American Academy of Actuaries

Issue Brief

Artificial Intelligence Governance: What Counts as AI?

NOVEMBER 2026

Key Points

- Governance is better guided by the risks a tool presents than by whether it carries an AI label. Factors that raise the oversight needed include learned behavior, probabilistic outputs, generative content, limited explainability, consumer-facing use, reduced human oversight, post-deployment adaptation, and unclear data provenance. These characteristics tend to accumulate as systems become more autonomous and opaque, so oversight should scale proportionally rather than depend on classification.
- A risk-based approach can protect consumers while supporting innovation. Rather than settling the taxonomy debate, stakeholders can build a shared set of risk characteristics to scope inquiries and controls. A central question guides the effort: What risks existed before a tool was introduced, and how have they changed? This keeps regulators focused on the highest-risk applications and lets insurers adopt new modeling approaches responsibly.
- Whether conventional models such as GLMs count as AI depends on who is asked. Regulators favor broad definitions to avoid oversight gaps; actuaries, developers, risk managers, and marketers draw the line differently. The result is inconsistent answers across insurers, heavy reporting on decades-old tools, added cost, and the risk that novel tools get buried.

American Academy of Actuaries
1850 M Street NW, Suite 300
Washington, DC 20036
202-223-8196 | www.actuary.org

Geralyn Trujillo
Senior Director, Public Policy

© 2026 American Academy of Actuaries. All rights reserved.

Introduction

Whether predictive models used in insurance, especially Generalized Linear Models (GLMs), should be considered "artificial intelligence" (AI) is a point of confusion in the current effort to develop regulations and controls for AI.

As states seek to develop AI governance frameworks, overly broad definitions risk sweeping in conventional tools that have been used for decades into regulatory regimes designed for far more complex and adaptive systems. On the other hand, excluding entire classes of models could create a gap in oversight. This paper proposes a path forward for discussion by putting technical definitions aside and focusing on practical considerations around the risks presented by a tool.

Definitions and Descriptions

The following definitions provide context for this paper.

Statistical Model	A simplified mathematical representation of relationships among real-world variables, entities, or events that uses statistical concepts and assumptions to describe data, estimate relationships, quantify uncertainty, or predict outcomes.[1]
Predictive Model	A statistical or computational model used to estimate an unknown or future outcome from available information. The label describes the model's purpose; by itself, it does not determine whether the model is classified as Machine Learning or Artificial Intelligence.[2]
Generalized Linear Model (GLM)	A family of Statistical Models with three core elements: a response variable whose distribution is drawn from the exponential family, a linear predictor, and a link function connecting the response mean to that predictor. GLMs are often used as Predictive Models.[3]
Machine Learning (ML)	A broad category of methods in which model parameters or behavior are estimated from data to improve performance on a task, rather than every relationship being explicitly encoded as a rule. Machine Learning is commonly treated as a subset of Artificial Intelligence, although classification conventions and regulatory definitions vary.[4]

1 "Actuarial Standard of Practice No. 56: Modeling"; Actuarial Standards Board (ASB); December 2019.
2 "Actuarial Standard of Practice No. 56: Modeling"; Actuarial Standards Board (ASB); December 2019.
3 "CAS Monograph No. 5: Generalized Linear Models for Insurance Rating"; Casualty Actuarial Society (CAS); 2016.
4 "Model Bulletin: Use of Artificial Intelligence Systems by Insurers"; National Association of Insurance Commissioners (NAIC); Dec. 4, 2023.

Artificial Intelligence (AI)	A machine-based system that, for explicit or implicit objectives, infers from its inputs how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. AI systems vary in autonomy and adaptiveness after deployment. Whether a conventional GLM falls within a particular AI definition depends on the definition and context; the label alone does not determine the appropriate level of governance.[5]
Generative AI	A class of Artificial Intelligence models that emulate characteristics of input data to generate derived synthetic content, including text, images, video, audio, code, or other digital content. Generative systems commonly use predictive inference as part of content generation.[6]
Agentic AI	An Artificial Intelligence system capable of pursuing objectives by planning and executing a sequence of actions, potentially including interactions with tools, data sources, or other systems, with some degree of autonomy.[7]

A Parable

Before diving into the discussion, let's begin with a parable.

A chef visits a botanist at his home. The botanist asks the chef if she would enjoy some fruit as a snack. The chef accepts. The botanist hands her a tomato. The chef is confused and clarifies that she wants to eat fruit. The botanist explains that, technically, a tomato is a fruit. The visit is off to a poor start.

Clarifications

The definitions section notes that both Generalized Linear Models and General Language Models share the acronym GLM.

Whether a particular GLM falls within a definition of Machine Learning or Artificial Intelligence depends on the definition and context. Some taxonomies place data-estimated statistical models within Machine Learning and define Machine Learning as a subset of AI. Actuarial practice often distinguishes conventional GLMs from newer Machine Learning systems because a GLM's variables, structure, and deployed parameters are deliberately specified, documented, and governed. This paper does not attempt to resolve that taxonomy. It focuses on the risks presented by a tool and the controls appropriate to its use.

5 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
6 "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile"; National Institute of Standards and Technology (NIST), NIST AI 600-1; July 2024.
7 "AI Agent Standards Initiative"; National Institute of Standards and Technology (NIST); 2026.

Members of the AI, Data Science, and Analytics Committee who authored this paper are: Tyson Mohr, MAAA, FSA, ACAS; Spencer Sadkin, MAAA, FCAS; Mario DiCaro, MAAA, FCAS; Melissa Hollis, MAAA, FSA; David McMichael, MAAA, FCAS; Bobby Jaegers, MAAA, FCA, ACAS; and Wanchin Chou, MAAA, FCAS.

The Problem: Miscommunication

Currently, when insurers are asked for information about their AI systems during market conduct exams, individual stakeholders interpret the scope of the request based on their own understanding, which leads to different individual interpretations. This makes it challenging for the parties to have a constructive conversation and can lead to confusion and inefficiency. The following example illustrates how this might occur.

Four Tools—Theoretical Example

Suppose an insurer is asked to name all the AI systems used in rating automobile insurance and provide several pages of information on each system. The insurer has four tools, labeled A through D below, and needs to determine whether they are in scope:

A. A multivariate analysis that calculates the guideline factors for model year, gender, and driver age. Actuaries select the factor to implement in each state based on the guidelines. Prior to the 1990s, using a method called "minimum bias procedures," these factors were calculated manually by iteratively attempting to balance all three factors across the book of business. In 1990, the insurer determined that the GLM function could be used to perform this multivariate analysis more efficiently and accurately (this is often called an "Actuarial GLM" or "Companywide GLM").[8]

B. A predictive model for determining whether to categorize new business as substandard. It is a logistic regression model that predicts whether applicants are likely to have a loss frequency over an established threshold. It was developed by a third-party vendor who specializes in building predictive models for underwriting. It has been used for 15 years.

C. A predictive model for rating. It is a GLM that produces a score based on predicted loss and expense costs for all policies. It was filed with state regulators and has been used for 10 years.

D. A tool that reviews filing objections and drafts responses. Actuaries use these responses as starting points for their actual responses. This tool started being piloted last year.

8 The Werner-Modlin "Basic Ratemaking" text explains this use of a GLM. The minimum bias procedures "involve performing several iterations of univariate analysis on rating variables, each time adjusting for the exposure weight and the indication of the previous variable in the sequence … The minimum bias procedures are not technically multivariate methods, and they were not necessarily based directly on statistical theory. However, many of the minimum bias procedures are actually a subset of the statistical method, GLMs. In fact, iterating the minimum bias procedure a sufficient number of times may result in convergence with GLM results, though many would argue the minimum bias procedures involve less computational efficiency" (p. 173-4).

How might various parties classify these tools?

Regulator: The regulator's preferred definition of AI explicitly scopes in all types of predictive modeling. The regulator prefers a broad definition to err on the side of inclusion and identify any risks of causing unfair discrimination or other adverse consumer outcomes. They consider all these tools in scope.

Actuaries: The actuary considers tools A through C to merely estimate relationships between variables and predicted outcomes. They find no characteristics that the actuary attributes to AI: they produce the same result every time given the same inputs, they are transparent, and they do not learn over time. The actuary does not consider them in scope.

The actuary considers Tool D to be AI but sees it as an internal tool used for efficiency that does not impact the rates charged to policyholders. They do not consider it to be "used in rating" and hence also not in scope of the request.

Model Developers: Technically, the skills used to conduct statistical analysis and develop predictive models are distinct from the skills used to develop generative or agentic AI. Tool A was built by actuaries using standard statistical analysis tools. They consider themselves to have produced a set of rating factors more efficiently than the previously available methods, but they do not view this as remotely resembling AI. The statisticians who built tools B and C, at both the insurer and the third-party vendor, likewise believe referring to their tools as "AI" misrepresents their skills.[9] AI engineers who developed Tool D are comfortable calling it AI because it is an LLM.

Model Risk Management: Imagine an insurer whose core model risk management practices were set up over the last 15 years to improve internal controls and meet regulatory/audit expectations. Two years ago, a separate set of AI policies were developed that apply to models that employ certain specified techniques. Tools are categorized as either AI or not AI so that the appropriate standards can be applied. Third-party and internal tools are treated the same. In the insurer's model inventory, only Tool D is categorized as AI.

Marketing: From a marketing perspective, it is important to emphasize to consumers that the insurer is sophisticated and competitive. While the individual tools would not be mentioned, marketers look for opportunities to convey that the rating plan is "driven by AI insights" and hence would be inclined to include all the tools. The third-party vendor also markets Tool C as being AI.

9 While developing this paper, an AI practitioner said he finds whether something is AI to be similar to how he views magic tricks. When you first see a magic trick, it looks like magic. But when you learn how the trick is performed, it doesn't seem like magic anymore. His experience with AI is similar—tools cease to seem like AI to him once he understands more about how they work.

The Impact

In practice, the scope of this inquiry could be much broader. It could cover many insurance functions and thousands of tools.

A lack of clarity about the scope of the question can lead to:

- Internal disagreement about categorization within the insurer.
- Depending on the scope, large amounts of information being produced on tools that have been in place for decades.
- Inconsistent information being delivered to the regulator across insurers.
- The regulator reviewing a large amount of information that is primarily about long-standing insurance practices; the novel risks presented by a tool such as Tool D may be obscured.

In summary, a lack of clarity about the scope of AI inquiries negatively impacts insurers and regulators by creating friction and, increasingly, workload. The resulting expenses are ultimately borne by policyholders and taxpayers. Technical definitions have not proved helpful in resolving this issue.

An Alternative Approach: Risk Characteristics That Influence Governance Requirements

To move beyond arguments related to definitions and techniques, an alternative approach is to discuss what risks are presented by AI systems that may not be sufficiently addressed by existing regulation and controls. The goal is to ensure that all parties are confident that relevant risks have been identified and mitigated through whatever combination of human judgement and technology is appropriate for the situation. The following outlines how this might occur.

This discussion uses GLMs as an example because this is a modeling technique for which parties may disagree about categorization as AI.

The level of governance applied to an analytical model or AI system should be proportional to the level of risk it presents. Rather than categorizing technologies as "acceptable" or "high risk," a risk-based framework evaluates characteristics that increase uncertainty, reduce oversight, or amplify the potential for harm. Systems exhibiting more of these characteristics generally warrant stronger governance, validation, documentation, monitoring, and regulatory scrutiny. This is true regardless of whether the system is completely automated or completely performed by humans.

Debates regarding whether a particular model should be classified as "AI" are often less informative than understanding the risks the model presents. Models built using similar techniques can have very different risk profiles depending on their use case, level of autonomy, explainability, consumer impact, and degree of human oversight. The key governance question is often not whether a model is definitionally AI, but what risks it introduces and what controls are needed to manage those risks.

The following characteristics illustrate factors that may influence the level of governance appropriate for a particular model or AI system. Stakeholders can collaborate to develop and refine such a list.

Business Rules vs. Machine Learning

Business rules are explicitly defined by humans and follow predetermined logic. The relationship between inputs and outputs is directly observable and can be readily inspected, tested, and modified.

Modeling methods lie on a continuum rather than in two mutually exclusive categories. In many traditional actuarial models, including GLMs commonly used for pricing and reserving, the modeler deliberately specifies the variables, structure, distribution, and link function. These choices can be documented, reviewed, and challenged, and the actuary must be prepared to defend their relationship to expected loss. Some Machine Learning methods automate more of the feature selection, interaction discovery, or representation learning process, which can make model behavior more difficult to document or interpret. These distinctions are relevant to governance even when a broad taxonomy classifies both approaches as Machine Learning or AI.

As systems move from manually defined rules toward learned behavior, governance requirements increase due to the need for model training oversight, validation, performance monitoring, drift detection, and periodic retraining.

Deterministic vs. Probabilistic Behavior

A system executes deterministically when the same inputs, parameters, state, and operating conditions produce the same output. This allows a reviewer to reproduce an output and verify that the implemented system is operating as intended.

A Statistical Model can represent a probabilistic relationship while still executing deterministically when producing a point estimate or score. Stochastic operation is different: it uses randomness or probabilistic sampling, so repeated runs under otherwise

equivalent conditions may produce different outputs. Keeping these concepts separate avoids describing a conventionally implemented GLM as operationally stochastic merely because it includes a probability distribution.

In addition, some AI systems may produce different outputs when presented with the same inputs. This can make independent verification more difficult and may create additional challenges for model validation, audit, and regulatory review.

As variability in system behavior increases, governance becomes increasingly important in establishing appropriate validation, monitoring, testing, and controls.

Predictive vs. Generative Outputs

Traditional actuarial and predictive models estimate a future outcome, such as claim frequency, mortality, lapse rate, reserve adequacy, or expected loss.

Generative AI systems produce derived synthetic content, including text, images, code, summaries, recommendations, or analyses. Generative systems commonly use predictive inference as part of that process, so "generative" and "predictive" are not mutually exclusive technical categories. The governance distinction arises primarily from the form, use, and risks of the output, including confabulation or hallucination, misinformation, inappropriate content, and disclosure of confidential information.[10]

Explainability vs. Transparency

Governance risk generally increases as a system becomes more difficult to understand and explain.

Models with transparent logic allow actuaries, reviewers, regulators, and auditors to better understand how outputs are produced. For example, a premium calculated using a GLM can typically be expressed as a base rate multiplied by a defined set of relativities, allowing a reviewer to identify the factors that contributed to an outcome and the impact of each factor. As complexity increases, it may become more challenging to identify errors, validate assumptions, explain outcomes, or assess the presence of unintended bias.

Third-party tools may suffer from decreased transparency as the developer takes steps to protect its intellectual property.

As explainability decreases, governance becomes increasingly important in providing visibility into model design, assumptions, limitations, and controls.

10 "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile"; National Institute of Standards and Technology (NIST), NIST AI 600-1; July 2024.

Internal Operational Use vs. Consumer-Facing Use

The context in which a model or AI system is used can significantly influence its risk profile.

Many organizations deploy AI for internal operational purposes, such as employee productivity, document summarization, software development, research assistance, workflow automation, or operational analytics. While these uses may create risks related to security, privacy, intellectual property, or operational reliability, they typically do not directly affect consumers.

Governance requirements generally increase as AI systems move from supporting internal operations to influencing customer-facing decisions and outcomes.

Degree of Human Oversight

Risk generally increases as decision-making authority shifts from humans to automated systems such as agentic AI.

Models used solely to inform a human decision-maker present different governance concerns than systems that automatically execute actions or make decisions without meaningful review. Human oversight can serve as an important control mechanism, particularly in higher-risk applications.

On the other hand, human oversight can also introduce subjectivity, cognitive biases, and a lack of transparency. For example, allowing an automated eligibility decision to be overridden by a human without a clearly documented justification could lead to favoring an influential applicant inappropriately. The goal of governance is to balance the outcomes of the entire system (human plus technology) along dimensions such as cost, errors, fairness, and transparency.

Fixed vs. Adaptive Models

Traditional actuarial models are generally implemented as "fixed" models. This means that once approved and placed into production, the variables, structure, and parameters remain unchanged until a formal review, update, or refiling occurs. As a result, the model operating in production is the same model that was validated, approved, and documented.

Some Machine Learning models and AI systems adapt after deployment by changing their parameters or behavior in response to new data or feedback. A deliberate, governed refit or vendor update creates a new fixed version; it is not the same as autonomous or continuous adaptation. As deployed behavior becomes more adaptive, governance becomes increasingly important to provide ongoing validation, monitoring, version control, testing, and transparency regarding changes.

These risk characteristics are not mutually exclusive and often accumulate within the same system. As analytical models evolve from predictive applications toward generative or agentic capabilities, additional risk characteristics may emerge without replacing those that already exist. As a result, governance requirements often build upon one another as system complexity, autonomy, opacity, and potential impact increase.

Known vs. Unknown Data Provenance

Predictive models can use structured or unstructured data and static or changing data sources. In many conventional actuarial applications, however, developers work with well-defined datasets and have substantial visibility into their sources, contents, transformations, and limitations. That visibility enables assessment of whether the data is appropriate for the model's intended use.

For other models, such as many modern AI tools, it can be unclear to the model developer and user what training data was used. The model could have broad access to the internet, raising privacy, intellectual property, accuracy, and cybersecurity concerns. The data may be unstructured (e.g., images, sounds, or handwritten documents). The data used to develop a model may not be representative of the population to which it will be applied, creating risks related to performance, fairness, or unintended outcomes. The data used to build third-party tools may also be less transparent.

As data provenance becomes less transparent or less well understood, governance should scale appropriately to ensure data sources are suitable, risks are understood, and limitations are recognized.

Existing Controls

Effective regulation encourages innovation while ensuring that governance is in place to avoid abuse or liability. Many of the risk characteristics discussed above may already be addressed, in whole or in part, through existing governance. AI-specific governance is necessary only to the extent that the following controls are not effective at mitigating the risks.

- **Regulatory filings:** Rating models and other types of models (e.g., underwriting models in some states) must be filed with state regulators. There are standard filing exhibits, such as GLM Questionnaires, which provide transparency on models. The models presented in filings are subject to external review, including by the NAIC Model Review team.

- **Market conduct exams:** State regulators have broad authority to examine insurance operations. This includes inquiring into how models are used to support the insurance functions. Regulators use these exams to probe business practices that may cause unfair discrimination and adverse consumer outcomes.

- **Model risk management (MRM) programs:** Insurers use MRM practices to help ensure that models work as intended at development and on an ongoing basis. Typical practices cover model purpose, design, data, methodology, validation, monitoring, documentation, change management, human involvement, manual overrides, and accountability. These practices are also promoted by auditors, rating agencies, and regulators. As one cross-sector reference point, the Federal Reserve, OCC, and FDIC issued Revised Guidance on Model Risk Management in 2026 (SR 26-2), superseding SR 11-7 and emphasizing a risk-based approach commensurate with model risk and use.

- **Actuarial Standards:** Actuaries evaluate rating models for their ability to predict expected losses and expenses.[11] Actuarial standards of practice (ASOPs) also define actuaries' responsibilities with regard to model development, including ASOP No. 12—*Risk Classification Systems*, ASOP No. 23—*Data Quality*, and ASOP No. 56—*Modeling*. There is currently no ASOP explicitly focused on AI, so the existing ASOPs must be used to interpret actuarial responsibilities while working with AI.

11 See Principle 4 of the CAS Ratemaking Principles: "A rate is reasonable and not excessive, inadequate, or unfairly discriminatory if it is an actuarially sound estimate of the expected value of all future costs associated with an individual risk transfer."

Conclusion

A lack of shared understanding about the scope of AI regulation is causing miscommunication among insurance industry stakeholders. Regulators prefer broad definitions to ensure they have transparency into any tools and practices that could adversely affect consumers. Insurers prefer a narrower definition of AI that recognizes the effectiveness of existing regulation of conventional actuarial methods and analytical tools, as well as the need for regulatory certainty across the states. This increases regulatory friction and expenses, makes it harder for regulators to focus on the most important use cases, and doesn't serve consumers.

This paper presents a way to advance this conversation by recognizing that technical differences are less important than risk characteristics. Many traditional actuarial techniques, including GLMs, already operate within mature governance frameworks that address issues such as validation, documentation, monitoring, and change management. As organizations adopt new modeling approaches and AI capabilities, governance and regulation should be based on the risks presented by the application rather than the technology itself. A core question is: What risks were present in this system before this tool was introduced, and how have these risks changed? This approach aligns oversight with risk while supporting continued innovation.

Appendix

GLM as AI Vocabulary

Purpose: These working definitions support this paper? They are synthesized from the cited sources and are not presented as authoritative definitions of the American Academy of Actuaries. Where legal, regulatory, technical, and professional taxonomies differ, the definition or note identifies the relevant scope issue.

Classification note: A model may be statistical, predictive, machine-learned, and within a particular definition of AI at the same time. Those labels describe overlapping characteristics; they do not by themselves determine the governance appropriate for a specific use.

Statistical Model: A simplified mathematical representation of relationships among real-world variables, entities, or events that uses statistical concepts and assumptions to describe data, estimate relationships, quantify uncertainty, or predict outcomes.[12, 13]

Predictive Model: A statistical or computational model used to estimate an unknown or future outcome from available information. The label describes the model's purpose; by itself, it does not determine whether the model is classified as Machine Learning or Artificial Intelligence.[14, 15]

Generalized Linear Model (GLM): A family of Statistical Models with three core elements: a response variable whose distribution is drawn from the exponential family, a linear predictor, and a link function connecting the response mean to that predictor. GLMs are often used as Predictive Models.[16]

Machine Learning (ML): A broad category of methods in which model parameters or behavior are estimated from data to improve performance on a task, rather than every relationship being explicitly encoded as a rule. Machine Learning is commonly treated as a subset of Artificial Intelligence, although classification conventions and regulatory definitions vary.[17, 18, 19]

12 "Actuarial Standard of Practice No. 56: Modeling"; Actuarial Standards Board (ASB); December 2019.
13 "What is Process Modeling?"; NIST/SEMATECH e-Handbook of Statistical Methods, National Institute of Standards and Technology (NIST); 2003.
14 "Actuarial Standard of Practice No. 56: Modeling"; Actuarial Standards Board (ASB); December 2019.
15 "Machine Learning Glossary"; Google for Developers.
16 "CAS Monograph No. 5: Generalized Linear Models for Insurance Rating"; Casualty Actuarial Society (CAS); 2016.
17 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
18 "Model Bulletin: Use of Artificial Intelligence Systems by Insurers"; National Association of Insurance Commissioners (NAIC); December 4, 2023.
19 "Machine Learning Glossary"; Google for Developers.

Artificial Intelligence (AI): A machine-based system that, for explicit or implicit objectives, infers from its inputs how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. AI systems vary in autonomy and adaptiveness after deployment. Whether a conventional GLM falls within a particular AI definition depends on the definition and context; the label alone does not determine the appropriate level of governance.[20, 21]

Generative AI: A class of Artificial Intelligence models that emulate characteristics of input data to generate derived synthetic content, including text, images, video, audio, code, or other digital content. Generative systems commonly use predictive inference as part of content generation.[22]

Agentic AI: An Artificial Intelligence system capable of pursuing objectives by planning and executing a sequence of actions, potentially including interactions with tools, data sources, or other systems, with some degree of autonomy.[23, 24]

Deterministic Execution: Operation in which the same inputs, parameters, system state, and operating conditions produce the same output. A model can have probabilistic assumptions while still executing deterministically when producing a point estimate or score.[25, 26]

Stochastic Operation: Operation that incorporates randomness or probabilistic sampling, so repeated runs under otherwise equivalent conditions may produce different outputs. This is distinct from a Statistical Model that represents uncertainty but produces deterministic scores when implemented.[27, 28]

Adaptive Model: A model whose deployed parameters or behavior can change after deployment in response to new data or feedback. Adaptiveness can be periodic or continuous and may invalidate assurances established for an earlier version of the model.[29, 30]

20 "Model Bulletin: Use of Artificial Intelligence Systems by Insurers"; National Association of Insurance Commissioners (NAIC); December 4, 2023.
21 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
22 "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile"; National Institute of Standards and Technology (NIST), NIST AI 600-1; July 2024.
23 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
24 "AI Agent Standards Initiative"; National Institute of Standards and Technology (NIST); 2026.
25 "What is Process Modeling?"; NIST/SEMATECH e-Handbook of Statistical Methods, National Institute of Standards and Technology (NIST); 2003.
26 "Deterministic Algorithm"; Dictionary of Algorithms and Data Structures, National Institute of Standards and Technology (NIST).
27 "What is Process Modeling?"; NIST/SEMATECH e-Handbook of Statistical Methods, National Institute of Standards and Technology (NIST); 2003.
28 "Deterministic Algorithm"; Dictionary of Algorithms and Data Structures, National Institute of Standards and Technology (NIST).
29 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
30 "Explanatory Memorandum on the Updated OECD Definition of an AI System"; Organisation for Economic Co-operation and Development (OECD), OECD Artificial Intelligence Papers, No. 8; March 2024.

Fixed (Non-Adaptive) Model: A deployed model whose parameters and behavior remain unchanged until a deliberate, governed update, refit, or replacement occurs. A model that is periodically refitted under change control is fixed between approved versions rather than autonomously adaptive.[31]

Autonomous Learning: Post-deployment adaptation in which a system changes its model or behavior based on new information without a person explicitly implementing each change. The degree of autonomy and the controls over adaptation should be described rather than inferred from the model family alone.[32, 33]

Neural Network: A computational model composed of connected units arranged in layers, with learned weights and transformations that map inputs to outputs. A deep neural network contains more than one hidden layer.[34]

Emergent Capability: A system-level capability or behavior that arises from interactions among system elements and is not specified as an individual rule. Claims of emergence should identify the observed capability and evidence rather than assuming that complexity or scale alone establishes it.[35]

Transparency: The availability of meaningful information about a model or system, including its purpose, design, data or inputs, assumptions, capabilities, limitations, operation, and deployment, at a level appropriate to the stakeholder and context.[36]

Explainability: The ability to provide understandable information about how a particular prediction, recommendation, content item, or decision was produced, including the factors and logic material to the outcome. Explainability concerns a specific outcome; Transparency is broader.[37]

Model Risk Management (MRM): The governance, policies, roles, controls, development practices, validation, monitoring, documentation, and effective challenge used to identify and manage risks arising from models throughout their lifecycle, with rigor commensurate with the model's risk and use.[38]

31 "Explanatory Memorandum on the Updated OECD Definition of an AI System"; Organisation for Economic Co-operation and Development (OECD), OECD Artificial Intelligence Papers, No. 8; March 2024.
32 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
33 "Explanatory Memorandum on the Updated OECD Definition of an AI System"; Organisation for Economic Co-operation and Development (OECD), OECD Artificial Intelligence Papers, No. 8; March 2024.
34 "Machine Learning Glossary"; Google for Developers.
35 "Emergence"; Computer Security Resource Center (CSRC) Glossary, National Institute of Standards and Technology (NIST); n.d.
36 "Transparency and Explainability"; OECD AI Principles, Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
37 Ibid.
38 "Supervisory Guidance on Model Risk Management"; Board of Governors of the Federal Reserve System, SR 11-7; April 4, 2011.

Model Validation: An objective evaluation of whether a model performs as expected and is reliable for its intended use, including assessment of its conceptual soundness, assumptions, methods, data, implementation, performance, and limitations. Validation rigor should reflect the model's use, materiality, and risk.[39]

Human Oversight: Human involvement and authority sufficient to understand and monitor a system, interpret its outputs, decide not to use or rely on it, disregard or override an output, intervene in its operation, and remain accountable for its use. Human-in-the-loop is one form of Human Oversight, not a synonym for the entire concept.[40, 41]

Hallucination (Confabulation): A Generative AI output that confidently presents erroneous or false content, diverges from the prompt or other input, or contradicts prior statements in the same context. The output need not be wholly fabricated to be a hallucination.[42]

Reversibility: The degree to which a system's output, action, or resulting effect can be identified, challenged, stopped, overridden, corrected, withdrawn, or otherwise undone. Reversibility depends on both system design and the surrounding human and operational process.[43, 44]

Editorial Approach

Definitions are paraphrased or synthesized for this paper rather than reproduced as quotations unless expressly indicated. Sources were selected primarily from actuarial standards and publications, insurance regulators, NIST, OECD, the European Union, and current interagency model-risk guidance. NIST AI 100-3 informed the decision to acknowledge multiple meanings where terminology is not standardized.

39 Ibid.
40 "OECD AI Principles"; Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).
41 "Regulation (EU) 2024/1689 of the European Parliament (Artificial Intelligence Act)"; Official Journal of the European Union; July 12, 2024.
42 "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile"; National Institute of Standards and Technology (NIST), NIST AI 600-1; July 2024.
43 "Regulation (EU) 2024/1689 of the European Parliament (Artificial Intelligence Act)"; Official Journal of the European Union; July 12, 2024.
44 "Transparency and Explainability"; OECD AI Principles, Organisation for Economic Co-operation and Development (OECD); May 22, 2019 (updated May 2024).

The American Academy of Actuaries is a 20,000-member professional association whose mission is to serve the public and the U.S. actuarial profession. For more than 60 years, the Academy has assisted public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.

This paper was developed thanks to the work of the Academy's volunteers. To learn more about becoming a volunteer, please visit actuary.org/volunteer.