Draft: 7/30/25

Big Data and Artificial Intelligence (H) Working Group
Virtual Meeting
July 16, 2025

The Big Data and Artificial Intelligence (H) Working Group of the Innovation, Cybersecurity, and Technology (H) Committee met July 16, 2025. The following Working Group members participated: Michael Humphreys, Chair, Shannen Logue, and Michael McKenney (PA); Doug Ommen, Co-Vice Chair (IA); Mary Block, Co-Vice Chair (VT); Alex Romero and Molly Nollette (AK); Tom Zuppan (AZ); Ken Allen (CA); Jason Lapham (CO); Wanchin Chou (CT); Omar Barakat and Yohaness Negash (DC); Rebecca Smid (FL); Renee Iverson (ID); Jack Engle (IL); Victoria Hastings (IN); Nathan Strebeck (LA); Caleb Huntington (MA); Marie Grant (MD); Sandra Darby (ME); Jeff Hayden (MI); Phil Vigliaturo (MN); Brad Gerling and Patrick Lennon (MO); Colton Mork (ND); Connie Van Slyke (NE); Christian Citarella (NH); Brandon Rocchio (NV); Kevin Yan (NY); Judith L. French and Matt Walsh (OH); Teresa Green (OK); John Haworth (OR); Matt Gendron (RI); Andreea Savu (SC); Emily Marsh (TN); Rachel Cloyd (TX); Eric Lowe and Michael Peterson (VA); Timothy Cornelius (WI); and Joylynn Fix (WV).

1. Adopted its Spring National Meeting Minutes

Block made a motion, seconded by Lowe, to adopt the Working Group's March 25, minutes (*see NAIC Proceedings – Spring 2025, Innovation, Cybersecurity, and Technology (H) Committee, Attachment One*) The motion passed unanimously.

2. Discussed the AI Systems Evaluation Tool

Commissioner Humphreys stated that the NAIC issued the *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers* in December 2023, and to date, 24 states have adopted the bulletin or pursued legislation, and an additional four states have adopted related activity. It was announced at the Spring National Meeting that the Working Group began having discussions on how to assess the risks associated with the use of artificial intelligence (AI). As for the next steps after the adoption of the bulletin, members of the working group have been working on a draft AI systems evaluation tool that provides regulators with an immediate resource for examining AI systems. The tool will provide regulators with an efficient and standardized data collection tool that can be used to assess and evaluate risk, along with providing guidance on what regulators can expect when using AI for insurer operations.

The tool is structured into four exhibits that could be incorporated into market and financial exams, while more permanent solutions are being developed. The tool allows regulators to progressively investigate AI governance, testing protocols, data sources, and financial implications while simultaneously serving as a checklist for insurers. Each exhibit is optional, offered to help regulators assess risk as they see fit.

Commissioner Ommen said the AI systems evaluation tool was exposed on July 7 for a 30-day public comment period ending Aug. 6. The Working Group is seeking feedback on the draft from interested parties, with the expectation that states will pilot the tool and provide insights on long-term regulatory solutions. The Working Group encourages interested parties to contribute to the discussion.

Block stated that in 2025, the Working Group will continue working on this tool and other tools to help evaluate insurers' use of AI with the goal of creating a standardized data collection tool to aid reviews. The four exhibits are designed to quantify how an entity is using AI systems. Which exhibit to use will depend on the focus of the regulator. The purpose of Exhibit A is to determine the degree of AI model use and the models' purpose and use

cases in order to identify models with higher risk. The questions on Exhibit A include how many models are in use, which ones have direct consumer or financial impact, how many of them are new, whether there were consumer complaints related to the models, and whether the company is planning to implement additional models. For example, insurers should use Exhibit A to identify reputational risk and consumer complaints. Exhibits B and C address the robustness of controls. Insurers can use the checklist in Exhibit B to assess financial risk and the checklist in Exhibit C to evaluate actions taken against the company's use of high-risk AI systems. Exhibit D is focused on the data used in AI models. Insurers can use all of the exhibits to identify adverse consumer outcomes.

Logue said that Exhibit B takes on both a narrative and a checklist format, depending on the preference of the regulator, to understand the level of governance and testing. The questions on Exhibit B include whether a company has established a governance program and the elements within that governance program, such as the responsibilities, assessment of effectiveness, and identification of AI systems that may have a consumer or financial impact, transparency disclosures, monitoring and risk mitigation procedures, and due diligence performed on third-party-provided AI systems.

Logue said that Exhibit C asks for specific information about high-risk models, such as how they were developed and tested, the level of human-in-the-loop involvement, and how they were reviewed for compliance. Exhibit D may be more likely to be used by market conduct in response to a consumer complaint or where there may be concerns about the types and sources of data used to develop AI models. She stated that the comment period will end a few days before the Summer National Meeting and will not be available at that time due to the timing. As for the next steps, the Working Group is looking for feedback and soliciting states to pilot the tool.

Commissioner Humphreys said that the focus of the Working Group's meeting at the Summer National Meeting will be discussing the comments received on the AI model law request for information (RFI). Then, in a later meeting to be scheduled, the Working Group will discuss the comments received on the AI systems evaluation tool.

Lowe clarified that Virginia's feedback on the AI systems evaluation tool was not meant to replace any of the work of the working group's members but was submitted as complementary.

3. Heard a Preliminary Summary of Comments from the RFI on an AI Model Law

Commissioner Humphreys stated that one of the items under consideration is a potential model law to promote consistent use of AI across states. The purpose of the RFI was to solicit feedback from all interested parties. It was exposed for a 45-day comment period, which ended June 30. The RFI gauged stakeholder reactions and asked for other goals to be considered that are not addressed by existing laws and regulations, and what key concepts should be considered if governance requirements should vary. The Working Group received 33 comment letters, which were posted to the Working Group's website. Some central themes included developing a model law with uniformity, consumer protection, and third-party vendors. On the other hand, some organizations expressed concerns or reservations, suggesting that the focus should be on continued adoption of the bulletin and providing guidance and clarity using bulletins rather than pursuing legislation.

Commissioner Humphreys said many states, including Pennsylvania, have developed legislation on AI use, specifically in health insurance; however, comments from a panel that would include health insurance, consumer, and trade group stakeholders could help move the conversation along. The Working Group would like to hear comments and feedback from all stakeholders to determine next steps and looks forward to furthering the conversation in mid-August.

4. Heard a Presentation from Lazarus AI on the Use of Agentic AI in the Insurance Industry

John Keddy (Lazarus AI) introduced his presentation on agentic AI by contrasting it with non-agentic AI. He explained that non-agentic AI includes reading handwriting, extracting complex medical information, models that perform reasoning and logic, creating contracts or documents, gleaning insights from documents, and analyzing legal contracts.

He explained that the big tech players that insurance companies use, such as Salesforce, Microsoft, and NVIDIA, are aggressively pushing agentic AI, which uses a series of AI analytic capabilities to execute processes to create value, reduce risk, reduce costs, increase productivity, and have faster turnaround for the consumer. Further, agentic AI is centered around an orchestrator, like a brain, that conducts AI tools and technologies. He stressed that people are a critical part of agentic AI. Agentic AI is an end-to-end multi-modal input and output process that executes a process and creates value.

Keddy continued with a demonstration of how agentic AI could be used in the claims process. He described a scenario where a car was stolen, and many inputs, such as documents, police reports, and video, were analyzed. Agentic AI can call out to other databases to compare images against an external source. In this case, the agentic AI determined that the claim was legitimate, but the original color of the car was incorrect, and there were inconsistencies in the model year. The expectation is that legacy systems will start to get thinner as the use of agentic AI increases in our economy. However, having a human in the loop is still critical.

Pearce asked about the skill sets needed by humans to ensure the right supervision expertise. Keddy responded that insurance domain knowledge is more important than ever, insurers must have a degree of skepticism, and authentication and technical skills are critical.

Umakant Narkhede (American Red Cross) commented that AI literacy is also a critical skill, and insurance professionals need to understand the risks associated with the models being used. Keddy agreed and added that there are two aspects. The insurance industry must upscale and gain knowledge on AI systems, and it is also incumbent on AI tech providers to provide explanations so that companies can make decisions.

Commissioner Humphreys asked how to prevent an agentic AI system from hallucinating. Keddy confirmed that errors are real concerns, just as humans make mistakes. In the same way that professionals have learned how to manage humans, there is a need to manage AI systems. It is important to understand the choices an AI provider has made and to continue to be active in third-party vendor management.

Commissioner Humphreys asked Keddy to provide written responses to the additional questions due to time constraints (Attachment 1).

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

SharePoint/NAIC Support Staff Hub/Committees/H CMTE/2025_Summer/WG-BDAI/2025 0716 Interim-Meeting_Open/Minutes-BDAIWG071625-Final.docx

NOTE: THE FOLLOWING VIEWS AND OPINIONS ARE EXPRESSED BY JOHN KEDDY (LAZARUS AI) AND ARE SOLELY THEIR OWN AND DO NOT REFLECT THE VIEWS, POSITIONS, OR POLICIES OF THE NAIC. THESE STATEMENTS SHOULD NOT BE CONSTRUED AS REPRESENTING THE STANCE OF THE ORGANIZATION.

To Commissioner Humphreys et al :

It was a pleasure to engage last week. I have provided some brief responses to the questions provided. I am available to discuss in more detail 1:1 or in any other group session. My sole intent is to support NAIC and the important work being done on this critical topic.

Stuart Jones (Florida Office of Insurance Regulation):

- This looks to be an advanced form of generative AI. Fascinating, for sure. But how do you keep the Agentic AI from hallucinating?
- (I believe you answered this one)

COMMENT : We did discuss but I will extend with a few comments here. For over 3 years I have been sharing that one way for people to get their minds around AI is to use HI (Human Intelligence) as a comparison. So for this question, ask "do Humans make mistakes ?" and of course the answer is yes. ("Hallucinations" are just one form of a mistake.)

When we hire a human do we check background and their training/education ? Yes – of course. With AI we need to have roughly analogous questions. Getting a basic understanding of how the models were trained is good but perhaps more importantly see how they apply this training/education.

Do humans need oversight ? Yes – of course. . All employees need to have proper management oversight (even the CEO of a corporation has governance and oversight.) So does AI. We need to have the proper guiderails in place. In the AI world we must think through "human in the loop" requirements early on and re-validate over time.

I could go on but think the analogy is clear. With Agentic AI (which provides us the ability to leverage AI more holistically), the topic of error minimization is ever more important and "human in the loop" will be a key technique.

- Frank Quan (Professor, University of Illinois):
    - Regarding the agentic AI, did your company develop and train the model in-house, or did you rely on a pre-trained version? If the model was trained on real insurance data, could you provide insight into the scale of the dataset and the computational resources involved? Additionally, are carriers generally open to sharing data for training purposes? Lastly, if the model is trained on data from a single insurer, does the agentic AI retain or transfer any insights when deployed for other clients?

COMMENT : Lazarus AI develops all of our own core AI technologies. Our core model training involved a very large corpus of business and medical records. Going beyond that would start to get into proprietary information and I am sure you can understand why we would be hesitant to go too far on this topic (without a business relationship, NDA etc.).

In general, speaking solely for my work at Lazarus AI, Insurers are not comfortable, in July 2025, with handing over internal information to train a model on. At Lazarus AI, as we use our models we do NOT train on any client data and do NOT want to. In my role, I am constantly asked about the potential training from insurer data. Due to the volume of this specific question, I conclude insurers (and others in Financial Services) are very concerned about this topic.

Now if we need to bring in a company's specific acronyms or some other very unique set of facts, we would work with that company on an auxiliary implementation (typically solved by techniques called RAG or VKG). We can discuss this more if of interest but I want to stay focused here on the question at hand.

To your last point, as noted above in my answer we do NOT train on our client's data and we do not retain. I must stress here as I am answering solely for Lazarus and not for other model providers. Other model providers may be training on client data (or using limited fine tuning techniques). Due to this possibility, in my view, every user of AI should be asking your exact questions and with agentic AI coming into the economy, the answers to your questions are ever more important .

- Matt Gendron (Rhode Island Insurance Division):
    - The Agentic AI tool shown seems very valuable. However, I would echo the concern raised by Stuart Jones above. And I wonder how it would adhere to the NAIC AI Guidelines of Fairness, Accountability, Transparency, and Ethical Considerations.

  COMMENT : In my view the work NAIC is doing on Fairness, Accountability, Transparency and Ethical Considerations is ever more important as our economy heads into a world of more AI including Agentic. Each component of AI being utilized by Agentic should be held to reasonable expectations---as-------should the humans.

  Notice in my example of how an insurer might use Agentic, our little demo ended by, essentially, saying, "human—not trying to accuse anyone—but suggest you look much harder at these specific inputs."

  As AI becomes a larger part of our industry, we must not lose sight of the human decisions made in all of this. We need to expect that companies will make different decisions making the work you note here more valuable.

- Peter Kochenburger (Southern University Law Center):
    - How transparent and testable are the assumptions and information used to ultimately generate a recommendation? Historically, some fraud indicators have been discriminatory.

COMMENT : The model that is most common in our current "Go to Market" has explainability of answers in common English (or any other language you would like). This helps users understand model sensitivity (language models are very sensitive to the way you phrase a question and how you provide context). This capability supports companies as they decide where to put "human in the loop."

I will also share that insurers are very intrigued when we show "no" answers. So, e.g., we will ask if there is evidence of a disease from a medical document. The correct answer is "no" but the explainability lets them see exactly what the model was looking for and why the model concluded "no" was the correct answer.

I can not speak for the entire AI industry but I think overall model providers have gotten more sensitive on the topic of transparency in answers. I can remember what seems like ancient days but was only about 36 months ago and I would commonly hear "the models are all a black box" "you can't understand what's going on." I would always

respond with " that's not going to work in heavily regulated industries." I don't hear those dismissive comments as much anymore but that doesn't mean all AI providers have same level of seriousness on this topic. In my view, users need to demand a "lay person" explanation of what is going on and being able to experiment a bit (testing the assumptions you note) so they can get comfortable with how recommendations are made.

Again, it was a pleasure to engage with you all and I found these questions very germane. As noted before I am always available to go deeper 1:1, speak with members of your team or participate in any group setting

Sincerely,

John Keddy, Lazarus AI

A: 33 Arch Street, 12th Floor

Boston, MA 02110

E: john@lazarusai.com

https://calendly.com/john-ked