1. Consider Adoption of its July 22 Meeting Minutes

Attachment A

–Commissioner Nathan Houdek (WI)

NAIC NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

Draft: 8/6/26

Big Data and Artificial Intelligence (H) Working Group
Virtual Meeting
July 22, 2026

The Big Data and Artificial Intelligence (H) Working Group of the Innovation, Cybersecurity, and Technology (H) Committee met on July 22, 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); Alex Romero (AK); Mark Fowler (AL); Tom Zuppan (AZ); Ken Allen (CA); Jason Lapham (CO); Wanchin Chou and George Bradner (CT); Nicole Crockett (FL); Weston Trexler (ID); Victoria Hastings (IN); Shaun Orme (KY); Caleb Huntington (MA); Raymond Guzman (MD); Sandra Darby (ME); Kate Stojsih (MI); Phil Vigliaturo (MN); Brad Gerling (MO); Connie Van Slyke (NE); Sean Rosene (NH); Randall Currier and Howard Wegener (NJ); Vanessa DeJesus (NM); Gennady Stolyarov (NV); Nishtha Ram (NY); Matt Walsh (OH); Nicole Nash (OK); Michael Humphreys (PA); Elizabeth Kelleher Dwyer (RI); Andreea Savu (SC); Haelly Pease (SD); Michael Schulz (TN); Rachel Cloyd (TX); Eric Lowe and Dan Bumpus (VA); and Lela D. Ladd (WY).

1. Adopted its June 1 Minutes

The Working Group met June 1 and took the following action: 1) received an update on the artificial intelligence (AI) systems evaluation tool pilot process and 2) heard a panel discussion on AI governance trends.

Fowler made a motion, seconded by Ladd, to adopt the Working Group's June 1 minutes (Attachment 1). The motion passed unanimously.

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

Manning provided an update on the AI systems evaluation tool pilot process. She said the pilot officially began in March and involves 12 participating states, whose insurance regulators continue to hold weekly calls to learn from each other, raise questions for the group, and receive training and presentations. She said that, in response to feedback from many stakeholders about the name of the tool, it has been renamed the AI Risk Evaluation Supplement to reduce confusion about what the supplement is and is not.

She reviewed the timeline for the remainder of the pilot through the Fall National Meeting. She said the individual state pilots have continued through June and July at varying speeds and that education and discussions have continued. The pilot will continue through September, and companies selected for the pilot should look for opportunities to provide input, including a centralized survey of the pilot experience, with each state coordinating its own invitations to companies individually. The next version of the supplement will be shared at the end of August, followed by two rounds of public exposure to provide ample opportunity for public input, in addition to feedback from participating companies, with the goal of adopting the supplement at the Fall National Meeting.

3. Heard a Panel Discussion on Perspectives from Actuarial Organizations on GLMs and AI Governance

Commissioner Houdek said the Working Group developed a draft version 4.0 of the supplement to help regulators better understand the scope and uses of AI systems and assess the potential risks of those systems as they are used in insurers' operations, including the data used to train them. He said one topic that interested parties and regulators have repeatedly raised is whether generalized linear models (GLMs) should be included or excluded from the scope of the supplement, as GLMs have been used in insurance company ratemaking for decades, and a

robust regulatory framework already exists for the review of GLMs used in ratemaking. However, GLMs are also used in other insurer operations, which have not been fully addressed by the current regulatory framework. He said the Working Group has also discussed the desire for an AI governance program to include components such as transparency and requirements for effective human-in-the-loop practices.

Commissioner Houdek asked the panelists about issues the Working Group should consider regarding AI governance and AI-related risks from an actuarial perspective.

R. Dale Hall (Society of Actuaries—SOA) said that insurers and actuaries are approaching AI through governance and risk management, digesting the NAIC *Model Bulletin on the Use of Artificial Intelligence Systems by Insurers*, along with existing corporate risk management frameworks. He noted that the SOA Research Institute participates in the U.S. Department of Commerce's AI consortium through the National Institute of Standards and Technology (NIST), whose AI risk management framework is widely discussed across the insurance industry, and said that it is useful that the supplement asks how AI systems integrate with a company's Own Risk and Solvency Assessment (ORSA) and enterprise risk management (ERM) assessments. He said actuaries view most AI tools as a new variation on models the profession has used for decades, so questions of data quality review, validation, and documentation are not entirely new. He cited Actuarial Standard of Practice (ASOP) No. 56 as a comprehensive guide for designing and evaluating models.

Henry Liu (Casualty Actuarial Society—CAS) said the risk that is most top of mind is bias, both in the statistical sense emphasized in actuarial training, which focuses on balancing models so that results regress toward the average, but also in a social sense, in that models should not introduce preferential treatment or injustice into business processes, which becomes easier to miss as the model becomes more complicated.

Spencer Sadkin (American Academy of Actuaries—AAA) said that bias relates to the actuarial fairness and explainability questions the profession has examined over the past 30 to 50 years and cautioned that overly prescriptive governance frameworks become unadaptable. He noted that AI and AI systems carry distinct definitions and urged frameworks built on strong principles rather than prescriptive requirements. Commissioner Houdek responded that the Working Group has discussed the same concern from a regulatory standpoint, given how quickly the technology is changing.

Commissioner Houdek asked the panelists to describe GLMs and how they differ fundamentally from other machine learning and generative AI algorithms.

Sadkin said that GLMs can be thought of as a more naturally explainable approach and that a fundamental difference is that whoever creates the GLM specifies the structure up front, before the model is created, unlike most other machine learning algorithms, where the algorithm is selected and then allowed to solve and produce the answer. He said generative AI differs in that it creates something as it goes, whereas the other models are supervised learning approaches intended to answer a specific question.

Liu said that GLMs are methods that assume linear, or curvilinear, relationships between the inputs and the output variable. GLMs have many real-world applications, but the constraints of this simple model form become apparent when relationships are more complex, which is where more advanced model types and training techniques come in. He said that a GLM is strictly a linear model form, whereas AI encompasses both model forms and modern computational techniques for training. He added that all of these models are attempts to describe relationships observed in the real world and that correlation does not mean causation.

Hall said that the starting point assumption of GLMs is that variables are fundamentally assumed to have a linear relationship, and that in an insurance context, GLMs allow actuaries to explicitly specify the input factors, called rating variables, that are assumed to have a mathematical relationship with an outcome for the determination of rates or premiums. He said linear regression dates back roughly 200 years and that GLMs were introduced on the actuarial risk modeling exam taken after the probability, statistics, and financial mathematics exams, where candidates learn to build, validate, and explain them. He said that rather than viewing GLMs and generative AI as entirely separate categories, there is a spectrum, with GLMs among the more transparent, easier to explain, and mathematically based approaches.

Stolyarov commented that GLMs do not fit the definition of AI because they are static, deterministic models in which humans completely prescribe the assumptions, the model structure, and the input data, so that the output is essentially determined by those human inputs, whereas true AI works in a way that cannot be precisely predicted by humans in advance. He referred to the definitions in the NAIC private passenger auto survey on artificial intelligence and machine learning circulated in 2021, which describes what AI and machine learning systems include and exclude, and said the distinction needs to be made more prominently as generative AI becomes more common. He illustrated the point, noting that a spreadsheet or a GLM performs computations in an exactly correct manner arithmetically, whereas a large language model may return a calculation accurate only to a few significant digits unless prompted to perform a precise computation, in which case it might call code capable of producing a precise result. He continued to state that emerging generative models require a different set of approaches and precautions to have confidence that the models are performing basic math correctly.

Liu responded that, within a universe of AI models, some are more unpredictable than others, but that the variability of large language model output is partly a commercialization choice by the builder, such that the model form is not inherently random but can be more flexible. He added that a GLM trained on a data set as large as those used for large language models would require computational and algorithmic simplifications, in that training on a sampled subset would not guarantee the same outcome, so a taxonomy based on the observation that GLMs feel more deterministic would ignore that GLMs can also be made more unpredictable depending on the data used for training. Sadkin commented that the question of whether GLMs are considered AI is the wrong discussion and does not matter, as academia considers GLMs to be machine learning techniques, and the distinction makes no difference to the end consumer affected by the models. He said the more important discussion concerns the different levels of scrutiny that may need to be applied based on the end user's familiarity with a model and how it is used, and he noted that the Academy is working on a paper on this topic.

Eric Ellsworth (Checkbook Health/Center for the Study of Services—CSS) asked whether there are clear definitions of when non-deterministic outputs are acceptable and when they are not. He said whether a model is AI may be a semantic question compared with what guardrails are required, expressed concern about the non-determinism of large language models, and asked how the process of understanding or mapping variability differs for a large language model compared with other approaches and at what point the logic of a process should be frozen and clear, compared to LLMs which can be fluctuational. Sadkin said the approaches used will differ depending on the use case, and that an actuary building a pricing model is almost always using a supervised learning technique in which the modeler selects the algorithm, often a GLM because GLMs naturally resemble rating plans, after which the model is static and has been assessed up front. He said the generative AI use cases he has seen in insurance are more often tasks such as compiling the 10 to 20 notes a claims adjuster may have into an overall summary, where the human-in-the-loop concept applies because the adjuster can review the individual underlying entries, reducing the need for stronger governance processes.

Liu added that the objective underlying the discussion of non-deterministic answers is avoiding catastrophic failures, such as a claim denial not based on solid evidence, and that human-in-the-loop means having someone

with human judgment and experience to prevent those failures, but any human process is also prone to randomness, as different claim examiners adjusting an identical claim may reach different outcomes, so it is important to be realistic about the level of human randomness already present in the business process that AI is replacing.

Hall said the matter comes down to asking good risk-evaluation questions, such as what data was used for training, how it was validated, what documentation exists, and what transparency can be offered. He said validating the use of a GLM is almost like turning in math homework, in that the data, the calculations, and the predictive value can all be shown, whereas layering broader generative AI topics on top calls for showing how the model was trained and that its output is auditable, transparent, reliable, and consistent across repeated runs, so that the appropriate request is to show the risk evaluation performed rather than only the mechanics of the mathematics.

Commissioner Houdek said transparency is frequently discussed by regulators, industry, and consumer representatives, but the definitions of these terms, including AI and AI systems, differ across these conversations. He asked what transparency should consist of when incorporated into a governance program.

Hall responded that transparency includes the model's purpose, its data sources, its methodology, and the validation performed, including identifying contexts in which a company is and is not comfortable using the model, and how the model is monitored and its change history to incorporate new information and retesting. He said that in an industry with protected classes, those types of questions make sense to avoid bias.

Sadkin said the ultimate concern behind transparency is protecting the consumer and challenged the view that GLMs are highly transparent while other techniques are opaque. He said that examining a decision tree shows it is entirely transparent because at each point the path goes either left or right. Such models are highly transparent, and diagnostics on variable importance features can be available on the back end.

Liu agreed that GLMs have an overhyped reputation for transparency, saying that once a model has more than roughly a dozen variables and 100,000 training records, a coefficient can be observed, but the reason it takes a particular value cannot be explained. He said there are both model-specific and model-agnostic diagnostics, such as a lift chart, which does not depend on which model generated the output, and that many of the ways regulators already evaluate GLMs have model-agnostic equivalents that may be more suitable for wider adoption, so that any model can be governed and made transparent.

Lapham agreed with the distinction between transparency toward consumers and the transparency regulators expect, but the two can bleed together because, as consumers become more aware that insurer decisions are informed by these tools, they come to regulators seeking an explanation, and it will be up to regulators to help them sort through it. He said documenting the changes made to a model over time is extremely important to transparency for regulators, and that it is also important to understand not only that a third-party modeling firm or the insurer has performed good model governance, but also, as has been observed in Colorado, when a carrier has decided not to use a model or not to use it for a specific purpose, and the rationale for that decision.

Liu responded that the transparency of the data feeding into a model is as critical as the choice of model, recalling that two decades ago, the use of credit in insurance was discussed, with reporting indicating that credit information could be inaccurate roughly 30% of the time. He said that feeding poor data, or data generated by vendors' AI systems, into a GLM is just as lacking in transparency and accountability as the concerns raised about AI within an insurance company, and he offered the example of an insurer using satellite imagery to assess the shape of his own roof, noting that he would want to know whether the assessment used his roof rather than his

neighbor's and would want the opportunity to validate and contest it. He said that as models take in more data, that data should also be subject to scrutiny and appropriate due diligence.

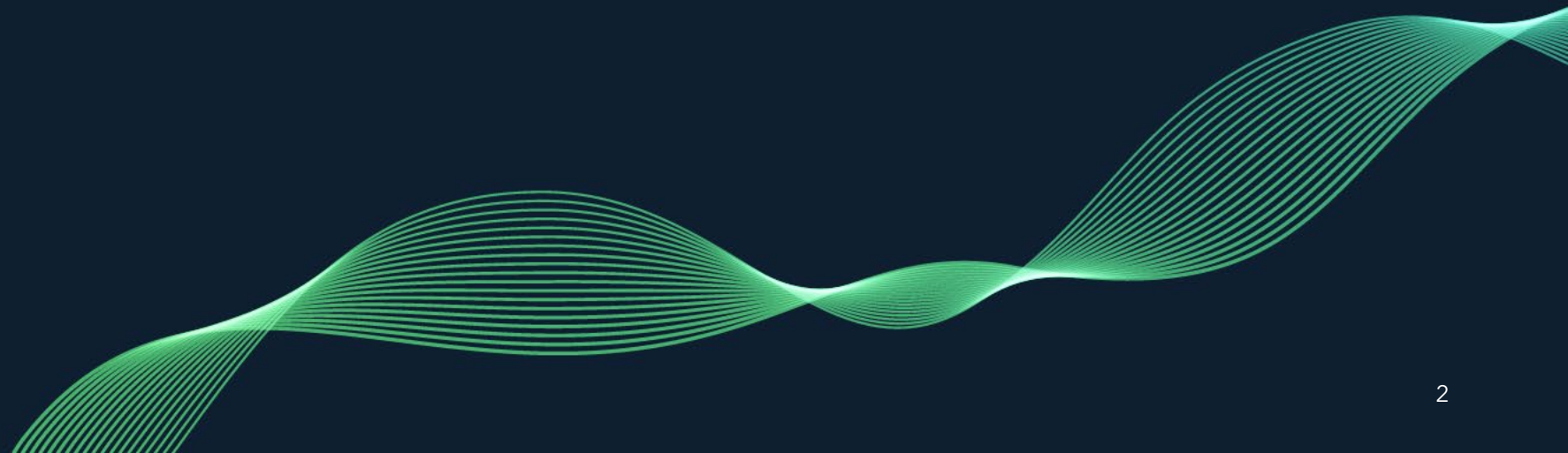
Commissioner Houdek relayed a question asking whether, given the potential for the characteristics of GLMs and AI to blend, there could be a model that is GLM-like but leaves it to AI to find and use at least one data endpoint in the service of a desired goal, such as minimizing risk in specific circumstances or keeping costs below a certain level, without further guardrails. Liu responded that such approaches exist because an AI model is not necessarily the most cost-effective option for a specific business process, so a simpler model may be used in production while AI is used to train on the back end. He said these are likely not being used for ratemaking, as he has not seen filings requesting approval of a process to derive rating values on the basis that whatever the process produces will be used, but that outside of ratemaking, AI-trained GLMs and other simple model forms are in use today within insurance and within the general economy.

Sadkin cautioned that discussions of AI should be crystal clear in distinguishing between AI models and algorithms, as the term appears opaque at present, and noted that the European Insurance and Occupational Pensions Authority (EIOPA) has discussed the same question of whether GLMs should be carved out and developed proposals on how that should work. He said his view, which he believes others share, is that such an approach expends unnecessary additional regulatory effort to categorize modeling techniques. Rather, it should be based on the level of potential risk, rather than on establishing a framework that allows matters to flow through naturally. Commissioner Houdek said the NAIC participates in the EU-U.S. Insurance Dialogue Project, an informal joint project with the EIOPA, in which AI regulation and supervision have been a focus area for the last couple of years.

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

2. Receive an Update on the Artificial Intelligence (AI) Risk Evaluation Supplement Pilot

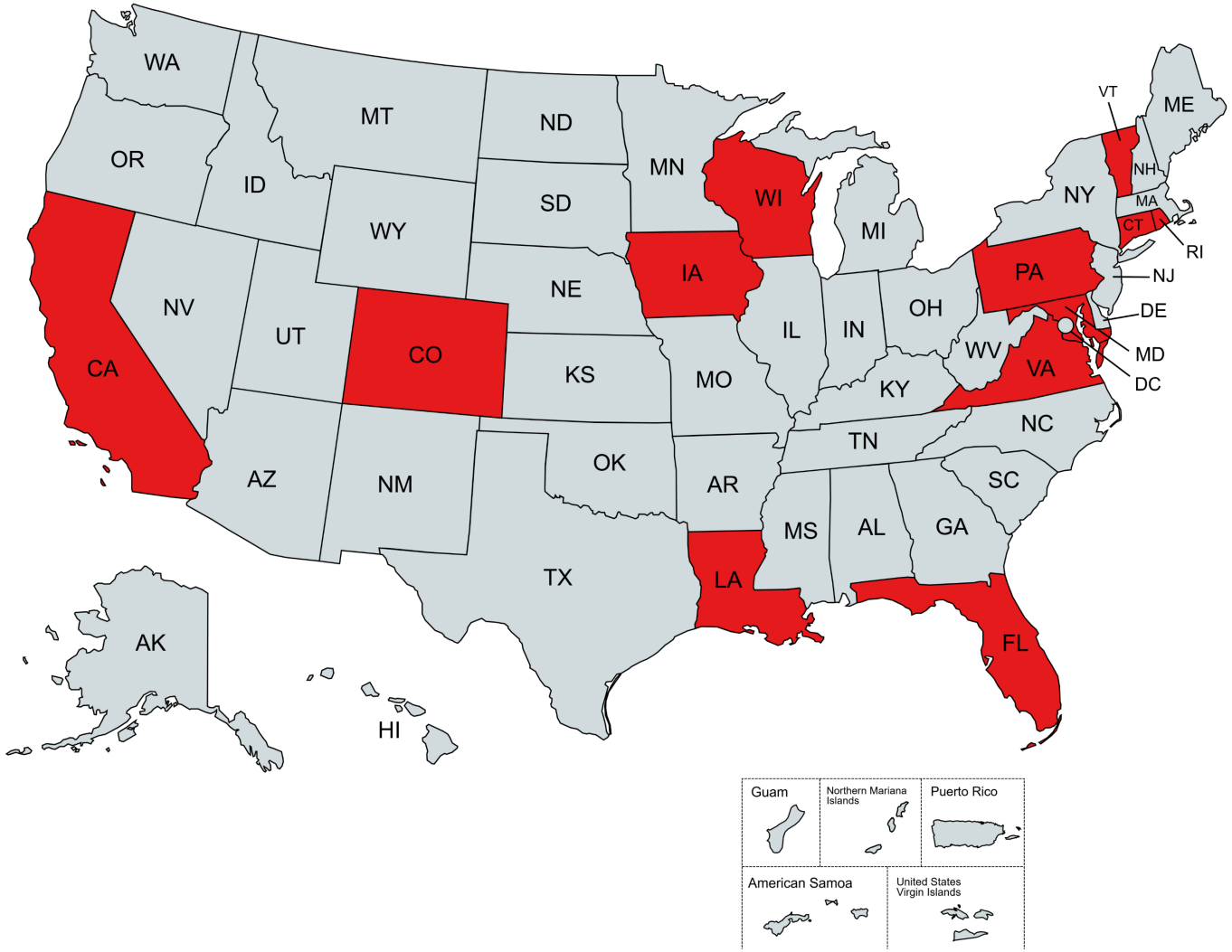
–Commissioner Nathan Houdek (WI)



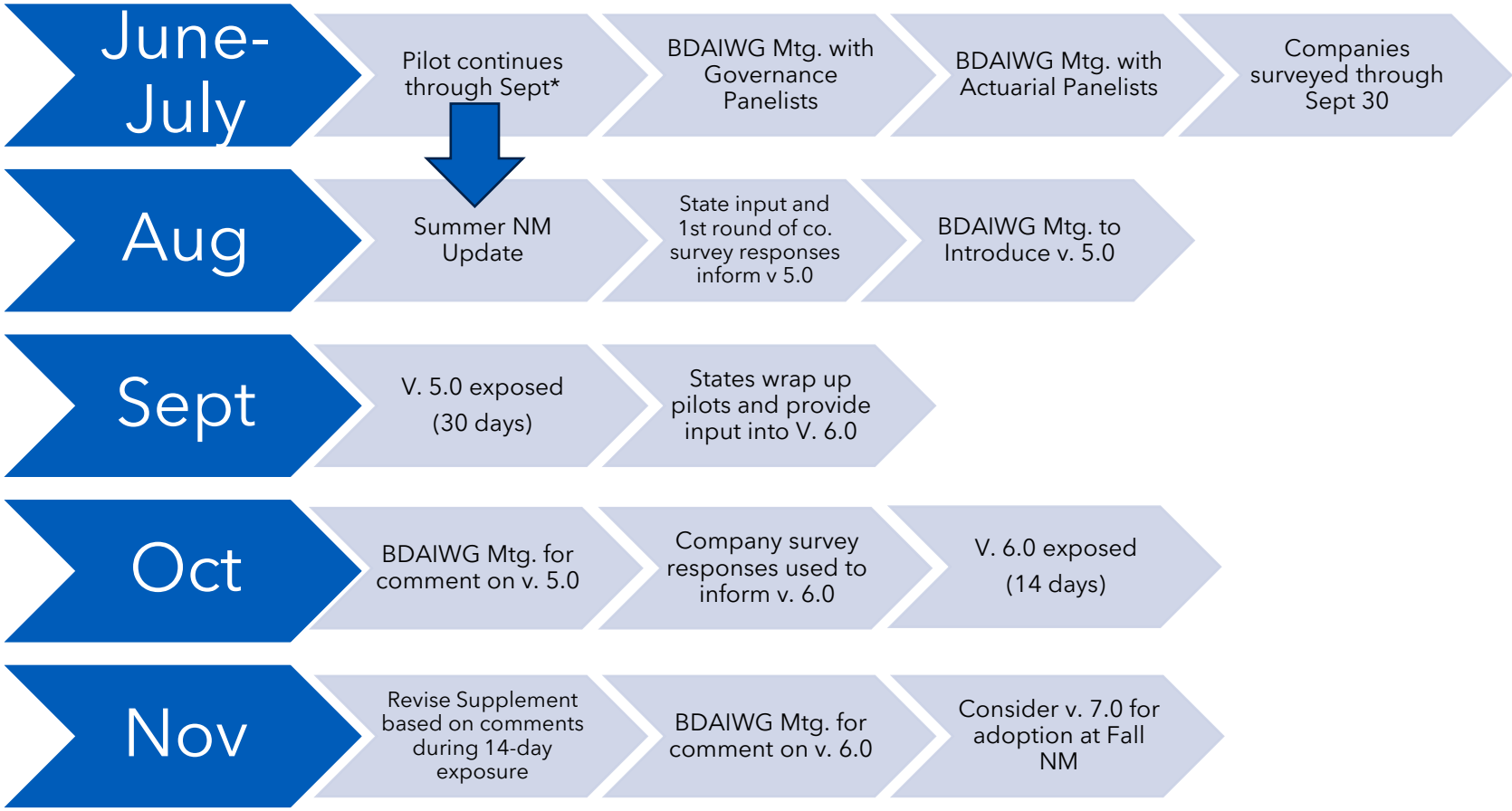
NAIC NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

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 Risk Evaluation Supplement 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 the Supplement drafts through adoption and future iterations.

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

Attachment B

–Edin Imsirovic (AM Best)

AM BEST

AI Governance in Insurance: Distinguishing Risk Across Predictive, Generative and Agentic Systems

Edin Imsirovic – Director, AM Best

NAIC

DISCLAIMER

Discussion material only; no change to criteria or guidance

This deck is for discussion. It does not change AM Best criteria, methodology, rating guidance, or create a formal AI governance assessment framework.

The discussion draws from public regulatory, supervisory, standards, model-risk, and AI-security materials. It uses those materials to address how governance expectations may change as AI systems move from prediction, to generation, to more action-oriented workflows.

Examples are illustrative. They are not a checklist, scoring tool, or statement of how AM Best will evaluate any insurer.

What We Mean by Predictive, Generative, and Agentic

Working definitions used throughout the discussion

Predictive AI	**Predictive AI uses data to estimate, sort, score, or flag something. In insurance, this can include pricing models, fraud scores, claims triage, risk selection, or tools that help decide which cases need more review. It usually supports a decision within a defined process, rather than creating new content or taking action on its own.**
Generative AI	**Generative AI creates a new content from the information it is given. It can summarize documents, draft language, extract details from notes or images, or turn unstructured information into a more usable format. Because the output can vary and may sound confident even when it is wrong, review, source checks, and clear use boundaries matter.**
Agentic AI	**Agentic AI can work through a sequence of steps toward a goal, rather than only returning a single answer. It may use tools, look up information, update systems, route work, or start the next step in a workflow. The governance question changes because the system may affect the process directly, so permissions, logging, human review, and the ability to stop or reverse actions matter.**

Two Lenses for Discussing AI Governance Quality

Innovation and ERM as existing analytical lenses

Innovative Lens

- Distinguishes adoption from durable operating integration
- Assesses inputs: leadership, culture, resources, and processes & structure
- Assesses outputs: measurable results and level of transformation, including change in underwriting, claims, fraud, and service workflows

ERM Lens

- Assesses whether governance is commensurate with the AI risk profile being created
- Considers product and underwriting, operational, regulatory/legal, and other relevant exposures
- Also highlights data integrity, execution, control lag, and third-party dependence

Common thread : The same AI deployment can create competitive differentiation and operational or strategic risk; the question is whether governance keeps pace.

How AI Governance Can Affect Value and Risk

Technology claims alone tell us very little.

What matters is whether AI-driven changes to the operating model last long enough to show up in measurable results, and whether governance keeps those same deployments from accumulating operational and strategic risk.

When governance keeps pace with deployment	When deployment outpaces governance
Expense and processing improvement	**Execution risk and control breakdowns**
Underwriting quality improvement	**Cyber and data exposure expansion**
Claims efficiency and consistency	**Vendor dependency and portability risk**
Fraud control and leakage reduction	**Regulatory, consumer, and explainability exposure**

Key point: Whether AI adds to operating performance or operational risk depends heavily on the governance around it.

Where AI-Related Risk Fits Within Existing ERM Categories

AI can create new pathways into existing risk categories

The same AI deployment that improves outcomes can also create Product & Underwriting, Operational, and Legislative / Regulatory / Judicial / Economic exposure when governance falls behind the deployment.

Product & Underwriting

AI driven pricing or risk selection error, model drift affecting segmentation, and unintended portfolio mix or profitability effects

Operational

Poor data quality, workflow breakdowns, claims handling errors, fraud exposure, cyber and IT resilience gaps, and weak oversight of third-party models

Legislative / Regulatory / Judicial / Economic

Consumer protection exposure, explainability and disclosure expectations, litigation or regulatory scrutiny, and changing legal or market conditions affecting pricing or strategy

Key point: AI governance risk already flows through ERM categories. The core question is whether risk-management capability is commensurate with the AI risk profile being created.

From Prediction to Generation to Action

Why one governance approach may not be enough

Wave 1

2000s–2010s

Structured data AI

Fraud detection, pricing refinement, and claims triage in bounded, structured workflows.

Wave 2

mid-2010s–2020

Computer vision + NLP

Photo-assisted claims, document classification, and intake automation for higher-complexity workflows.

Wave 3

2020–present

Generative AI

Narrative inputs converted into summaries, structured outputs, drafts, and recommendations.

The governance problem changes

Wave 4

2023–present

Agentic AI

Planning, tool use, workflow orchestration, and sequential decision-making with reduced human intervention.

Governance catching up

Increasing governance complexity

Key point: Each wave expands what the system can do, from bounded prediction, to generated output, to multi-step action. Governance has to expand with it.

Three AI Capabilities, Three Governance Challenges

Inputs, outputs, and control challenges across AI capability tiers

	Predictive / analytical AI	Generative AI	Agentic AI
Typical inputs	Structured data, tabular features, defined signals	Narratives, documents, images, mixed context	Context + goals + tools + system instructions
Typical outputs	Scores, classifications, rankings, recommendations	Summaries, drafts, extracted fields, explanations	Action sequences, workflow steps, external tool calls
Governance challenge	Validation, fairness, drift, documentation	Hallucination, boundary control, prompt variation, update effects	Authorization, rollback, logging, human override, task boundaries

Key point: Each capability creates a different governance problem, and later capabilities carry forward many of the earlier risks.

When Controls Fail: Illustrative Patterns by Capability Class

What can go wrong, and what shows the controls are working

Predictive / Analytical AI

WHAT CAN GO WRONG

- Discriminatory or opaque adverse decisions at scale
- Pricing / selection drift that distorts portfolio quality
- Reserve or performance deterioration before drift is caught

EVIDENCE THAT CONTROLS WORK

- Evidence that monitoring results change thresholds or model use
- Independent validation or challenge
- Explainability for adverse or high impact decisions

Generative AI

WHAT CAN GO WRONG

- Hallucinated summaries or extracted fields driving poor handling decisions
- Privacy or boundary failures in documents, prompts, or outputs
- Systematic content errors affecting claims, underwriting, or customer outcomes

EVIDENCE THAT CONTROLS WORK

- Output review calibrated to decision criticality
- Vendor updates trigger an impact review before release
- Demonstrated data boundary controls preventing sensitive policyholder information from entering unapproved inputs or workflows

Agentic AI

WHAT CAN GO WRONG

- Unauthorized or poorly bounded actions affecting coverage, claims, or service
- Compounding multi step errors before detection or human review
- Weak reversibility and correlated failures across common platforms

EVIDENCE THAT CONTROLS WORK

- Tested kill switch, rollback, and full action logging
- Clear evidence of human review before consequential actions
- Defined action boundaries and override authority

Key point: The failures matter, but the stronger test is whether the controls work in practice.

Novelty Alone Does Not Determine Risk

Five factors that help judge the risk of a specific use case

01
Decision criticality

How directly does the output affect a consumer or financial outcome?

02
Degree of autonomy

Is the system assisting, informing, or executing?

03
Opacity / explainability

Can the insurer reconstruct why a specific output or action occurred?

04
Speed of change

How quickly can behavior drift or vendor updates change results?

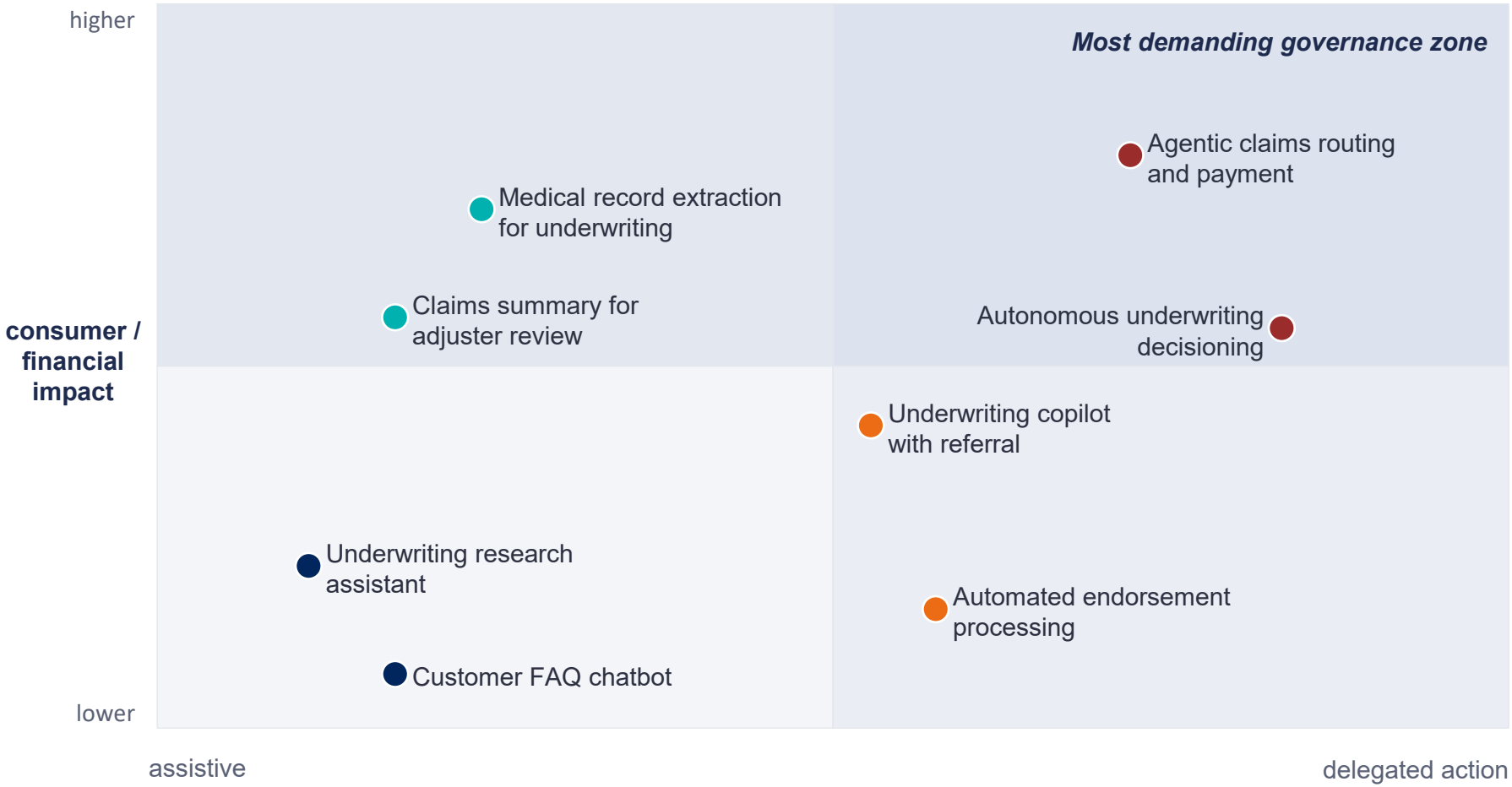
05
Third-party dependency

How much can the insurer independently audit, test, and override?

Key point: Risk depends on the use case such as its impact, autonomy, opacity, rate of change, and third-party reliance.

The Risk Gradient: An Illustrative View of Impact and Autonomy

Illustrative view using two factors: impact and autonomy



Reading the chart

This chart uses two of the five factors. It is a simple way to organize the discussion, not a complete assessment.

The examples are illustrative. The same technology can sit in different places depending on how it is used.

Key point: Impact and autonomy help show why the same technology can require very different governance.

How Governance Considerations Expand Across the Three Capability Tiers

Controls build on one another as capability increases

Predictive AI	Generative AI	Agentic AI
• Independent validation • Bias / proxy-discrimination testing • Drift monitoring, thresholds, and revalidation triggers • Documentation for review • Vendor contracts with audit rights	• Input / output controls • Prompt governance and versioning • Data-boundary controls for sensitive policyholder information • Human review calibrated to risk • Update-impact assessment and revalidation before release	• Task-scope boundaries • Approval thresholds by action • Capability-change testing, tested rollback, and kill switch • Full action logging • Defined override authority and human escalation pathways

Key point: Earlier controls still matter, but they are not enough on their own.

What Stronger Governance May Look Like

Five signs that governance can hold up over time

01

Governance that survives turnover

Review cadences, escalation protocols, and accountability survive leadership transitions and staff turnover.

02

Data governance that works in practice

Insurers can trace inputs, assess data suitability, and govern third-party data in practice, not just on paper.

03

Evaluation tied to decisions

Pre-deployment testing, red-teaming, and ongoing application-level evaluation are tied to defined evidence thresholds for release, escalation, or rollback.

04

Monitoring that changes behavior

Monitoring results trigger changes in thresholds, controls, model use, or workflow boundaries.

05

Resources that grow with use

Dedicated AI governance talent, infrastructure investment, and control resourcing scale with deployment scope as AI use expands.

Key point: Strong governance still works when systems drift, vendors change, and people move.

Controlled Integration vs. Fragile Tool Layering

Why the operating environment matters, not just the model

Controlled integration

- Workflows, governance, and accountability are redesigned together
- Improvements spread across data, triage, pricing, service, and fraud controls
- Governance sits in the business process, not beside it
- New capability and controls grow together

Fragile tool layering

- AI tools are dropped onto unchanged legacy processes
- Benefits stay isolated
- Controls sit outside the day-to-day workflow
- Automation expands faster than the governance infrastructure required to oversee it

Key point: The deepest governance risk is often not the model itself. It is the operating environment around the model.

Shared Dependency and Concentration Risk

Common vendors, models, data, cloud providers, or other shared infrastructure can create correlated risk

What makes vendor dependence risky?

- The insurer may not fully understand or control underlying model behavior
- Contracts often fall short on audit rights, update notice, and data transparency
- Portability and exit options matter more as AI becomes embedded in critical workflows
- Vendor model updates can materially alter system behavior without triggering internal governance review
- Common vendors, models, data sources, or orchestration layers can create correlated risk across multiple carriers simultaneously

Predictive AI	Requires validation, meaningful audit rights, and update-notification provisions, though gaps remain common in practice
Generative AI	Greater sensitivity to vendor model updates and data-boundary drift
Agentic AI	Additional risk from delegated action, correlated failures, and harder override

Key point: The real question is whether the insurer can understand, test, and override the vendor's system.

Operational vs. Documentary Governance

Illustrative questions that test whether governance works in practice

Inventory and classification	• **Can the insurer produce a complete AI inventory?** • Are systems classified by impact and autonomy? • Does each risk tier have clear governance expectations and controls?
Validation and monitoring	• Before release, are evaluation criteria and re-testing triggers clear? • **Do threshold breaches trigger documented responses?** • Is validation conducted independently of the deployment team?
Human oversight	• Is human review substantive given volume and time constraints? • **Can the insurer show examples where human review changed the outcome?**
Vendor and agentic controls	• Are audit rights usable, update notices reliable, and rollback procedures tested? • Are task boundaries and kill switch procedures explicit? • **What happens if a vendor or system update materially changes behavior?**

Key point: Documentary governance answers with policies. Operational governance answers with evidence.

The Central AI Governance Question

The real question is whether a carrier can keep AI use within clear limits, see how it is behaving, and stop, reverse, or recover when needed, as systems change, autonomy grows, and human oversight becomes harder.

Regulatory Backdrop

Selected public materials shaping the discussion; not a complete list

United States

National Association of Insurance Commissioners (NAIC) Model Bulletin (2023): governance, testing, documentation, third-party oversight

NAIC AI Systems Evaluation Tool (2026 pilot in 12 states): AI use, governance, high-risk systems, and data

New York Department of Financial Services (NYDFS) Circular Letter 7 (2024): fairness testing and governance for underwriting and pricing

Colorado External Consumer Data and Information Sources (ECDIS) rules: state example of algorithm and data governance in insurance

Federal Reserve Board / Office of the Comptroller of the Currency (OCC) SR 11-7: long-standing model-risk baseline

National Institute of Standards and Technology (NIST): AI Risk Management Framework and Generative AI Profile

International and standards

European Union AI Act (2024): risk-based AI framework (certain insurance uses treated as high risk)

Office of the Superintendent of Financial Institutions (OSFI) Guideline E-23: model-risk governance across the lifecycle; effective 2027

Prudential Regulation Authority (PRA) Supervisory Statement 1/23: banking model-risk principles (widely read across finance)

International Association of Insurance Supervisors (IAIS) application paper and European Insurance and Occupational Pensions Authority (EIOPA) supervisory opinion on AI (2025)

International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 42001 (2023): AI management system standard

What to watch

NAIC Evaluation Tool pilot running in 12 states; adoption anticipated at the 2026 Fall National Meeting

Financial Stability Board (FSB) sound-practices final report AI sound-practices final report expected October 2026

EU high-risk compliance dates reset to late 2027 and 2028

Generative and agentic workflows moving faster than older model-risk rules

Discussion

Disclaimer

Disclaimer

US Securities Laws explicitly prohibit the issuance or maintenance of a credit rating where a person involved in the sales or marketing of a product or service of the CRA also participates in determining or monitoring the credit rating, or developing or approving procedures or methodologies used for determining the credit rating.

No part of this presentation amounts to sales / marketing activity and AM Best's Rating Division employees are prohibited from participating in commercial discussions.

Any queries of a commercial nature should be directed to AM Best's Market Development function.