

Draft: 8/26/24

Big Data and Artificial Intelligence (H) Working Group  
Virtual Meeting  
July 29, 2024

The Big Data and Artificial Intelligence (H) Working Group met July 29, 2024. The following Working Group members participated: Michael Humphreys, Chair and Shannen Logue (PA); Kevin Gaffney, Vice Chair and Mary Block (VT); Jimmy Gunn (AL); Alex Romero and Molly Nollette (AK); Tom Zuppan represented by Lori Munn (AZ); Ken Allen (CA); Michael Conway represented by Jason Lapham (CO); Andrew N. Mais represented by George Bradner (CT); Karima M. Woods (DC); Rebecca Smid (FL); Weston Trexler (ID); Erica Weyhenmeyer (IL); Amy L. Beard represented by Victoria Hastings (IN); Doug Ommen represented by Jared Kirby (IA); Tom Travis (LA); Sandra Darby (ME); Raymond Guzman (MD); Caleb Huntington (MA); Jeff Hayden and Jake Martin (MI); Jacqueline Olson and Phil Vigliaturo (MN); Cynthia Amann (MO); Connie Van Slyke (NE); Scott Kipper represented by Nick Stosic (NV); Christian Citarella (NH); Adrienne A. Harris represented by Kaitlin Asrow (NY); John Harrison represented by Tracy Biehn (NC); Jon Godfried represented by Colton Schulz (ND); Judith L. French represented by Matt Walsh (OH); Elizabeth Kelleher Dwyer (RI); Michael Wise (SC); Carter Lawrence represented by Emily Marsh (TN); J'ne Byckovski and Rachel Cloyd (TX); Scott A. White represented by Dan Bumpus (VA); Nathan Houdek represented by Lauren Van Buren (WI); and Bryan Stevens represented by Lela Ladd (WY).

1. Adopted its Spring National Meeting Minutes

Commissioner Gaffney made a motion, seconded by Superintendent Dwyer, to adopt the Committee's March 16, minutes (see *NAIC Proceedings – Spring 2024, Big Data and Artificial Intelligence (H) Working Group*). The motion passed unanimously.

2. Received an Update on the Working Group's Health Insurance AI/ML Survey Work

Commissioner Humphreys provided an update on the status in the development of the health insurance AI/ML surveys, which included tailoring the prior surveys' questions to health insurance, proceed with a pilot study, and issue the survey later this year. He reiterated that the purposes of the health AI/ML surveys are to understand how industry is using AI, how the use of AI is governed, and how the products and systems are being developed to guide future discussions on next steps. Commissioner Humphreys stated that the group has had some conversations with consumer representatives and are currently finalizing conversations with a handful of large major medical carriers that will participate in the pilot program to give feedback on the survey questions. By the Spring National Meeting the group will have the analysis and report complete for discussion at the group level and publicly.

Birny Birnbaum (CEJ) asked what the plan was for reissuing the surveys to receive updated responses. Commissioner Humphreys deferred this question to Shannen Logue (PA) to answer.

Josh Goldberg (HCSC) asked to confirm that the launch of the survey is planned for November 11 with a due date of January 15. Commissioner Humphreys confirmed.

Shannen Logue (PA) stated the group met with consumer representatives on May 13 to receive feedback and stated that the survey will be issued for public access on October 4. She stated that the health surveys will include questions relating to data usage, arrangements with third parties, coordination of governance with existing health provider governance standards, and will be tailored to the use of AI in operational functions of health insurers. She explained the group's intentions are to ensure that the questions align with the NAIC Model Bulletin.

Regarding the auto surveys, the group will conduct regulator-only follow up discussions with selected personal auto carriers. Among those carriers, for those that initially responded that they do not currently use AI/ML in their operations, the group will follow up to ask whether they have begun to use AI or ML in which operations and in which capacity. For the selected carriers that originally responded they are currently using AI/ML, follow up questions will be asked about any changes in their use of AI/ML, whether they have begun to use generative AI, their degree of human involvement, efforts to identify and mitigate model drift, and their uses of third-party systems. The group anticipates completing the first round of follow up interviews by October 31 and anticipates repeating the surveys every two to three years.

Birnbaum asked whether the plan consists of following up with selected companies who provided anomalous responses between auto and home who indicated that they have certain uses or that they were engaged in using AI/ML. Logue confirmed that is correct. Birnbaum expressed that repeating the surveys on a regular schedule would result in more consistent responses.

Lucy Culp (Leukemia & Lymphoma Society) asked whether Other Health, like Short Term Plans Accepted Benefits, will be included in the surveys. Logue responded that the surveys will start with comprehensive major medical plans (individual, the small group, large group as well as student health), but then there could be a second round of surveys.

### 3. Received a Presentation on the Society of Actuaries' Research on Inference Methods

Dorothy Andrews (NAIC) covered several aspects of the Society of Actuaries (SOA) paper on inference methods, explained the theory of the BIFSG method, and included examples of the results of the method. Andrews discussed the underlying data used by the BIFSG method, its limitations, and concerns regarding its accuracy. She showed how the BIFSG method has been applied to a variety of studies and applications, including health care decision making, mortgage and non-mortgage lending patterns, academic research, taxation, and financial credit access issues. She explained a few of the performance metrics used, and introduced the concepts of the probabilistic and statistical types of inference methods. She clarified that the BIFSG method is a Bayesian probabilistic approach. She explained that the BISG only uses surnames, geo-location, and census bureau demographics data to estimate race, while the BIFSG additionally uses first names to estimate probabilities of race and ethnicity. The BIFSG method has been applied on data from mortgage applications and voter registration rolls and has shown improvement over the BISG method in accuracy and coverage. The BIFSG method was used to find that the incidence of missing race and ethnicity data is higher among non-Hispanic and Hispanic blacks than other groups.

Andrews then walked through the mechanics of how the probabilities are calculated in the BIFSG method using Bayesian theory, and provided the results of estimated probabilities of race for Miguel Romero (NAIC), Scott Sobel (NAIC), and herself. She explained why her estimated race was incorrect considering her first and last names and her location of residence. In that example, she provided insights into how bias can be embedded in reference/training data. She provided another example that referenced a study where the researchers found the BIFSG method overestimated the earned income tax credits take-up rate for whites, and underestimated the rate for blacks; it underestimated average tax rates for whites but was fairly accurate for blacks, Hispanics, and other groups; and it predicted higher audit rates for whites than non-whites, which is in conflict with actual audit rates. She clarified that the BIFSG method was designed to perform inference on a large group of people, not to infer the race at an individual level.

Sylvia Yee (DREDF) asked about whether the method would work well on people of mixed race. Andrews responded that the method may not be as accurate on people of mixed race, and for people who live in very diverse communities.

Birnbaum commented that perfect is the enemy of good, in that there is a technology that has been used in regulatory applications that, while may not be perfect, may be fit for purpose to assess bias in AI applications and insurance applications. Further he stated that while there is always room for improvement, there is no reason for the NAIC not to endorse testing for racial bias using the BIFSG method.

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

SharePoint/NAIC Support Staff Hub/Member Meetings//H CMTE/2024\_Interim Meetings/Minutes-BDAIWG072924.docx

# Considerations for Implementing Artificial Intelligence in Insurance

Prepared for:

**NAIC Big Data And Artificial Intelligence (H) Working Group**

**Tom Prince, FCAS, MAAA**

November 12, 2024



## Who we are



**Tom Prince, FCAS, MAAA**  
**Principal**  
Consulting Actuary  
[tom.prince@milliman.com](mailto:tom.prince@milliman.com)

**“We can only see a short distance ahead, but we can see plenty there that needs to be done.”**

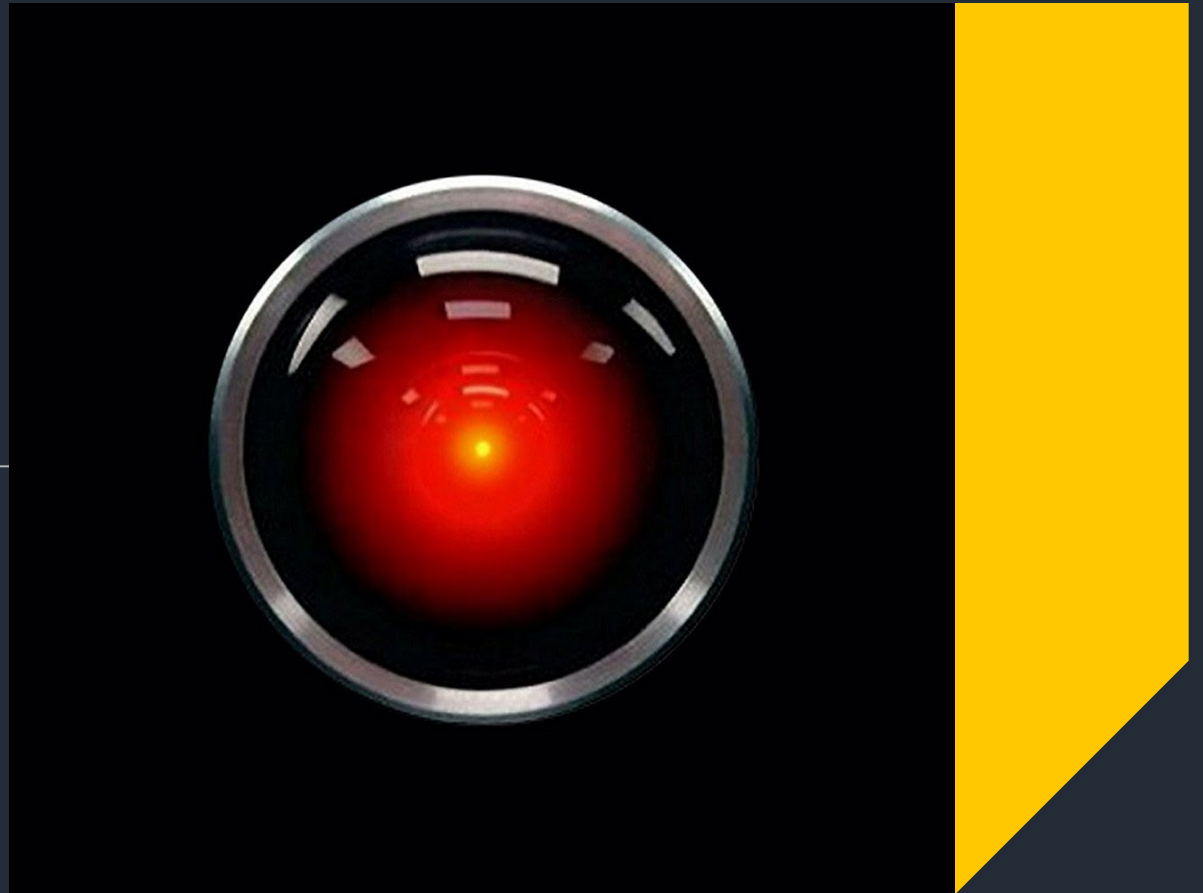
— Alan Turing

# Considerations for Implementing AI in Insurance

## Agenda

<b>5</b> <b>Overview</b>	<b>8</b> <b>Use case study</b>	<b>11</b> <b>Build vs Buy</b>
<b>14</b> <b>AI Risk</b>	<b>20</b> <b>Wrap up / Q&amp;A</b>	

Image source: <https://www.nytimes.com/2018/03/30/movies/hal-2001-a-space-odyssey-voice-douglas-rain.html>



# AI in insurance: a brief overview





# Common Language

What are we even talking about?

## Glossary of common terms

Source: NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers



### Machine Learning (ML)

a field within artificial intelligence that focuses on the ability of computers to learn from provided data without being explicitly programmed.

### Artificial Intelligence (AI)

models and systems that perform functions normally associated with human intelligence

### Generative Artificial Intelligence (GAI or Gen AI)

a class of AI Systems that generate content [...] that is similar to, but not a direct copy of, pre-existing data or content.

### AI System (AIS)

A machine-based system that can [...] generate outputs such as predictions, recommendations, content [...] or other output influencing decisions

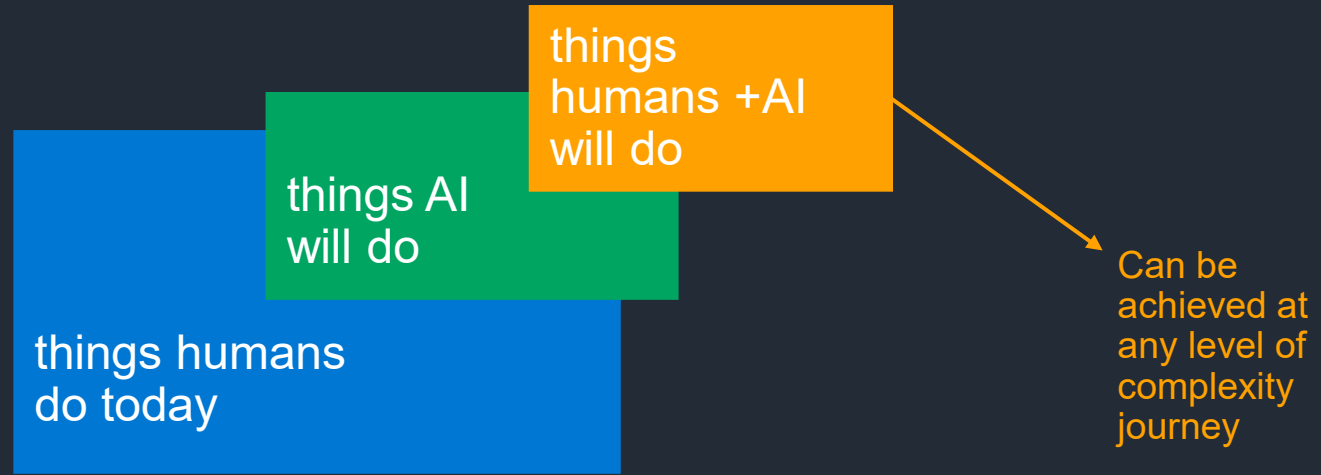
### AI Actors

all persons or entities facilitating the business of insurance that play an active role in the AI system life cycle

### Third Party

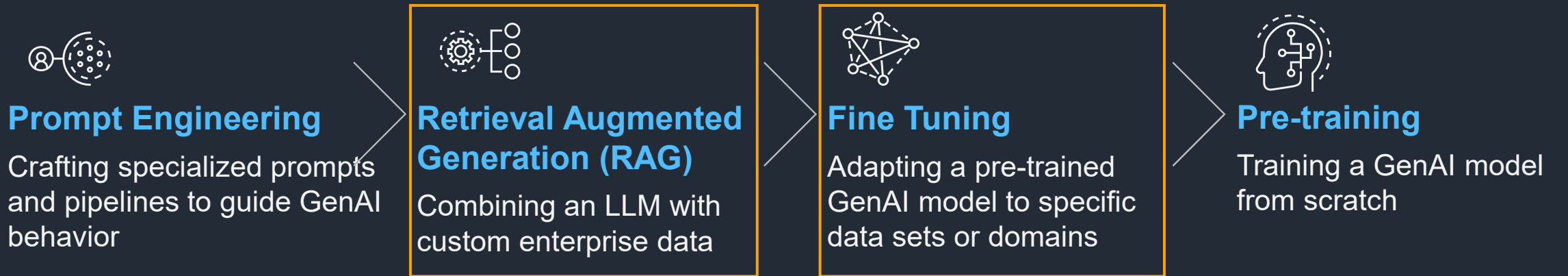
an organization other than the Insurer that provides services, data, or other resources related to AI

# The GenAI Journey



## COMPLEXITY

Flowchart source: Databricks Big Book of Generative AI, <https://www.databricks.com/resources/ebook/big-book-generative-ai>



# A brief look at two Gen AI use cases in insurance



# Use case 1: Build Customized GPTs

## Applications and Considerations

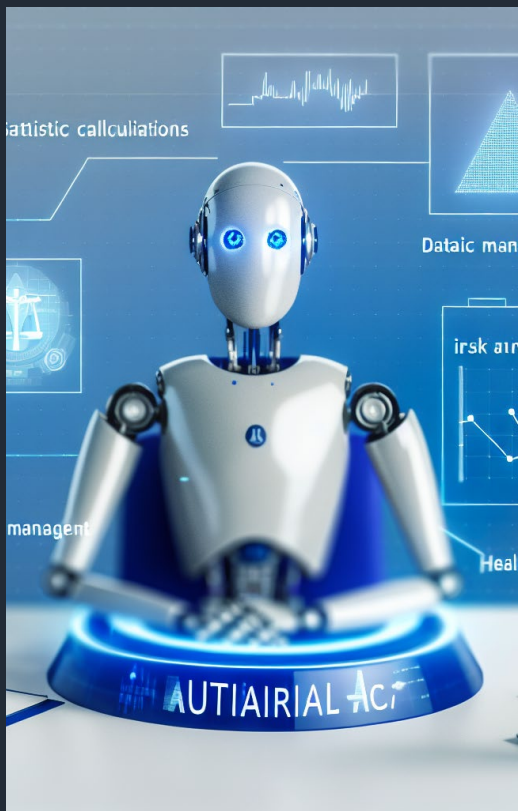


Image Source: Milliman ChatGPT

### Use Cases

Build an AI chat bot that answers difficult questions to assist Milliman professionals and our clients

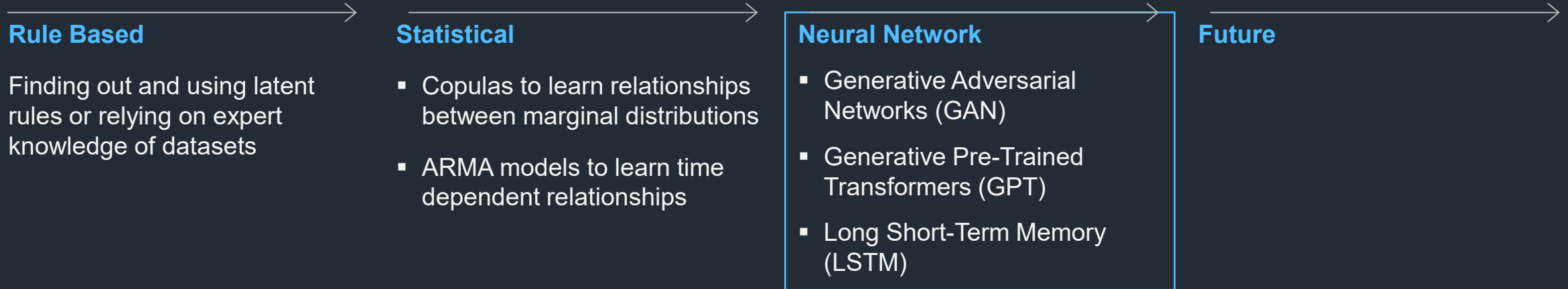
- Legal compliance
- Actuarial guidance
- Rate/product filing intelligence

### Lessons Learned


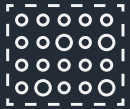


- Data is key – garbage in, garbage out
- Use most powerful foundational LLM available
- Pre-built solutions often can't handle broad and complex subject matter
- Refinements needed to minimize errors and hallucinations – keep a human in the loop (HITL)!
- Build an extendable framework for applications beyond a chatbot (e.g., document creator, agentic AI system)
- Outside resources valuable, but vendor management a must

# Use case 2: Creating Synthetic Claims Data with Generative AI Methods

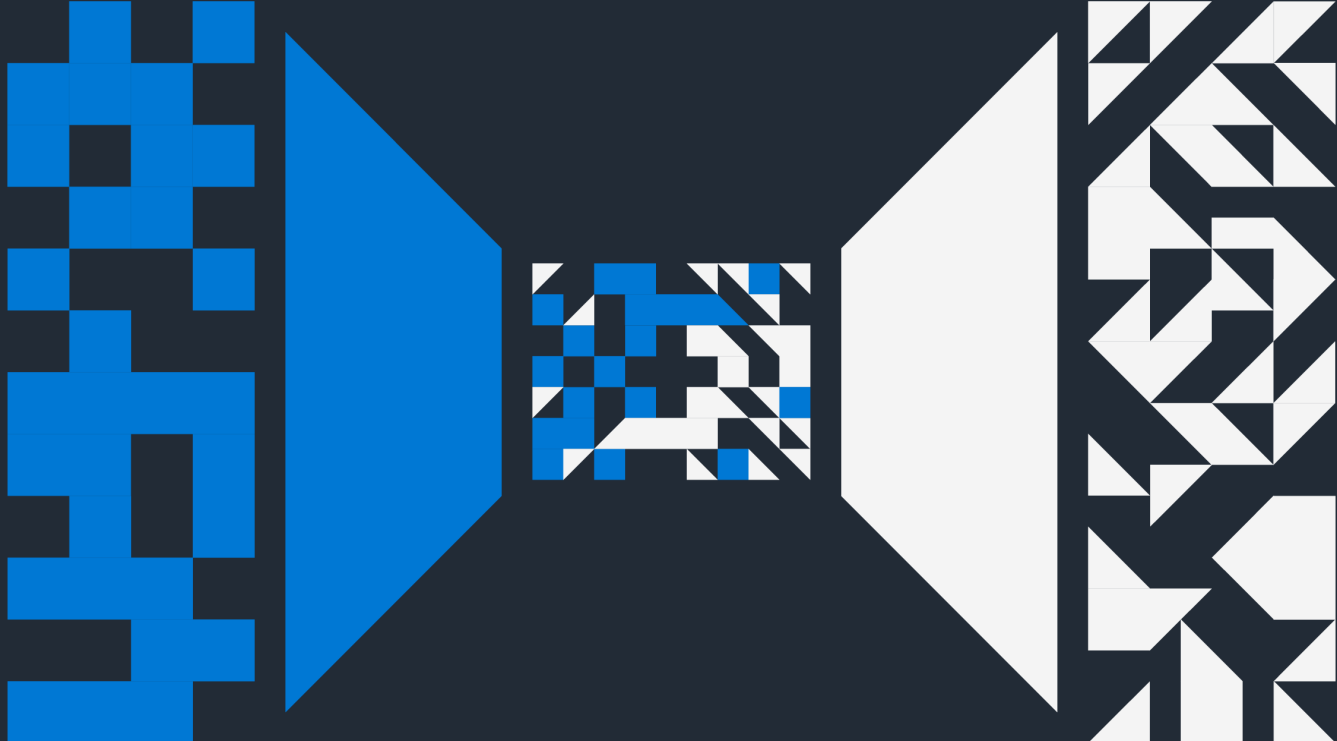
## Evolution of Models



## Why Synthetic Data?

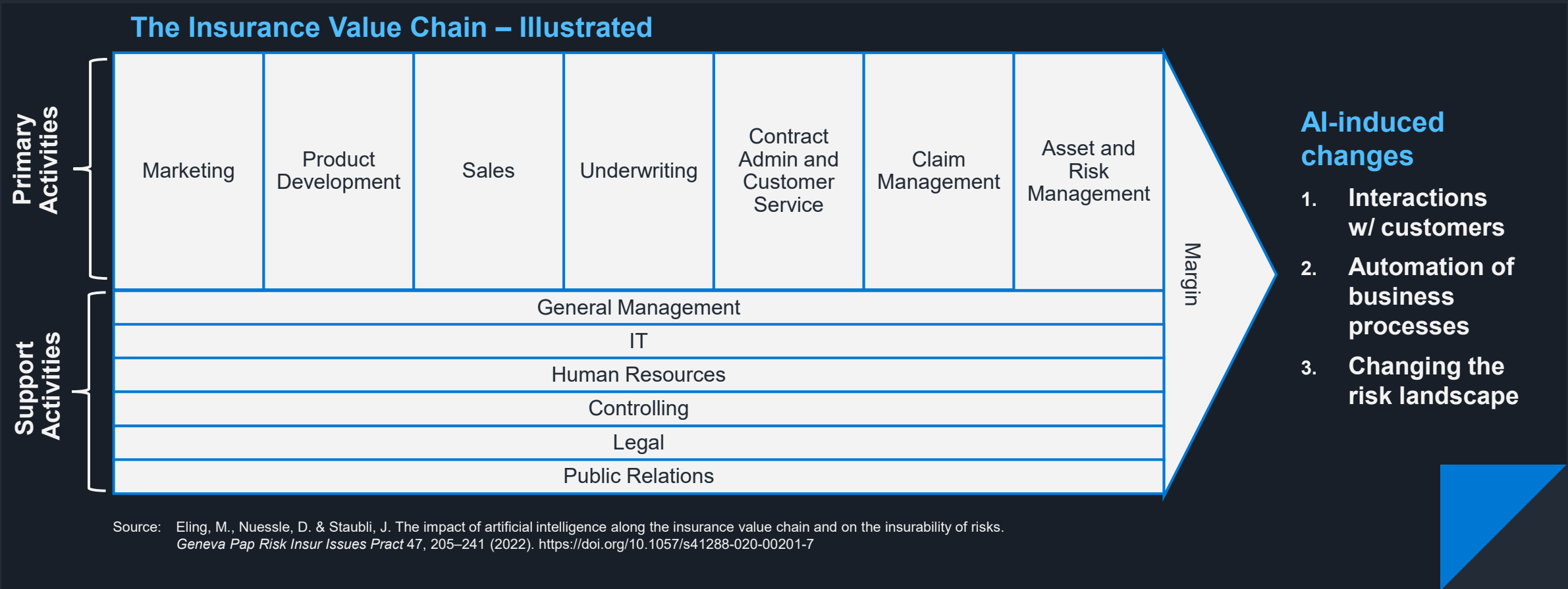
 <p><b>Data Security</b></p> <ul style="list-style-type: none"> <li>▪ Privacy protection</li> <li>▪ Regulatory compliance</li> <li>▪ Sharing and collaboration</li> </ul>	 <p><b>Increased Data Volume</b></p> <ul style="list-style-type: none"> <li>▪ More data when insufficient or difficult to collect</li> <li>▪ More targeted samples by characteristics</li> </ul>	 <p><b>Feature Embeddings</b></p> <ul style="list-style-type: none"> <li>▪ Summary of data features</li> <li>▪ Concretizing real/abstract ideas into numerical representations</li> </ul>	 <p><b>Reduced Storage Footprint</b></p> <ul style="list-style-type: none"> <li>▪ Storage of parameters only</li> <li>▪ Reduced amount of input data</li> <li>▪ Data generation on demand for various models</li> </ul>
--	---	--	--

# Build vs Buy: a fundamental decision in AI systems



# AI and the Insurance Value Chain

An extremely brief overview

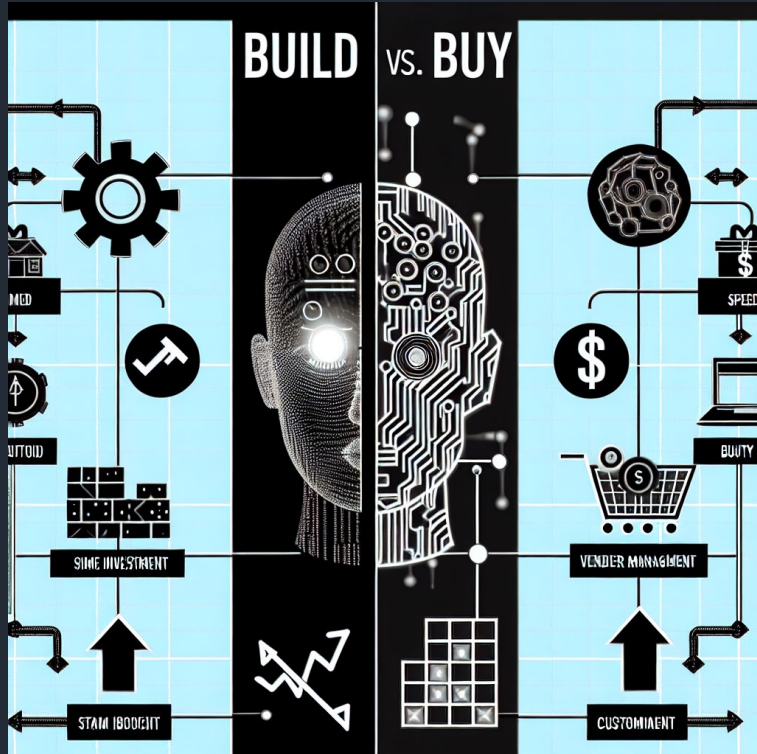


Source: Eling, M., Nuessle, D. & Staubli, J. The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *Geneva Pap Risk Insur Issues Pract* 47, 205–241 (2022). <https://doi.org/10.1057/s41288-020-00201-7>

# The Build vs Buy Decision

Considerations for a critical choice in AI systems

Image Source: Milliman ChatGPT



## 1

### Evaluating AI Appropriateness

- Is AI the right tool or a shiny object?
- Data:
  - Do we have it?
  - Can we get it?
  - What is it?

## 2

### Problem Type: Core Business vs. Support Functions

- Using AI in support functions is table stakes
- Tailored solutions needed for core business

## 3

### Evaluate organizational skill sets

- Is development an option – training or hiring
- Build Intangible capital
- Not just technical
  - vendor mgmt.
  - change mgmt.

## 4

### Optimize Value and Mitigate Risks

- Speed vs Quality – are you outpacing the foundational models?
- Purpose and Target
- Building for buy in – engage your stakeholders

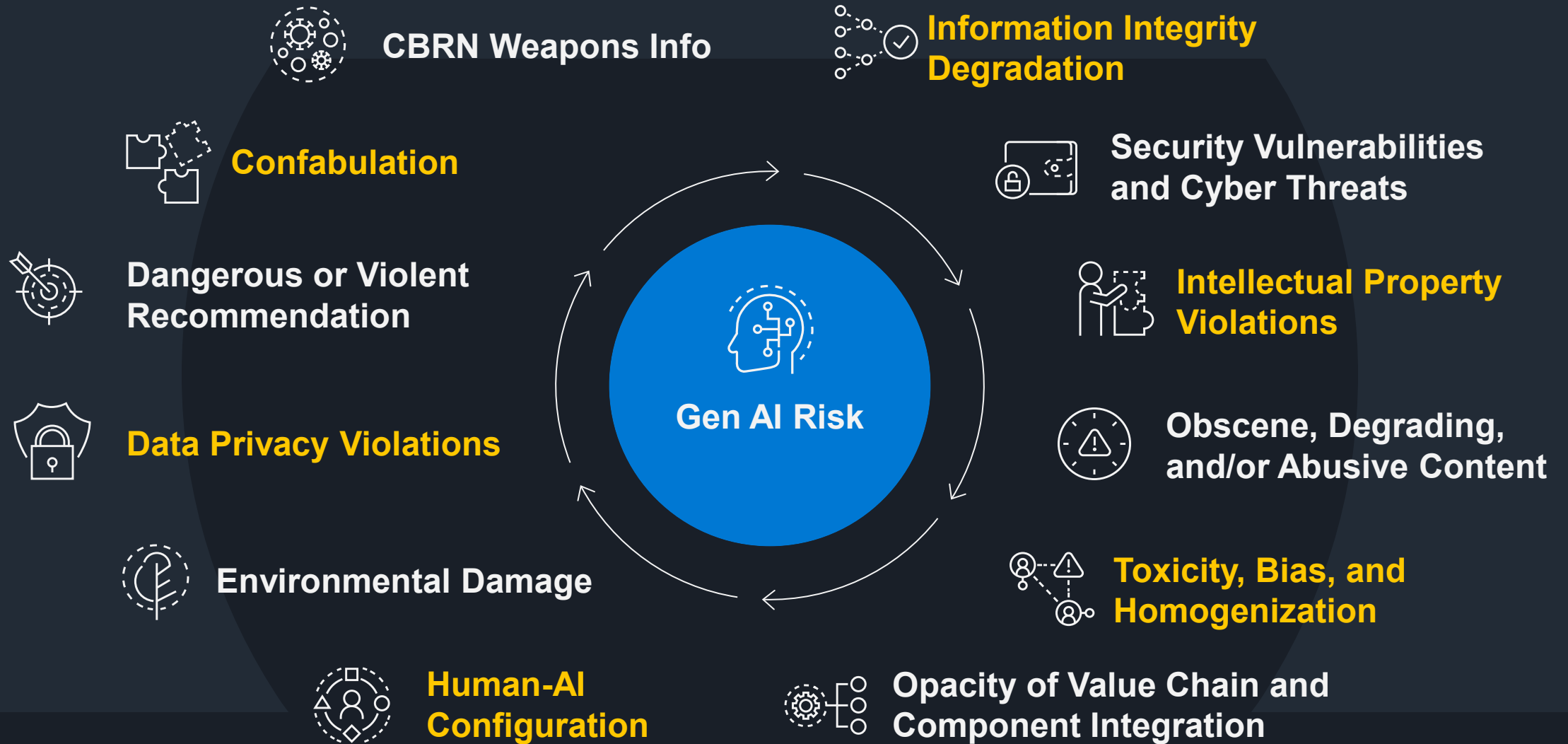


# Complex Risk: AI Risk Differentiators



# Generative AI Risk Categories

Source: NIST Generative AI Profile, <https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf>



# What is Data?

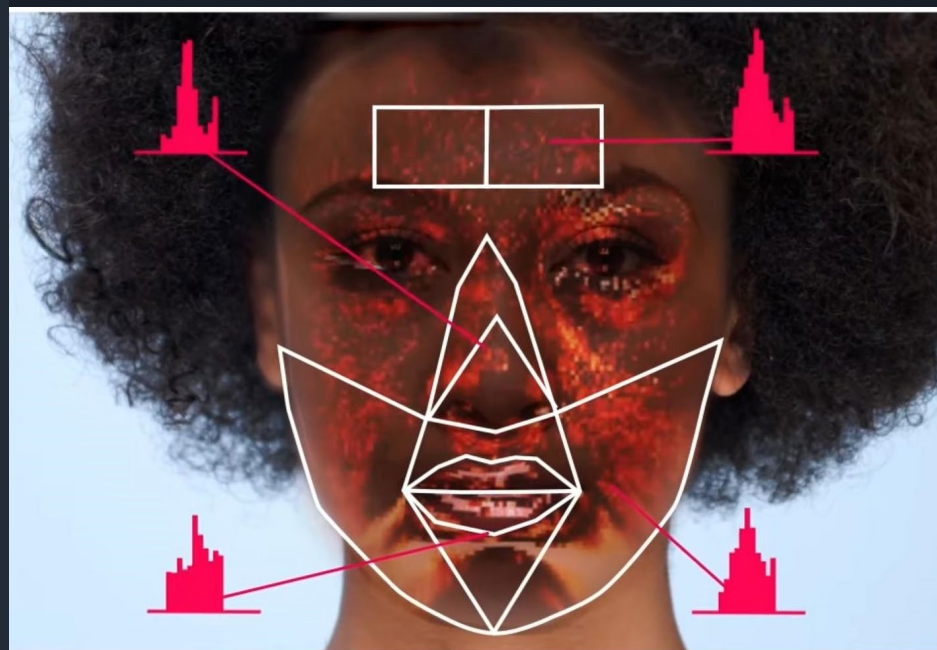
AI will reshape the data available to insurers and the safeguards around it

A cell phone “selfie” becomes...



## Transdermal Optical Imaging

- High blood pressure/hypertension
- Emotional state/stress levels



Source: <https://futurism.com/neoscope/blood-pressure-tech-analyze-selfie>

# NIST AI Risk Management Framework Core

Composed of Four Functions

Designed to help organizations **manage AI risk** and **promote trustworthiness** while designing, developing, deploying, or using AI systems



1

## Govern

A culture of risk management is cultivated and present

2

## Map

Context is recognized and risks related to context are identified

3

## Measure

Identified risks are assessed, analyzed, or tracked

4

## Manage

Risks are prioritized and acted upon based on a projected impact

# AI Use Guiding Principles

NIST and NAIC recommendations for gold standards

## NIST

- Fair – harmful bias managed
- Accountable and Transparent
- Explainable and Interpretable
- Safe
- Secure and Resilient
- Privacy-Enhanced
- Valid and Reliable

## NAIC

- Fair and Ethical
- Accountable
- Transparent
- Safe, Secure, and Robust
- Compliant



# Transitioning to Business Analytics for Risk Mitigation

Expanding on structured frameworks, a Bayesian-net-driven modeling approach creates the ability to make business decisions for risk mitigation

## Structured Framework

- Defines risk areas and enables risk register reviews.
- Helps identify relevant risks and assess control effectiveness.
- **Akin to a checklist**
- **Silent on Risk Tolerance – a prerequisite for effective decision making**
- **Necessarily broad – priorities may not align and focus can be lost**

## Importance of going beyond structured framework

- Lug nuts more important than air filters
- Model connections between processes, risks, and opportunities.
- Provide a comprehensive view of risks and opportunities.
- Highlights the financial impact of controls, threats, and opportunities.
- Supports data-driven decisions for risk reduction investments.

# Q&A

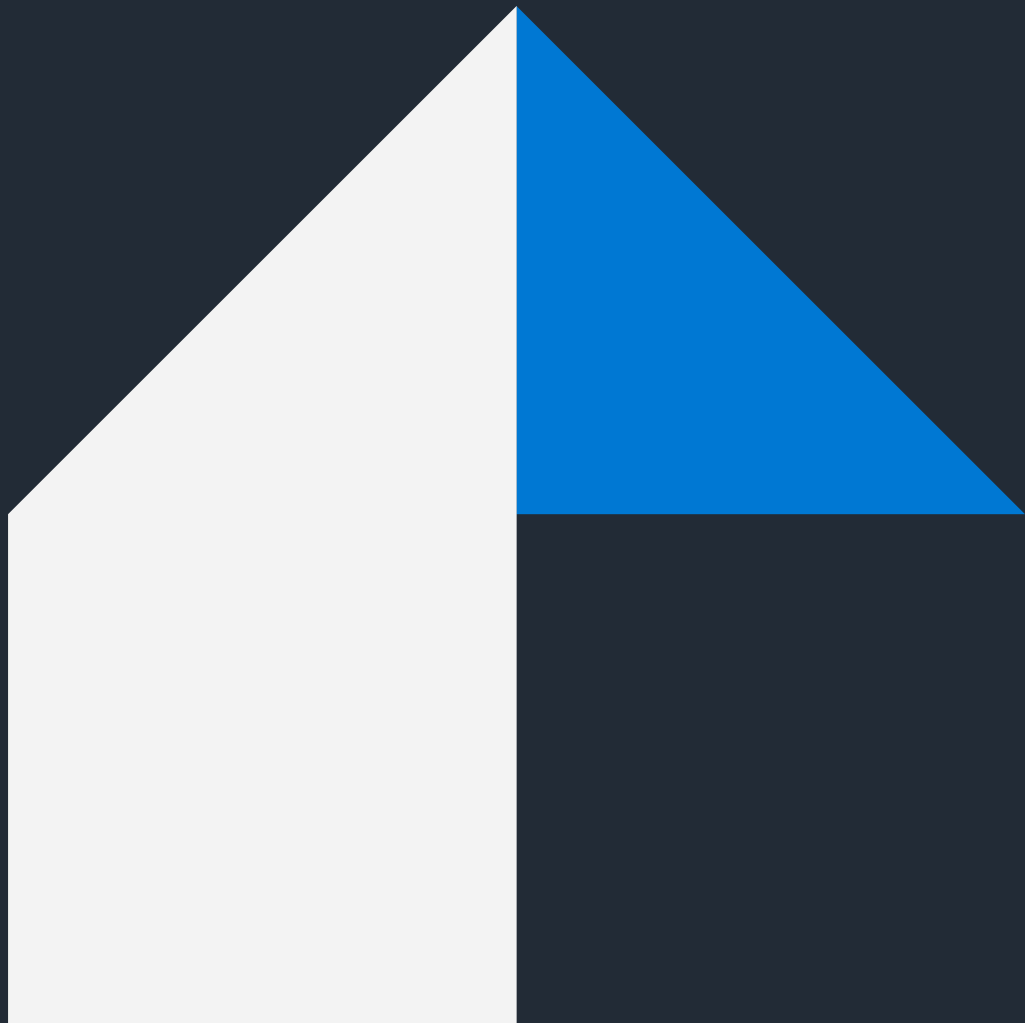




# Thank you

Tom Prince, FCAS, MAAA

[tom.prince@milliman.com](mailto:tom.prince@milliman.com)





# AI Systems Evaluations Working Group

# What's Next?

In Process

Options to further update regulatory framework  
(i.e., Possible Next Steps)

Third Party  
Oversight

AI System  
Evaluation

Regulatory Filings

Prohibited  
Practices/Attributes

Required Practices  
(Disclosure to  
consumers)

Other  
Considerations



How should this work proceed?

# AI Systems Evaluation and Training Collaboration Forum

## Draft Charges Based on Needs from Working Groups

1. Identify existing tools, resources, materials, and training that will assist and guide regulators in their review of AI Systems used by licensees, including an insurer's AI Program. This includes establishing a coordinated work plan and timeline for further development of those resources.
2. Develop new regulatory tools or regulatory guidance to assist regulators in their review of AI Systems used by licensees, including an insurer's AI Program.
3. Coordinate the development of review and enforcement tools, resources, guidelines, and training related to AI Systems for regulators across the NAIC.

# AI Systems Evaluations - 2024/2025 Plan



## How do we evaluate AI Systems?

- **Short-term:** What initial tools, resources, and education is necessary to meet the immediate needs of enforcing the Bulletin and assess the use and risks of AI?
- **Longer-Term:** How should an overall AI regulatory framework be developed—incorporate into the Market Conduct Exam Handbook or as a stand-alone Handbook?

## Possible Timeline & Coordination

### **AI Systems Evaluation**

2024

Understand evaluations work from other groups, determine need for short-term tools

2025

Discuss market regulation process and recommend updates to D Committee

2026 &  
beyond

Support implementation of proposals