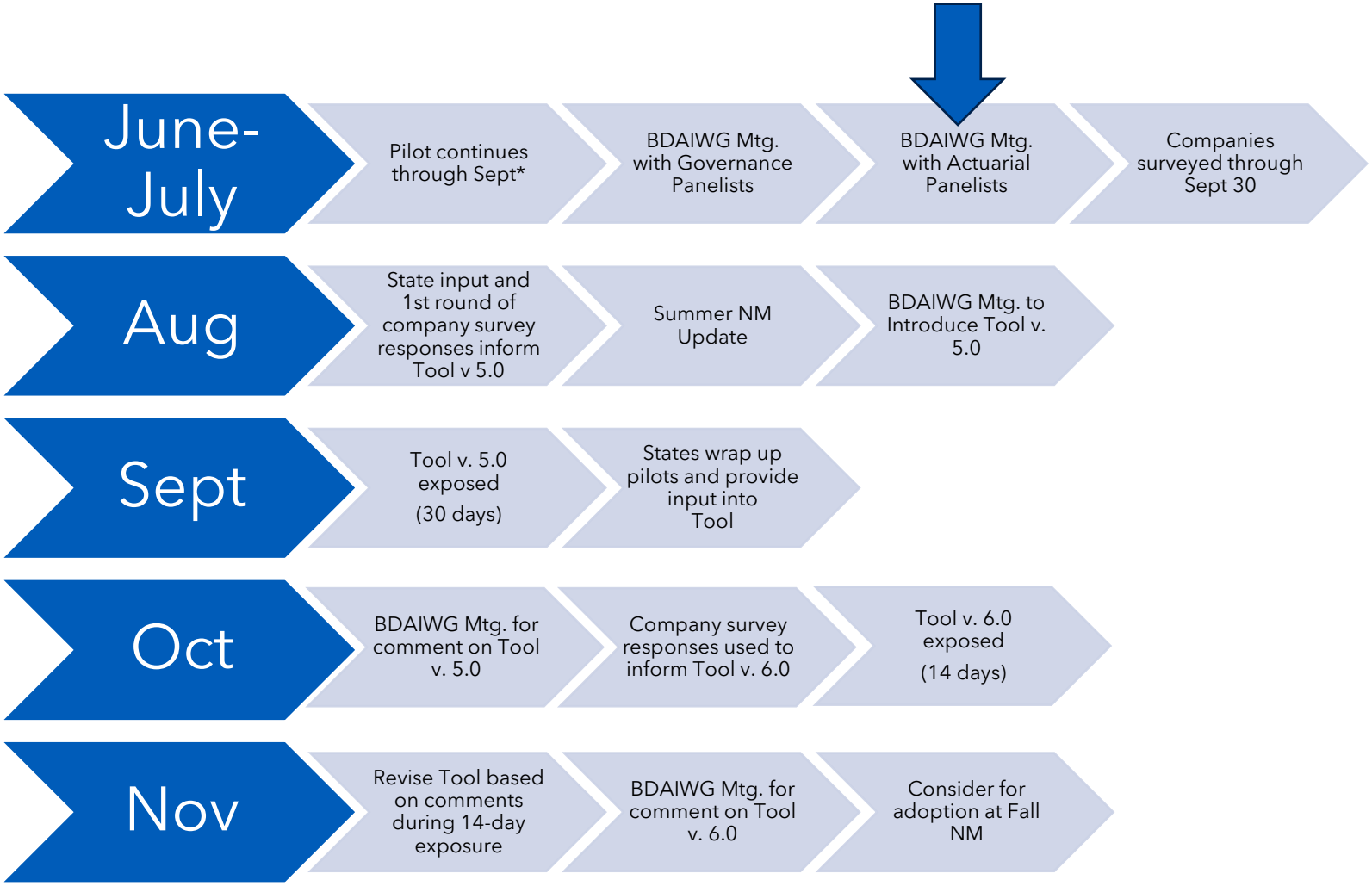
–Coral Manning (WI)

Pilot Participation States

1. California
2. Colorado
3. Connecticut
4. Florida
5. Iowa
6. Louisiana
7. Maryland
8. Pennsylvania
9. Rhode Island
10. Vermont
11. Virginia
12. Wisconsin



AI Systems Evaluation Tool Pilot—Timeline



*Timing for each state's pilot will vary and every pilot may not be completed by 9/30/26. Information gleaned from the pilot experience will inform Tool drafts through adoption and future iterations of the Tool.

3. Hear a Panel Discussion on Perspectives from Actuarial Organizations on Generalized Linear Models (GLMs) and AI Governance

–Dale Hall (Society of Actuaries–SOA)
–Henry Liu (Casualty Actuarial Society–CAS)
–Spencer Sadkin (American Academy of Actuaries–AAA)