Draft: 8/23/26

Big Data and Artificial Intelligence (H) Working Group
Columbus, Ohio
August 13, 2026

The Big Data and Artificial Intelligence (H) Working Group of the Innovation, Cybersecurity, and Technology (H) Committee met in Columbus, OH, Aug. 13, 2026. The following Working Group members participated: Nathan Houdek, Chair, and Coral Manning (WI); Doug Ommen, Co-Vice Chair (IA); Mary Block, Co-Vice Chair (VT); Richard Fiore (AL); Alex Romero (AK); Lori Munn (AZ); Chris Erwin (AR); Ken Allen (CA); Jason Lapham (CO); Joshua Hershman (CT); Nicole Crockett (FL); Matt Kilgallen (GA); Weston Trexler (ID); Ryan Gillespie (IL); Jake Vermeulen (IN); Saima Sanjida Shila (LA); Sandra Darby (ME); Mary Kwei (MD); Caleb Huntington (MA); Kate Stojsih (MI); Phil Vigliaturo (MN); Brad Gerling (MO); Connie Van Slyke (NE); Gennady Stolyarov II (NV); Sean Rosene (NH); Dave Wolf (NJ); Vanessa DeJesus (NM); Angela Hatchell (NC); Colton Schulz (ND); Matt Walsh (OH); Nicole Nash (OK); Gary Jones and Michael McKinney (PA); Andreea Savu (SC); Tony Dorschner (SD); Michael Schulz (TN); Marianne Baker (TX); Eric Lowe (VA); Ellen Potter (WV); and Lela D. Ladd (WY).

1. Adopted its July 22 Minutes

The Working Group met July 22 and took the following action: 1) received an update on the Artificial Intelligence (AI) Risk Evaluation Supplement pilot process and 2) heard a panel discussion on perspectives from actuarial organizations on generalized linear models (GLMs) and AI governance.

Stolyarov made a motion, seconded by Trexler, to adopt the Working Group's July 22 minutes (Attachment One-A). The motion passed unanimously.

2. Received an Update on the AI Risk Evaluation Supplement Pilot

Commissioner Houdek said the 12 states participating in the pilot have been engaged since March and have met nearly weekly to discuss their individual state experiences implementing the pilot. He said that, based on feedback received during the pilot and input from subject matter experts, the supplement will be revised with an updated version ready for a public exposure comment period in early September.

He said each participating state has been engaged in the work over the spring and summer, and that some states used the supplement as part of a market conduct or financial examination that was already planned, while others used it as an ad hoc questionnaire outside the typical examination process. Each state has the flexibility to modify or add questions to meet its specific needs. The pilot will continue through September, and feedback has been collected from companies involved in the pilot, with each participating state also reaching out to its pilot companies to solicit input through a centralized survey of the experience. Written feedback has been received, which will help inform revisions to the supplement. Opportunities for verbal feedback will be provided during the public meetings following the two exposure periods planned for this fall, with the goal of finalizing the supplement and adopting the final version at the Fall National Meeting.

Crockett commented that the pilot process has been effective due to the continued engagement and strong collaboration between the states.

 

3. Heard a Presentation from AM Best on AI Governance in Insurance

Edin Imsirovic (AM Best) presented on AI governance in insurance, stating that the presentation was for discussion purposes and does not change AM Best's criteria, methodology, or rating guidance, nor does it create a formal AI governance assessment framework. He said the discussion on AI governance draws on public regulatory and supervisory standards, model risk, and AI security measures to address how governance expectations may change as AI systems move from predictive to generative to action-oriented workflows. He said the NAIC *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers* and the AI Risk Evaluation Supplement pilot provide a foundation for a company's AI governance policy toward how governance operates.

Imsirovic defined three working terms: 1) predictive AI, which uses data to score, rank, classify, or flag something, often used in pricing models or claims triage, usually supporting a decision inside a well-defined process; 2) generative AI, which creates new content, such as summarizing documents, drafting language, extracting details, or turning unstructured data into something more useful, but its output can sound confident when it is wrong; and 3) agentic AI, which goes a step further and can work through a series of steps toward a goal, use tools, look up information, update a system, route work, or move a workflow forward. This system may affect processes directly rather than support a human decision. Each creates a different governance problem.

Imsirovic said AI does not create a separate risk category, as the exposures already flow through existing categories, such as product and underwriting, operational, and regulatory and legal risk, and that the enterprise risk management (ERM) question remains whether risk management capability is commensurate with the risk profile.

He said that well-governed AI deployment may support outcomes such as expense and processing improvement, underwriting quality, claims efficiency and consistency, and fraud control, while risks of execution, control breakdowns, cyber and data exposure, vendor dependency, and regulatory and explainability exposure can arise from deployment. He said what matters is whether governance mitigates the risks.

He said that if AI affects pricing, risk selection, segmentation, or portfolio mix, that points to product and underwriting risk. Data quality, claims-handling errors, cyber exposure, workflow breakdowns, or weak oversight of third-party models fall under operational risk and consumer protection. Explainability, litigation, or regulatory scrutiny can flow through the legislative, regulatory, judicial, and economic categories. One use case can span several categories; for example, a claims AI tool that begins as an operational issue becomes a regulatory issue if consumers are treated inconsistently and creates financial consequences if errors affect payments or reserve information. He said the categories are familiar, but the path into them is changing rapidly, as carriers move from predictive models to generative models to agentic actions. As a result, risks can emerge faster, become harder to understand, and move further without human intervention.

He described the development of AI capabilities as a compounding progression rather than a timeline. He said that using AI for fraud detection and pricing, or for computer vision and language processing used in claims and document work, was foundational, and that the governance problem at this stage remained closer to traditional data and model governance. He said that generative AI changed that, as those systems work with less structured information, produce less predictable output, and may be harder to explain, presenting a different governance problem rather than a harder version of an old one. He said that agentic AI changed it again, as those systems can act rather than only recommend, and one error can trigger others and move through a workflow before a person can intervene; therefore, a single governance approach is not enough.

He noted that predictive AI raises familiar questions about whether the model was properly tested, is fair, is documented, and whether its performance is changing over time. Generative AI raises questions about whether the output can be trusted, whether sensitive information stays where it should, and what happens when the underlying model is updated. Agentic AI raises questions about the system's access, what gets recorded, where a person steps in, and whether an action can be stopped or reversed. The risks build, such that an agentic system may still carry predictive and generative risks in addition to new permission and action risks, and that the comparison should be read as a comparison of governance problems rather than as a simple risk ranking.

He said that for predictive AI, monitoring should lead to action, so that when a model drifts or produces a weak outcome, the natural thing to look for is whether the company adjusted a threshold, limited the model's use, or added human review, and a dashboard that never changes anything is weak evidence. For generative AI, the question is whether controls hold up through the live workflow, including whether the level of review increases when the output matters more, whether the company tests the system again when a vendor updates the model, and whether the company can show that sensitive information is secured. For agentic AI, the evidence has to be more concrete, including whether permissions are tested, whether actions are logged for company reconstruction, whether there is a kill switch to roll the process back to a safe state, and whether a person has the authority to override actions. He said two companies can have the same policy but very different governance in practice.

He said the predictive, generative, and agentic labels do not indicate how much is at stake. He identified five factors: 1) how important the decision is, including whether the output directly affects a consumer or a financial outcome; 2) how much the system can do on its own; 3) whether the company can explain why the system produced a particular result or took a particular action; 4) how quickly the system can change, whether gradually through drift or suddenly through a vendor update; and 5) how dependent the insurer is on a third party, including whether the company can independently test, challenge, or override the system. Underlying these factors is proportionality, in that governance around a system should become stronger as stakes, autonomy, lack of transparency, speed of change, or third-party dependence increase, and he said this is consistent with the NAIC *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers* and with recent work from the European Insurance and Occupational Pensions Authority (EIOPA), which also takes a risk-based and proportionate approach to AI governance in insurance. Thus, predictive AI is not inherently lower risk, and agentic AI is not automatically unacceptable. The label indicates the type of system, while the use case indicates more about the level of risk and the governance it may need.

He said lower-impact tools that mainly assist a person, such as an underwriting research assistant, still need governance, but the likely consequences of an error are more limited, and a person remains close to the process. Where the system is still assisting a person but the output matters more, such as extracting information from medical records for underwriting or preparing claim summaries for adjusters, the review has to be meaningful. This raises the question of whether the reviewer has enough information, time, and authority to question the result rather than accept it. He said that where the system has more autonomy but the impact is operational, such as narrowly defined automated endorsement processing, the questions become whether the boundaries are clear, whether actions are logged, and whether errors can be contained and reversed. He said the most demanding governance zone combines greater autonomy with more important consumer or financial outcomes, such as an agentic system routing and paying claims or making underwriting decisions. Controls such as human override, action logging, rollback, and clear limits on the system's actions become especially important.

He said controls are built in layers, with earlier controls staying in place as each layer is added. Predictive AI governance begins with model risk controls, including independent validation. Revalidation triggers mean deciding in advance which kinds of changes should lead to another review or testing, such as changes to a predictive model

or its data. For generative AI, a company should know what prompts are used; otherwise, the output will be harder to explain. For agentic AI, the level of approval should depend on the action, as looking up information is not the same as changing coverage or issuing payment. Some controls depend on vendor contracts, such as testing rights and notice of important updates, which can be much harder to obtain.

He said that where governance lives can make a difference, as two companies can use the same AI tool and have similar controls but end up with very different levels of risk in how the tool is integrated into the business. Governance should be built into the process. Fragile tool layering occurs when a tool is added to an existing process without changing the workflow or its accountability. He said companies do not choose that setup deliberately but arrive at it gradually. As a pilot works, more people begin using it, the use case spreads, and only later does the company realize that the process around the tool never caught up. A well-designed predictive model in a weak operating environment can create more risk than a more advanced AI system that is properly integrated and controlled.

He said shared dependency is broader than reliance on a single vendor, as many carriers may depend on the same vendors, models, data sources, cloud providers, or other infrastructure, and noted that the Financial Stability Board (FSB) has highlighted third-party dependency and provider concentration as potential financial stability concerns. For a single carrier the issue comes down to three questions: 1) visibility, whether the company understands how the system behaves; 2) control, whether the company can test and challenge the system, be notified when something important changes, and step in when needed; and 3) portability, whether it is difficult to replace or unwind a tool once it becomes part of an important workflow. A vendor change can affect how a system behaves, even if the carrier makes no changes, and risks can compound if the company does not review it. The larger point is at the market level, as each carrier may see only its own dependency, but if many carriers rely on the same underlying provider, a single weakness can affect many companies at the same time. This does not mean third-party use is a problem, as vendors can offer useful tools, but the key is whether insurers can understand what the system is doing and have the authority to oversee it and step in when something changes. The issue is whether oversight works effectively.

He identified four areas where a company should be able to show more than a policy or framework: 1) AI inventory, where knowing which decisions each system touches and how much it can do on its own shows which systems need the most attention; 2) validation and monitoring, where the question is what happens when a trigger is hit, including any tests and approvals performed before the system went live and what the company did when monitoring later found a problem; 3) human oversight, where a company should be able to show an example in which a person challenged the result, changed it, or stopped the process; 4) vendor and agentic controls, where the question is whether the company can see a vendor change, test its effect, stop the system, or return to a safe version. Policy alone is not enough; rather, operational governance is necessary.

He said the question is whether the use of AI stays bounded, monitorable, governable, and resilient as systems change, take on more autonomy, and become harder for people to oversee. There is still no single AI rulebook, but regulators want to know what AI is being used for, what decisions it affects, how it was tested, who is accountable, and evidence showing that controls actually work. He said that the last point represents the real shift, as much of the earlier work was principles-based, while newer work is moving toward implementation and examination.

He said modern AI systems may include the model, the prompt, the information it retrieves, memory, tools, and permissions, so the control boundary becomes much larger than the model itself. He said the UK AI Security Institute (AISI), working with the Bank of England, examined what developers are building rather than only testing models in a laboratory, reviewing more than 177,000 agent tools built over the preceding 16 months. Two findings

stood out: 1) tools that take action by changing something in an external system grew from about one-quarter of usage to 65% of usage over that period; and 2) while most action tools support medium-stakes work, such as editing a file (finance was an exception), with high-stakes financial roles having more tools that can change things than would be expected, including growth in the number of servers that can execute payments from about 46 to more than 1,200 in a single year. He said memory fits within that same inventory question, raising data governance questions about what is kept when a system carries information from one task to the next, how long it is kept, and whether it can be corrected or deleted when a policyholder asks, with the difference being that the data now sits inside the system rather than beside it, so identity, delegated authority, the ability to revoke access, and what the system remembers may all become core governance questions rather than technical details.

He said that capability can change even when the model does not. For example, the AISI tested frontier models on multistep cyberattack scenarios and increased the compute budget from 10 million tokens to 100 million tokens, which improved performance by 59% with no retraining, additional fine-tuning, or special expertise. This means a company could keep the same model and still materially change the risk by changing what surrounds it, whether through more time, more attempts, better tools, or wider permissions, such that approving the model may not be the same as approving the deployment.

Commissioner Houdek asked where companies struggle the most in governance. Imsirovic said that one of the biggest areas for improvement is building or revising the operating model around the use of AI, as too many companies use various tools without redesigning the operating model.

Block commented that agentic AI seems like a transition from human to AI models, back to another human being doing the work. Because agentic AI operates independently, it feels like a continuum in which supervision of a nonhuman employee is required. She appreciated the discussion of having a plan for a quick kill switch, for what happens when a system fails, and for the process to shut a system down, and that she views this as part of the ERM disaster recovery process to be treated the same way as planning for when a key employee disappears. Imsirovic agreed and said that agentic AI is not as widespread as generative AI at the moment, in part because companies are cautious about deploying agentic systems due to operational and regulatory concerns. He said most of the deployments he sees are in lower-stakes roles, such as a frequently asked questions (FAQs) chatbot or research assistance, and that the controls companies most often discuss involve keeping a human in the loop throughout the process and ensuring the AI system is monitored and tested, with the caveat that agentic AI remains mostly in early pilot stages.

Commissioner Ommen said a number of Working Group members are receiving different responses through that structured effort. He asked whether Imsirovic is beginning to see policies and procedures standardized in writing or whether there is still a great deal of communication through meetings. Imsirovic responded that the amount of capability an AI system has draws a different amount of scrutiny within insurance companies, noting that companies typically tell him they do not use AI for underwriting or, if they do, that it is monitored, and that in underwriting, it is usually used to turn unstructured data into structured data to enrich the submission process. Another area is in the claims process, helping adjusters with claims summaries, and that the most consistent answer he receives to governance questions is that a human is in the loop. Commissioner Ommen asked whether that response comes in conversation or whether written procedures are beginning to appear. Imsirovic responded that most of it comes through conversations, as AM Best does not have an official process for gathering data on AI governance, although that may change. He said some companies volunteer more information than others and that most are keen to describe how AI will improve their processes, but he has not seen a great deal of that documentation.

Stolyarov asked whether Imsirovic is seeing insurers adopt a consequential approach to AI governance in which the impact of a particular failure is considered not only for a consumer but for the enterprise as a whole. He illustrated the distinction with a chatbot that interacts with consumers and hallucinates information, which could mislead consumers and draw regulatory concern but is fundamentally correctable because a person at the insurer could intercede and provide the correct information, compared with an AI agent that, in an effort to be helpful, proactively and irreversibly deletes the company's claims information to free storage space, which is a failure from which the enterprise as a whole could not recover. The member cited the July 2024 CrowdStrike incident, in which a new employee implemented a systemwide update that disabled the systems of airlines and other enterprise clients, as the kind of catastrophic failure governance needs to prevent, and asked whether insurers are prioritizing those systemwide consequences. Imsirovic responded that the level of sophistication depends in part on the sophistication of a company's use of AI systems, as companies that entered the space earlier had more time to develop proper governance frameworks for those tools and tend to be higher-rated with well-established ERM frameworks. He said his concern is with companies that are newer to the technology and lack that history, which may implement it without having thought it through.

Commissioner Hershman noted Imsirovic had identified three types of AI, that those systems are leveraged for different use cases carrying different risks to the consumer or the company, and that different governance structures should surround those different types of systems and risks. He asked how the Working Group should think about the different levels of governance and how governance tools should be distinguished, such as by high risk versus low risk or by the type of AI being used. Imsirovic responded he would look for a pattern rather than one document, including a useful AI inventory, clear testing and approval standards, monitoring that leads to action, examples in which human review changed the outcome, evidence that the company can respond when vendor systems change, and whether the company is changing its operating model around the technology. He said the answer returns to the impact chart and the five factors: The higher the impact of a particular tool, the more governance that tool will need. He said an agentic tool used for narrowly defined endorsement processing may not require as much governance as the same type of tool used to pay claims or to automate underwriting, so the use case and what is at stake should determine the level of governance around a particular technology.

4. Discussed Other Matters

Commissioner Houdek reminded participants that an updated version of the supplement will be exposed for a public comment period at the beginning of September.

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

NAIC Support Staff Hub/Member Meetings/H Cmte/2026_Summer/WG-BDAI/Minutes-BDAIWG081326-Final.docx

 