Artificial Intelligence Governance Framework

General Actuarial Practice

DRAFT

International Actuarial Association Artificial Intelligence Task Force Deliverable (2024)

Internal IAA Document

Artificial Intelligence Governance Framework

General Actuarial Practice

A comprehensive governance framework

on Artificial Intelligence Actuarial Work

DRAFT

This document was prepared by the Governance workstream of the Artificial Intelligence Task Force (AITF) of the International Actuarial Association (IAA). The AITF was established to implement the Statement of Intent for IAA Activities on Artificial Intelligence, as adopted by the IAA Council on 8 March 2024.

Please note the following:

- This document is currently a draft and may be adjusted and/or combined with other documents and has not been subject to the formal due process required for it to be considered an official publication of the IAA.
- This document is intended to encourage understanding and debate of the issues raised therein. It is not an International Standard of Actuarial Practice (ISAP), nor does it set standards or requirements which any individual or organization is expected to consider or observe, or with which they are expected to comply. This is the case notwithstanding any language in the paper which, but for this clause, might suggest otherwise. This statement takes precedence over any such wording.
- This document is intended solely for internal use of the IAA and its member associations and is not for public distribution. It may not be circulated, quoted, or referred to in any public or external forum without prior written consent. The content by does not necessarily represent an official position of the IAA.

For any questions or comments, please contact us at: aitf.comments@actuaries.org

Please visit the AITF Website for more information.

International Actuarial Association

605 - 75 Albert Street, Ottawa, Ontario K1P 5E7
Tel: +1-613-236-0886
Email: secretariat@actuaries.org
www.actuaries.org

© International Actuarial Association/
Association Actuarial Internationale

TABLE OF CONTENTS

DRAFT

DRAFT

1. Introduction

1.1. Purpose

This document aims to provide material that helps and educates actuaries in safeguarding responsible Artificial Intelligence (AI), while raising awareness of the risks that must be managed when implementing, deploying, and using AI systems. Actuaries have long been at the forefront of managing uncertainty, utilizing a combination of skills in areas such as probability theory, advanced mathematics, statistics, and, importantly, ethics. As key players in decision-making within the financial industry and the field of social protection, actuaries' concerns about managing risks appropriately and contributing to societal well-being have become even more relevant with the rise of AI. AI is no longer an emerging topic; it has already found its place within the actuarial profession, particularly in data analytics, predictive modeling, and risk management. Given the distinctive nature of the actuarial profession - encompassing technical skills, ethical standards, and professionalism - actuaries are well-positioned to contribute to the development of AI systems and to oversee the AI landscape as a whole. While this paper will continue to evolve over time, the responsible use, distribution, and implementation of AI systems have yet to be fully adopted and embraced by the industries in which actuaries work. This document will build upon traditional governance areas impacted by AI, and the following sections will describe best practices for governing and mitigating risks related to data, modeling, and the outcomes produced by AI systems once deployed. Note that throughout this document, the terms “AI”, “AI systems” and” AI models” may be used interchangeably.

1.2. Importance of AI Governance

The importance of developing and adopting a governance framework for a more AI-centric landscape is crucial to actuarial work. The complexity of AI introduces a range of challenges. One major concern is the potential for bias in algorithms, particularly regarding gender or ethnicity. This issue is amplified by the large datasets used to train AI systems, which often incorporate historical, human-influenced data, including language, societal feedback, and human interactions. Unintentionally, algorithms can inherit and perpetuate biases from these training materials, as is frequently observed in large language models (LLMs). Since models do not "understand" the underlying issues, they are shaped by how data is processed, the design of the model, implementation decisions, and output configurations. For instance, the pricing and risk assessment processes actuaries are involved in may inadvertently capture protected characteristics of customers. Consequently, when bias is inherited from data sources, it can result in both direct and indirect discrimination.

Firstly, gaining a comprehensive understanding of how to manage and mitigate the risks associated with the use of data and modeling practices is increasingly important. Robust governance structures and controls are essential to ensure that the use of data and the implementation of actuarial models are conducted appropriately. Having an AI governance framework in place will help address potential errors, identify weaknesses, maintain consistency across actuarial services, and ultimately support better decision-making. Consistently applying this governance framework throughout an organization will foster productive dialogue among relevant stakeholders, providing high-quality information, models, and assumptions that are fit for purpose. Secondly, an AI governance framework can help actuaries meet regulatory requirements specific to their organization's region and sector. Increasingly, regulators around the world are focusing on the risks associated with AI system implementation, particularly in areas such as data governance, model governance, testing, and validation. By acting as facilitators of proper governance framework implementation within their organizations, actuaries can further contribute to mitigating existing risks, enhancing productivity, and improving organizational efficiency. This, in turn, allows for a stronger focus on delivering high-value analysis aligned with business needs.

2. Key Components of the Framework

To understand the key components of AI governance frameworks, actuaries may reference regulatory guidance for model governance (e.g. regulations established within the U.S. banking sector for model risk management (MRM) first promulgated by the Board of Governors of the Federal Reserve System in 2011, establishing an initial standard for a bank's adoption of model risk management practices.) Since then, many financial services firms throughout the world, including insurance companies, have developed model risk management frameworks that have adopted or adapted such guidelines to align with the size and complexity of their organizations. Foundational to these frameworks is the definition of a model, which often references the components of a model that the model governance framework addresses, namely an input component (e.g. data and assumptions), a processing component (e.g. a calculation engine or algorithm), and an output component (e.g. report or data and information relied upon by users of the output.).

For illustration, the IAA already provides a definition for a model. It is clear from this definition that many AI systems would be captured within this definition and thus be subject to the expectations set out in established model governance frameworks.

Model governance, in its purest form, refers to the roles and responsibilities of senior leaders responsible for the oversight of model development and use within an organization and the policies that guide them. Such roles and responsibilities typically extend to those charged with the effective challenge of models within an organization, such as within a "3 lines" structure, including leaders from the business, risk management, internal audit, and the organization's board of directors.

However the term "model governance framework" often extends beyond the purest form of governance and includes the processes and procedures related not only to a model's development, implementation, ongoing use and eventual retirement, but also to the risk management processes that aim to mitigate model risk throughout a model's lifecycle, capture an organization's model risk appetite, and conduct internal audits that confirm that the organization's management of model risk is performed in a manner consistent with expectations.

Organizations that have established model governance frameworks (used interchangeably with model risk management frameworks) have typically introduced corporate policies or guidelines for the purpose of clearly articulating model risk management expectations, including: a) the roles and responsibilities associated with model governance including escalation protocols, b) key terms used within the framework and their definitions, c) the organization's model risk rating methodology for the purpose of identifying "high risk"

DRAFT

models, d) key governance and risk management processes, e) the applicability of the framework to third party models, and f) expectations for model documentation and model risk reporting.

Actuaries should consider the applicability of any internal model governance policies or practices when developing, managing the risk of, or using AI systems. In some cases, actuaries will observe that the AI system they are currently developing or using or planning to develop or use does not fit the definition of a model and thus would not be swept into an existing model governance framework. In this circumstance, governance guidelines for non-model AI may separately exist as a reference or the model governance framework could be consulted for guidance on good governance practices. In addition, actuaries may turn to existing data governance frameworks, often adopted in accordance with regulations established for the safe and secure use of data.

The following sections highlights the key components of a model governance framework while highlighting actuarial considerations when developing or using AI systems.

2.1. Roles and responsibilities

As stated, a strong model governance framework begins with the clear articulation of roles and responsibilities of the senior leaders involved in the development or appropriate use of models, including the management of model risk. Such an articulation of roles might include: the board of directors, senior management committees, the Chief Risk Officer (CRO, business functional leaders responsible for the use of models, model owners who have the responsibility for the appropriate development and use of the model throughout its lifecycle, and the head of Internal Audit.

Typically, strong model governance program will reflect a risk-based view, with model risk oversight following a three-level approach:

1st Level	Strategy/Oversight	Board of directors
2nd Level	Management	Committees & Policies, Executive Management, Key Functions,
3rd Level	Operational	IT & Business Management

The rapid evolution and the usage of AI systems may require an AI related adjustment of the existing governance structure, depending on the new material risks arising as a result of the use of AI systems within an organization. In addition, new roles and responsibilities may need to be established to effectively manage the widespread use of AI throughout the organization.

Actuaries currently hold a variety of leadership roles with the responsibility for the development and appropriate use of high-risk models including AI systems. In some cases, actuaries serve as members of teams reporting to such leaders. As a result, the actuary involved in the development or use of AI should understand the organizations expectations for the management and mitigation of model risk at all levels within the organization.

Board of directors

Board-level executives bear the ultimate responsibility for the use of AI and thus for appropriate governance framework which includes developing a new and/or updating

DRAFT

existing governance frameworks (e.g. IT, risk management including model risk management, etc.), with clear communication about the strategy for and policies relating to the use of AI within the organization. In order to accomplish this, board level executives must have sufficient knowledge of how AI is being used and its potential risks. For this reason, the board should receive adequate and continuous training on AI risks and opportunities and consult with available and appropriate subject matter experts. In addition, relevant policy documents approved by the board provide the framework for the development and the usage of AI in the organization (e.g. goals, principles, values, processes, requirements, responsibilities, etc.). The risk management framework approved by the board should address all material AI activities and their related risks.

Committees & Policies

For existing senior management committees consisting of key governance roles (e.g. CRO, CFO, functional leaders etc.), organizations may need to revise current policies and procedures to ensure the development and use of ethical and trustworthy AI. Especially for high impact AI, an entity may evaluate whether it is worthwhile to implement a standalone AI / Data / Ethics Committee, which takes care about the development, deployment or procurement of AI systems or whether these responsibilities are best managed by existing committees. Regardless of the decision to create new or utilize existing committees for the oversight of AI, an organization should consider including a minimum number of members with a "fit and proper" profile on AI.

Risk Management Function, Actuarial Function, Internal Audit, Compliance Function

For the above key functions, it is obvious that AI-related issues must be integrated in the corresponding positions in an adequate manner respecting the very different roles and responsibilities of each function. Given a high-risk AI System, the impact on the Risk Management (RM) Function (including model risk management) could be very different compared to the impact on the Compliance Function. While the RM Function has to find a measurement for the risk, which could be very time consuming and complicated, the Compliance Function need to ensure, that the deployment and the operation of the AI System is in line with existing internal policies and external regulations. In principle the need for the adjustment for each function can be evaluated using an appropriate GAP-Analysis.

It’s important to note that some organizations are sufficiently small to not have formal model risk management practices in place. In these instances, internal audit may adopt certain “second line” function to evaluate the effectiveness of governance and controls over high-risk models, including AI, systems in accordance with the organization’s policies and guidelines.

Model Owner

The model owner owns all aspects of the model, including development, performance monitoring, compliance, and maintenance. They ensure the model meets specific business objectives, aligns with strategic goals and business objectives, complies with regulations and internal policies, and remains effective in its purpose. and provides value throughout the model’s entire lifecycle, from inception through decommissioning. They are responsible for the data within the models they develop or use and must approve how that data is defined and utilized. Usually, the model owner is responsible for the complete value chain (cf. illustrative example below).

DRAFT

Responsibilities of the model owner could include:

- Define the responsibilities for the model stakeholders including data scientists, IT, compliance, and business units
- Assign stewards responsible for maintaining, updating, and monitoring models throughout their lifecycle
- Establish and/or approve workflows for key stages of model development, such as pre deployment, post-deployment, and significant updates
- Set up accountability structures to ensure that the models are used responsibly and in alignment with organizational values
- Assemble the right team (data scientists, compliance officers, IT security, and business leaders) to support governance and works closely with the team to ensure the model adheres to internal policies and external regulations
- Regularly update stakeholders on model performance, risks, and any significant changes
- Decides when the model should be retrained, updated, or retired, based on performance and relevance
- Manage models and approve their definitions and requirements
- Review and approve specifications related to the model standard and revisions to ensure that any key changes are sufficiently understood and their impact fully assessed and integrated
- Ensure consistency and quality across different types of models and data (i.e. actuarial assumptions)

The model owner's responsibility includes the alignment with the IT-Security-Officer, who is responsible for safeguarding digital assets, network and data from cybersecurity threats. Furthermore, they have to ensure that data and model security are embedded in the information security management framework of the entity. Model security includes the input as well as the output of the model and the complete process of data processing during the workflow of the model.

2.2. Model risk ratings

An essential tool within a strong model governance framework is the model risk rating methodology. When appropriately developed and applied, this tool enables the actuary to determine whether the AI system that they are developing or using, initially determined to be within the scope of the model governance framework, is categorized as high, medium, or low risk. While risk rating methodologies vary by organizations, they typically include risk criteria such as adverse financial impact and degree of complexity among other characteristics to assess risk. Model governance frameworks often have more significant governance requirements for high-risk models, with low-risk models typically requiring minimal independent oversight.

DRAFT

Since AI systems pose different risks to organizations and their customers than do traditional models, model risk rating methodologies likely have or need to evolve to adapt to the changing model risk environment. For example, such methodologies might consider 1) the degree of transparency and explainability, 2) the level of autonomy, and 3) the degree of reliance on third party data or systems.

2.3. Key governance and risk management processes

As shared, model governance frameworks articulate key governance and risk management processes and practices by functional area, with a particular focus on those relating to high-risk models.

The model risk management processes captured within a model governance framework typically include: a) model risk identification processes via the application of the risk rating methodology, which are frequently performed in cooperation with functional leaders and model owners; b) model risk assessments via independent model validations and ongoing (model performance) monitoring; c) risk mitigation via the resolution of validation outcomes (if any) and the establishment of additional model control processes; and d) risk reporting to business leaders and the board of directors.

Actuaries that are developing AI systems or performing independent validation of these systems will need to focus more heavily on input data and outcomes testing than when developing or independently validating traditional actuarial models. See section 3 for a discussion of several unique considerations when developing or validating AI systems, including addressing data quality and privacy, the transparency and explainability of both data and algorithms, and outcomes testing.

2.4. Independent validation of an AI model

Validation in the context of this governance framework is intended to provide documented evidence that a model shows certain desired properties (e.g. generalization, accuracy) and does not show undesirable properties (e.g. bias, overfitting). This validation should be performed by individuals who did not develop the model, unless this imposes a burden that is disproportionate to the model risk. Evaluation criteria such as performance, stability, and reproducibility are consistent across these types of models, but validation of AI may require an actuary to adapt validation methods to the specific type of AI. This adaptation may be required for e.g. the usage of unstructured data, previously manual steps are performed automatically, the degree of explainability, dynamic update of the model or missing transparency. Some further validation aspects are listed below.

Aspect	Explanation
Goals and requirements	Validating AI models are similar to general model validation criteria. These criteria include: - Model behavior, e.g. reliability, ethical aspects - Model's strengths, e.g. robustness - Model's weaknesses, e.g. biases in the input data - Model's assumptions, e.g. consistencies
Validation quality	Provide documented evidence that a model shows certain desired properties (e.g. accuracy and adequacy, completeness and

	consistency) and does not show undesirable properties (e.g. bias, overfitting). Validation should be performed by individuals who did not develop the model, unless this imposes a burden that is disproportionate to the model risk.
Validation Method	Tools and methods required to validate AI models depend on the characteristics of the AI model. Some are - Trained vs. pre-Trained models (because of huge data usage) - Observability of stability in training and inference - Type of unpredictability - Frequency of model modification and/or recalibration - Severity and consequences of model results
Model validation checklist	Before starting the validation, the scope should be clear, which could be defined via a checklist. This includes e.g. - *Which acceptance criteria must be applied?* - *Does the model fit these acceptance criteria?* - *How undesirable model behavior is treated?* - *How is uncertainty quantified in the model output?* - *How, known bias in the training data is handled?* - *Are stress, sensitivity or scenario tests appropriate?* - *Which burden (manpower) is acceptable?* - *What is necessary for ongoing validation activities?*
Limitations of validation	Ultimately, the quality of validation is measured by how "well" the adequacy of the models can be assessed. With the help of the validation methods used, it must be possible to assess whether the model fulfills the purpose for which it is to be used. In accepting the validation results inherent uncertainty levels about the quality of the model must be tolerated

DRAFT

The model is trained, validated and tuned using the training and the validation data sets respectively. The main objective of this step is to ensure that the model generalizes well to unseen data upon its deployment in the real world. For this purpose, multiple experiments should be carried out on the model using all three data sets – training, validation, and testing to bring out its best abilities and minimize the changes it undergoes post-deployment.

2.5. Applicability of framework to third party vendor AI models and data

Third party or vendor models do not need to follow this exact framework, but it is the actuary's responsibility to confirm that the governance standards of the vendor are sufficient for the purpose of the model. All items in this document should be considered when using a third party or vendor model, and additional information may need to be requested to ensure that the model fulfils, and continues to fulfil, the standards for actuarial use. Necessary due diligence should be conducted by the relevant stakeholders and governance needs to be defined in the contracts. Third party data should be checked for sufficient representation across groups and existing historical biases. Data privacy, accuracy, regulatory compliance should be adhered to, and ownership clarity should be clearly defined. For external models, the actuary should independently validate the model results for accuracy, interpretability, fairness and alignment with actuarial standards. Models need to be auditable and vendor accountability should be established through contractual terms.

When using third-party data or vendor models, actuaries must ensure that the governance standards employed by the vendor are adequate for the intended purpose. While vendors may not follow the same frameworks actuaries use, it is the actuary’s responsibility to confirm that the model meets ethical, regulatory, and actuarial standards. This involves thoroughly assessing the vendor's governance processes, validating the model's assumptions, and ensuring data quality and transparency. Additional documentation or assurances may be required to confirm the model aligns with actuarial principles and is suitable for deployment.

Ongoing oversight is essential to ensure the model continues to perform effectively and ethically. Actuaries should establish processes for monitoring model outputs, identifying and addressing issues, and requesting updates or recalibrations as conditions evolve. Collaboration with vendors is critical to securing necessary documentation, testing outputs for fairness and accuracy, and negotiating any required modifications to align the model with actuarial use cases.

Ultimately, actuaries remain accountable for the use of vendor models and must apply professional judgment to evaluate their adequacy. By adhering to a proportional governance approach - tailored to the risk and impact of the model's application - actuaries can ensure reliable and fair outcomes while maintaining compliance with actuarial standards and regulatory requirements.

2.6. Human supervision and oversight

Establishing appropriate human oversight of AI systems is critical to ensuring they function as intended and do not produce adverse effects. Human oversight involves direct involvement in the design, operation, maintenance, adaptation, or application of AI systems. While AI increasingly automates tasks and processes, human involvement remains necessary at all stages of the AI model lifecycle to provide checks and balances.

Both human and AI decisions can be biased - AI due to potential bias in training data, and humans due to inherent predispositions. Consequently, a well-calibrated balance between human judgment and AI automation is essential for certain tasks to mitigate risks and improve outcomes.

The design of human oversight should be proportionate to the specific AI use case, reflecting the nature, scope, and complexity of the associated risks, while considering existing governance frameworks. The method of oversight depends on the stage of the AI application lifecycle. During the design phase, developers must address biases in training datasets and determine the appropriate level of automation for use cases, such as pricing or underwriting, where human review may be required. Once operational, oversight shifts to monitoring daily processes, controlling system performance, addressing incidents, and adapting to changes as per established procedures. This may involve employees reviewing key metrics to assess the AI system's impact on vulnerable groups or requiring human validation of AI outputs before implementation.

DRAFT

3. Governance Over an AI Model‘s Lifecycle

3.1. Overview

Strong model governance is not achieved by a series of procedures that are applied after an AI system has been developed, but rather as integrated practices and procedures that are applied within every phase of a model's lifecycle, including the design, development, implementation, ongoing monitoring and ultimately retirement of the AI system. Such practices and procedures are designed to achieve a common set of objectives, including fairness and non-discrimination, safety and security, robustness, and compliance with regulations, Actuaries have an essential role to play in establishing and maintaining strong model governance over AI systems.

For the purpose of establishing and maintaining strong model governance, actuaries should be aware about the fundamental characteristics of three categories of AI models:

1. Transparent AI: Statistical models (e.g. GLM, GAM, Decision Tree), and other models in which the calculation of the output is transparent and easily explained to the public.

2. Explainable AI: GBM, Neural Networks, vector support machines, and other models in which the relative importance of input characteristics can be explained, but the exact calculation of an individual output is opaque and not easily explained to the public

3. Opaque AI: BART, Large Language Models, image recognition, and other models in which the relative importance of input characteristics and the calculation of an output is not easily explained to the public.

In reality, an AI model may exist on a spectrum between Transparent and Opaque. For example, a decision tree with a single split is considered transparent, while a decision tree with thousands of splits may only be considered explainable due to its complexity.

As a result, the type of model chosen will have a significant impact how readily the organization's objective of transparency and explainability may be achieved, and the types of testing, validation, and ongoing monitoring procedures that need to be applied. Actuaries will need to be focused on the unique governance considerations of the types of AI systems chosen during the AI system's design and development phases and throughout the model's lifecycle. Model governance it is particularly necessary to ensure the model is able to adapt to a changing environment without many changes in its anticipated results. When a model is deployed in the real world, the data fed to it becomes very dynamic. Apart from the data, there might be changes in technology, business goals, or a drastic real-world event like a pandemic that have an impact on the AI model's performance. Frequent refreshments of the AI model is necessary to keep up with changes in data, technology, and regulatory expectations. And when the AI system is no longer fit for purpose, model governance ensures that the model will be appropriately retired.

DRAFT

DRAFT

3.2. Designing the AI system

Designing an AI system and the associated processes is a critical step to ensure the effectiveness of a strong governance framework. A well-thought-out design begins with determining the alignment between the AI system's capabilities with business needs. When developing specific AI models, effective collaboration and cooperation between teams within an organization are essential. Actuaries, who play a central role in governing the applicability of models, must engage with development teams from the outset. This early collaboration ensures that the model's lifecycle comprehensively addresses key considerations, such as identifying limitations and clearly defining intended outcomes.

Furthermore, using AI responsibly involves navigating a range of complex challenges. For instance, bias in AI systems may stem from factors such as ethnicity or gender. Since AI is trained on extensive datasets influenced by human behavior, language, and societal norms, algorithms can inadvertently inherit and perpetuate the biases present in their training data, as observed in Large Language Models (LLMs). These models lack an understanding of underlying issues and rely heavily on predictive methodologies shaped by their design, inputs, and outputs. As a result, the rights of individuals—such as women or other marginalized groups—can be disproportionately affected by the way these algorithms operate.

For actuaries, who are responsible for risk assessment models like pricing or reserving, it is crucial to ensure that protected characteristics are adequately considered. This proactive approach will lead to a more ethical and effective application of AI systems.

The following sections will explore fairness and transparency in greater detail.

3.2.1. Fairness and bias

Unfair and biased algorithms result in both direct and indirect discrimination. While the definition of discrimination is highly contextual and varies on a case-by-case basis, institutions and regulators worldwide have attempted to standardize such definitions. For instance, the Organization for Economic Co-operation and Development (OECD) has recently published[1] recommendations to clarify definitions of AI systems, focusing on responsible stewardship and international cooperation for trustworthy AI. Similarly, the European Commission[2] has developed the Ethics Guidelines for Trustworthy AI, and the Monetary Authority of Singapore previously issued principles aimed at promoting Fairness, Ethics, Accountability, and Transparency[3] (FEAT).

Other related policies and laws that may positively impact the functioning of AI systems include regulations on data protection (GDPR in Europe), insurance distribution (IDD in Europe), and commercial practices (UCPD, still to be verified).

Direct discrimination is relatively straightforward to define, as it involves unfavorable treatment of individuals by misusing their protected characteristics as a factor. Indirect discrimination, however, is more complex. In actuarial pricing practices, features or proxies that do not directly involve protected characteristics may still be used in models. These

[1] OECD, May 2024, Recommendation of the Council on OECD Legal Instruments Artificial Intelligence [1]

[2] Ethics guidelines for trustworthy AI, 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [2]

[3] Monitoring Authority of Singapore, Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector [3]

variables can inadvertently reflect protected characteristics, potentially disadvantaging certain policyholders.

AI systems are subject to feedback loops and dynamic data collection, which can lead to alterations in model behavior over time. If models influence the underlying data, an algorithm initially considered free of discrimination can eventually evolve into one that is discriminatory[4].

Some examples of vulnerabilities to consider when addressing discrimination, or factors that affect vulnerable groups, include the following:

- **Associated Characteristics**: Age group, low income or poverty, low level of education, migrants, young people (e.g., students).
- **Lifestyle Factors**: Unemployment or homelessness, divorced or single parents, over-indebted persons, prison inmates, accident victims, victims of domestic violence, etc.
- **Health-Related Factors**: Disabilities, hereditary diseases (e.g., genetically determined conditions), individuals with mental illnesses, etc.
- **Digital Illiteracy**: Lack of digital skills, difficulties accessing online or digital services, or challenges understanding services provided online.

If dealing with such factors and related characteristics, individuals may encounter numerous cumulative obstacles to financial inclusion. For example, actuaries need to implement appropriate preventive measures within their modelling frameworks.

Effective and transparent data governance is crucial for ensuring fair and non-discriminatory treatment of consumers. This involves key elements such as transparency, explainability, and robust data management practices. Additionally, principles like robustness and performance, as well as human oversight, are essential for fostering a responsible and meaningful approach to AI, helping to address deficiencies within the data. Actuaries should design and refine algorithms that prioritize explainability and transparency to uphold fairness, while steering clear of practices that might unduly influence consumer behavior.

A significant challenge in promoting fairness and non-discrimination in AI systems lies in understanding and evaluating the impact of corrective measures implemented to improve outcomes.

So, when designing and setting-up actuarial AI models, or systems, one should interlink bias and discrimination to fairness.

3.2.2. Transparency and Explainability

The professional use of AI comes with a set of responsibilities, including transparency, explainability, accountability (to be discussed in section 3.5 Implementing the AI system), and robustness. Ensuring that actuarial professional ethics align with these responsibilities is crucial, and adopting a sound system-building governance will result in an explainable and transparent AI. Moreover, addressing the challenges of fairness and discrimination is imminent.

[4] [4] European Union Agency for Fundamental Rights, Bias in algorithms: Artificial intelligence and discrimination: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf [4]

DRAFT

Actuaries need to deal with multiple stakeholders; therefore, ensuring these concepts are addressed accordingly will result in better decision-making and well-understood processes.

Transparency issues can lead to opaque systems that are difficult to validate and can have consequences for all stakeholders and customers involved. It is crucial that the way a system works and how it is documented are clearly communicated to all parties involved. Disclosing the right amount of information will ensure that the system is well understood, uses the right data sources, is correctly implemented, thoroughly tested, and its outputs are clearly explained. In some regulatory environments, such as under Europe's AI Act, GDPR, or Solvency II, traditional actuarial models are already somewhat aligned with transparency requirements. Generally speaking, actuarial models have often struggled with being fully disclosed or well presented to a wider audience. This issue stems partly from the complexity of such models and the variety of assumptions needed based on business requirements. These challenges are further amplified when adopting AI models, where the degree of randomness and uncertainty inherent in AI system operations adds another layer to the methodology. Therefore, it is important to clearly outline the intended purpose of the model, highlight the accuracy of the algorithms used, the robustness of the infrastructure on which the model operates, the appropriate software testing methods, and the necessary documentation that reflects human intervention, choice of assumptions, and the degree of uncertainty (consider feature selection, model drift, etc.). Ultimately, one should be able to replicate the results of a model comfortably.

Explainability in AI often pertains to how model choices and assumptions are manifest in the system's reasoning process and contribute to achieving the intended output. The degree of explanation required varies by application. For example, some AI chatbots that do not significantly impact decision-making and pose no consequences to individuals may not require extensive disclosure; their outputs are generally self-explanatory. However, more complex models necessitate robust explainability measures that align closely with the system’s intended use and the methodology for deriving its outputs. Numerous explainability methods exist, though they are often complex to implement and require a deep academic understanding. Actuaries are uniquely positioned to serve as a bridge in this context, possessing the expertise to comprehend complex mathematical methods and the dynamics of their specific business cases. Many current methods utilize visualization techniques and can be categorized into two types: local and global metrics. Local metrics, such as Individual Conditional Expectation (ICE), Local Interpretable Model-agnostic Explanations (LIME), and Shapley Additive Explanations, often focus on the influence of individual features on predictions. Global metrics, on the other hand, assess how features affect the system overall and include tools like Partial Dependence Plots, Feature Importance Stability, or Fairness metrics.

Conceptuality, explainability, and transparency are integral to fairness and non-discrimination in AI systems, serving as key prerequisites for identifying issues within these areas. These elements foster trust and good governance, particularly in financial services, making explainability a fundamental component for ethical AI use. The following sections will provide an overview of key issues in developing an AI System.

3.3. Developing the AI System

Actuaries who are part of an AI system development team may be involved in some or all of the AI system’s development steps, from designing the AI system to meet the organization’s

DRAFT

business objectives, to training and evaluating the model, and to documenting the model. The actuary should be aware of the unique model governance objectives of each step of the AI development process.

In some organizations, actuaries may be part of AI development teams that have adopted formalized DevOps methodologies or are using DevOps systems. Such methodologies and systems focus on both software development" Dev" and technical operations "Ops", and typically guide or support developers with data preparation, training and testing, automated deployment, and ongoing monitoring. In such cases, actuaries will need to ensure that the organization's model governance policies and practices are being consistently applied within the DevOps environment.

3.3.1. Gathering and Preparing the Data

Since AI models are only as accurate as the data fed to them, it becomes crucial to identify the right data (sources, types and formats) to ensure model accuracy and relevance. It is the responsibility of the AI system's Model Owner and development team to ensure that it has access to data that is appropriate for the AI system's intended purpose and that there are no regulatory, or ethical constraints with respect to the collection and/or use of the data.

When actuaries are gathering and preparing data for the development of an AI system or using an existing AI system for their purpose, they should understand the data governance practices of their organization, namely the policies, processes, and practices that have already developed to guide how data is managed, used, and protected. The goal is to ensure data quality, security, privacy, and overall compliance with policies and procedures.

Key Components of a Data Governance are:

1. Policies and Standards - guidelines defining data management rules, including access, sharing, and retention.
2. Ownership and Stewardship – the assignment of roles (e.g., data owners, stewards) to ensure accountability and proper data handling.
3. Quality Management - ensuring accuracy, completeness, consistency, and timeliness of data.
4. Compliance Management – alignment with existent local regulations, or regulations that apply to the region in scope (e.g. Europe's General Data Protection Regulation, California Consumer Privacy Act etc.) Technology and Tools - platforms for data cataloging, monitoring, and compliance reporting.
5. Risk Management - identifying and mitigating risks, such as data breaches or non-compliance penalties.

Good data governance practices enable informed decision-making, enhance regulatory compliance, and reduce risks associated with poor data practices. Once the organization's data governance policies and practices are understood, actuaries who are responsible for the development or use of AI systems will need to ensure that they rely upon quality data that is unbiased and maintained in a secure environment with the appropriate data security measure in place.

DRAFT

Since datasets used for building models may come from multiple sources including third parties and have a variety of characteristics, such as: including personally identifiable information, being structured or non-structured, or including generated / synthetic data, the actuary should maintain strong data accountability practices which includes understanding the lineage of data and keeping a data provenance record.

Critically, actuaries developing or using AI systems need to focus on potential bias in data as it can lead to unfair, unethical, or inefficient outcomes when AI systems are deployed in decision-making processes. As discussed in section 3.2, addressing data bias is critical for ensuring that AI systems are transparent, accountable, and aligned with human rights and societal values. Such biases could include historical bias, sampling bias, label bias and measurement bias.

In addition, data sourced for the development of an AI system should meet the following, non-exhaustive requirements:

- be current – up to date
- be appropriate, given the intended use of the model
- be consistent – not contradictory
- be reliable - the origin should be known or traceable)
- be accurate and reliable
- be complete - records and/or attributes
- be authentic - not altered in any form
- be representative of the various constituency groups expected to use or be affected by the use of the model

Actuaries with the roles of Model Owner or serving as part of a development team should be transparent with respect to the collection of the data and should ensure that the data is secure (tamper-free) and that collection of the data has been authorized (requisite permission obtained) and use of the data is limited to the purpose for which it was collected.

Since different datasets should be used for the training, testing, and validation of the AI system, such datasets should be periodically reviewed and updated.

In addition to an actuaries' focus on an organization's data governance practices and data quality objectives, data privacy and data security are also key considerations in model development and are (at a minimum partially) the responsibility of the actuary as Model Owner and member of the development team.

Data privacy refers to the practice of ensuring that personal information is collected, processed, stored, and shared in a way that respects an individual's rights and expectations of confidentiality. It is a key component of data protection and cybersecurity, aimed at preventing unauthorized access, misuse, or exposure of sensitive information.

Key principles of data privacy include:

- Transparency - clearly inform users how their data is collected, used, and shared.
- Consent – obtain consent to use the data

DRAFT

- Data Minimization **-** collect and retain only what is needed for the specific purpose.
- Security – put in place adequate measures to protect data against breaches and unauthorized access.
- Accountability **-** demonstrate compliance with privacy laws and regulations, for example.

Data security refers to the measures and practices implemented to protect digital information from unauthorized access, corruption, theft, or destruction throughout its lifecycle. Effective data security not only protects sensitive information but also:

- Ensures compliance with legal and regulatory requirements.
- Preserves organizational reputation and trust.
- Mitigates financial losses from data breaches or cyberattacks.

Key aspects of data security are:

- Confidentiality **- e**nsuring that sensitive information is accessible only to authorized individuals or systems.
- Integrity - protecting the data from being altered or tampered with, ensuring its accuracy and trustworthiness
- Anonymization and pseudonymization.
- Availability - ensuring that the data is accessible to authorized users when needed.

When developing or using an AI system, actuaries need to introduce common data security measures including:

- Access Control - limiting access to data through authentication (e.g., passwords, biometrics) and authorization mechanisms.
- Encryption - converting data into a secure format to prevent unauthorized access during storage or transmission.
- Firewalls and Antivirus Software - protecting systems from malware, viruses, and unauthorized network access.
- Data Backup - regularly creating copies of data to ensure recovery in case of accidental deletion, cyberattacks, or hardware failure.
- Data Masking - obscuring sensitive data to prevent exposure during testing or processing.
- Auditing and Monitoring - tracking access and changes to data to detect and respond to suspicious activities.

Data robustness refers to the quality and stability of data used to train, validate, and test a model. It ensures that the model performs consistently across various conditions, even when exposed to noisy, incomplete, or adversarial data. Robust data are**:**

- Accurate - represent real-world scenarios as precisely as possible.

DRAFT

- Diverse - include diverse scenarios, covering edge cases and anomalies.
- Consistent - uniform in structure, format, and meaning.
- Complete - Missing data or incomplete records should be addressed appropriately.
- Resilient to Noise – the model should perform well despite the presence of errors or perturbations in the input data.

Data robustness can be enhanced through:

- Data Augmentation - Introducing variations (e.g., noise, transformations) to simulate diverse conditions.
- Outlier Detection and Handling - Identifying and managing anomalies to prevent skewed results.
- Use of Synthetic Data - generate data to fill gaps or simulate rare cases.
- Adversarial Testing - Exposing the model to intentionally perturbed inputs to identify vulnerabilities.

DRAFT

3.3.2. Training and evaluating the AI Model

This step of **training** and evaluating (testing and validating) the AI system relies upon training and testing data that has been gathered and prepared as discussed in 3.3.1, using an iterative process until the model performs as intended. For example, should the AI model perform poorly on the training data, the actuary will have to improve the model which can be achieved by selecting a better algorithm, increasing the quality of data, or feeding more data to the model. Should the model not perform well on testing data, then the model might fail to generalize well to unseen data.

Evaluation criteria such as performance, stability, and reproducibility are consistent across these types of models, but validation of AI may require an actuary to adapt validation methods to the specific type of AI. This adaptation may be required for several reasons:

- AI may involve the transformation of unstructured data.
- Previously manual steps such as variable selection and dimensionality reduction may be included in an automated fashion and must be evaluated.
- In addition to out of sample validation based on test statistics, explainability may or may not be essential to the actuarial or business purpose validation of a model. This explainability may be required by internal or external parties.
- It can be difficult to evenly compare models at different levels of transparency. An actuary may need to weigh predictive power vs. explainability in the validation process.
- The model is trained, validated and tuned using the training and the validation data sets respectively. The main objective of this step is to ensure that the model generalizes well to unseen data upon its deployment in the real world. For this purpose, multiple experiments should be carried out on the model using all three

data sets – training, validation, and testing to bring out its best abilities and minimize the changes it undergoes post-deployment.

Testing AI and ML models in actuarial work is essential due to the unique complexities and risks associated with these technologies. Unlike traditional actuarial models, which often rely on established statistical methods, AI and ML models can exhibit unpredictable behaviours and biases that may not be immediately apparent. Key elements to consider during testing include accuracy, fairness, robustness, and explainability. Actuaries must ensure that models not only perform well on historical data but also generalise effectively to new, unseen data. This requires a comprehensive approach to testing that includes defining clear objectives, creating diverse test cases, and ensuring high-quality data preparation to minimize bias and enhance reliability.

The role of the actuary in testing AI and ML models is multifaceted. Actuaries are responsible for validating model performance, assessing the ethical implications of model outputs, and ensuring compliance with regulatory standards. They must leverage their expertise in risk management to identify potential pitfalls in model design and implementation. This involves not only evaluating the technical aspects of the models but also understanding the broader implications of their use in decision-making processes. Actuaries should actively participate in the development of testing protocols and contribute to the establishment of governance frameworks that promote transparency and accountability in AI applications.

To effectively test AI and ML models, actuaries can employ a variety of techniques and metrics. Common approaches include functional testing, bias and fairness testing, and robustness testing against adversarial inputs. Metrics such as precision, recall, and F1 score can be used to measure accuracy, while statistical tests can help evaluate fairness across different demographic groups. Explainability techniques, such as SHAP (SHapley Additive exPlanations) values, can provide insights into model decision-making processes. Continuous testing and monitoring are also crucial, as they allow actuaries to track model performance over time, identify emerging biases, and implement necessary adjustments through feedback loops and retraining protocols.

The testing of AI and ML models in actuarial work requires a rigorous and structured approach that goes beyond traditional methodologies, which is why actuaries should continue learning about managing the risks of these models. Actuaries play a critical role in ensuring that models are reliable, fair, and ethically sound. By employing a range of testing techniques and metrics, they can navigate the complexities of AI and ML, ultimately contributing to more informed and responsible decision-making in the actuarial field. For more guidance on testing AI and ML models, please refer to [GUIDANCE NOTES ON TESTING].

3.3.3. Documenting the AI Model

Given the inherent complexity and lack of transparency of AI systems – often referred to as the “black box” effect – it is essential to document the key attributes of the model through the model life cycle. Documenting the AI system is crucial whether the model is developed internally or outsourced to third parties.

DRAFT

DRAFT

While documenting the model, it is important to adhere to the principle of proportionality, to ensure that the level of detail in the documentation is proportionate to the potential risk and implications of the AI system.

Categories	**Attributes**
Data	• Data flow and data inventory: Document the flow of data through the AI system, including sources of data, data transformations, and storage locations. Create a comprehensive data inventory that lists all datasets used, their origins, and how they are integrated into the model. • Data quality and data governance: Outline the processes and standards for ensuring data quality, including validation checks, data cleansing methods, and criteria for data acceptance; Describe the governance framework that oversees data management practices, including roles and responsibilities for data stewardship. • Data usage: Specify how data is used within the AI model • Data Limitations and weaknesses: Identify any limitations or weaknesses in the datasets used, such as biases, gaps, or inaccuracies. Document how these limitations may impact model performance and the steps taken to mitigate these issues. • Data privacy and security: Describe the measures in place to ensure data privacy and security, including compliance with relevant regulations; Document data encryption practices, access controls, and anonymization techniques used to protect sensitive information.
Model development	• Model overview: Provide a high-level overview of the AI model, including its purpose, functionality, and the problem it aims to solve. This should also include a description of the target audience and stakeholders involved. • Model selection, model training and validation: Document the criteria and rationale for selecting specific algorithms and techniques for the AI system. Detail the training process and validation techniques used. • Model performance and evaluation: Outline the metrics used to evaluate model performance; Document the results of performance evaluations. • Model interpretation: Provide explanations of how the model makes decisions, which helps stakeholders understand the model's outputs and the rationale behind them.

	• Model assumptions, risk and limitations: Clearly state the assumptions made during model development, the potential risks and limitations of the model. • Version control and management: Maintain a version history of the model, documenting changes made over time, including updates, improvements, and bug fixes.
Model deployment and maintenance	• Model deployment process: Describe the steps taken to deploy the model into a production environment. • Monitoring and feedback loop: Outline the plan for ongoing monitoring of model performance after deployment. • Model maintenance: Specify the maintenance plan for the model, including regular updates, quality checks, and the criteria for model adjustments.
Other disclosures	• Ethical considerations and legal considerations: Document any ethical considerations relevant to the AI model, including fairness, accountability, and transparency; Outline compliance with legal requirements, including data protection laws and industry regulations. • Reliance on other data or models: Specify any reliance on external data sources or other models that impact the AI system; Document the nature of this reliance, including how it affects the model's performance and any associated risks.

3.4. Seeking approval of the AI system

The formal approval of an AI system prior to its implementation and deployment is an essential part of a strong model governance framework. The type of approval required may vary based on the nature and intended use of the AI system as captured by its model risk rating. Higher risk AI systems may be subject to more rigorous review and approval processes. High-risk AI systems may be subject to independent model validation (see section 2.4), with final approval given by the appropriate governance committee once any exceptions produced as the result of the independent model validation are resolved.

Actuaries may be involved with this process as a member of the development team responsible for independent the validation team the information they need to conduct their review, as members of the independent validation team, or as members of the oversight committee responsible for the AI system's final approval.

3.5. Implementing the AI system

When implementing AI systems, accountability is key to their success. Accountability can refer to different areas when implementing an AI system - the <u>mechanism</u> put in place to ensure who is responsible and accountable for an implemented system; the <u>operational effectiveness</u> – ensuring the system is transparent and adheres to ethical standards;

DRAFT

auditability – ensuring the design, data, processes and algorithms are compliant, fair and accurate; and, redress – a very important aspect when implementing AI systems, referring to the human intervention, which at times should be able to address and correct any harm and errors caused by AI systems.

Trustworthiness, compliance, and accountability are essential for implementing AI models and require adherence to best practices. As DevOps methodologies were mentioned in Section 3.3 – Developing the AI System, following practice-based frameworks such as, but not limited to, MLOps (Machine Learning Operations) is crucial. MLOps frameworks cover the entire lifecycle management of a system, including development, deployment, monitoring, and retraining. By integrating DevOps practices, data engineering, and machine learning algorithms, an MLOps framework enhances the synergy between these domains. The CI/CD (continuous integration and continuous deployment) practices within this framework automate workflows across different organizational segments, particularly among various stakeholders. This automation ensures that roles and responsibilities are maintained, even when using concurrent models or scaling models to embrace technological innovations.

The benefits of employing a framework such as the MLOps framework when implementing AI systems are manifold. It fosters efficiency through streamlined collaboration among diverse disciplines such as actuarial sciences, data science, IT professionals, and lawyers. Automation of various stages—ranging from data preparation and feature engineering to training, testing, deployment, and monitoring—minimizes manual errors. Furthermore, it enables standardization of processes across different lifecycle stages of a model, ensuring consistency. Scalability is another advantage, as it allows AI systems to be dynamically adjusted based on business needs for enhanced performance. Lastly, compliance is ensured, as the AI models implemented adhere to the legal and regulatory requirements specific to their application region.

In addition to establishing a system management framework, it's essential to integrate the system into the organization's culture. Furthermore, as outlined in section 2.4 - Roles and Responsibilities, it's crucial to clearly define and thoroughly document the involvement and motivation of various stakeholders during the implementation phase. Moreover, the primary obstacle to digital transformation is often the lack of know-how. Actuaries must facilitate the interconnection of various disciplines and stakeholders, particularly when certain parties are less accountable for the system itself. This can be achieved by organizing workshops, creating straightforward documentation, and providing cross-functional training. Cross-functional training is an effective method for improving a team's skills, collaboration, and performance. It allows team members to experience different roles, functions, and perspectives within the organization and to learn from one another.

Given AI's scalability and adaptive nature, it's crucial to implement standards, best practices, and quality tools from the start. Actuaries must ensure AI systems comply with legal requirements early on to mitigate industry-specific regulatory risks and liabilities towards model users in case of defects.

3.6. Ongoing monitoring of the AI system

Continuous model monitoring and retraining are essential practices to maintain the performance, reliability, and relevance of AI models, especially as data distributions shift or

DRAFT

DRAFT

new information becomes available. AI systems inherit biases from the environment they operate in, therefore not directly accounting for potential harm. Thresholds need to be set for model performance metrics and drifts and alerts should be triggered.

Such thresholds need to account for well-defined KPIs, business objectives and a set of parameters that can be easily adapted to address improvements on the data and model quality. Once a model is being deployed into a well-established framework, such as the MLOps mentioned in section 3.5, monitoring the performance, accuracy of the model and its alignment with the business goals becomes necessary. However, the following points need to be considered:

1. Continuously monitor model performance, use model logging and access information through a dynamic dashboard.
2. Define parameters and thresholds (e.g. grid search, optimization, fairness awareness etc.) that adapt and help in measuring model performance and accuracy at any future iteration point in time.
3. Create an environment to monitor training and test datasets, changing data environments and changing business goals to be able to address model drifts, deficiencies or potential harm.
4. Use the environment and monitoring as a function to redress the model, should the model be negatively changed by its environment.

In addition, one needs to be continuously concerned about the changes in **model performance** and data robustness (see section 3.3.1 for more detail). If variability or unforeseen changes occur, those aspects can be monitored. Model performance measures how effectively an AI system achieves its objectives based on specific metrics. It reflects the model's ability to generalize from training data to unseen scenarios. Some key metrics and monitoring elements to use when evaluating model performance are:

1. Accuracy - proportion of correctly predicted instances.
2. Data and model drifts, which relate to the data and model alteration during time, such as changing the distribution of features, new patterns, on which models become unable to perform appropriately.
3. Precision and Recall - relevant for imbalanced datasets where false positives or false negatives matter. (false positives in data science are the results that indicate that a condition exists, while it actually does not exist; false negatives is the opposite, a test result indicating a condition does not exist, while it exists)
4. Automation and proactive notifications when dealing with performance issues
5. Monitorization of memory consumption, latency and GPU/CPU usage

Actuaries must ensure that all these aspects are considered, fed into audit and test trails, and are implemented in a stable technical environment. Moreover, once the processes of an AI system can be monitored, an automated governance framework should be established to ensure that regulatory requirements and conformity assessments are completed in a timely manner before a model is deployed.

Having a strong foundation for AI system monitoring also enables actuaries and organizations to decommission models more effectively. When **retiring an AI system**, the metrics and overviews gathered provide valuable insights. Clear performance assessments, safety reviews, fairness tests, and financial viability tests help communicate findings to stakeholders. They also facilitate post-mortem analysis, offering lessons for future processes, promoting knowledge transfer among teams, and mitigating any residual risks arising from retired systems.

4. Additional Considerations

4.1. Training and Education

To build an organization-wide culture of AI governance, training at different organizational levels is essential. Bridging the gap between disciplines and functions is important, as strengths and weaknesses are unavoidable. Teams responsible for the use of AI should have received appropriate training to reduce the likelihood of deployment errors.

Executive Leadership (C-Suite, Board)

- **Goal**: Ensure strategic alignment, risk management, and oversight.
- **Focus**: Ethical AI, risk and compliance, strategic goals, and accountability.
- **Format**: Executive briefings and workshops with case studies.

Senior Management (Dept Heads, AI Leaders)

- **Goal**: Embed governance practices within departments.
- **Focus**: Operationalizing AI governance, bias mitigation, resource needs, and risk management.
- **Format**: In-depth workshops and scenario simulations.

AI Practitioners and Data Scientists

- **Goal**: Equip with skills to design, test, and monitor governed models.
- **Focus**: Bias detection, data privacy, explainability, robustness testing, documentation.
- **Format**: Live practice sessions, coding workshops, and toolkits.

IT and Operations Teams

- **Goal**: Consistent governance in deployment and monitoring.
- **Focus**: Model lifecycle, data quality, monitoring, security.
- **Format**: Training in MLOps, data governance, and technical workshops.

Legal, Compliance, and Risk Management

- **Goal**: Oversee compliance and risk.
- **Focus**: Regulatory knowledge, risk assessment, data privacy, audits.

DRAFT

- **Format**: Legal briefings, scenario-based compliance training.

All Employees (Awareness Training)

- **Goal**: Foster awareness of AI ethics and governance.
- **Focus**: AI basics, ethical implications, data security, reporting mechanisms.
- **Format**: E-learning modules, webinars, and ethics workshops.

Ongoing Education

- Offer certification programs, refresher courses, and industry seminars to keep skills and knowledge up to date.

This structured training ensures AI governance is practiced effectively and aligns with organizational goals across all levels.

5. Conclusion

This governance framework breaks governance into few relevant parts – data, training, validation, implementation, and human involvement throughout. Each step requires interpretation to be applied to various forms of AI depending on its complexity and application. Because of this interpretation, AI Training mentioned in section 4 should not be undervalued or overlooked. Knowledge of AI will develop a culture of appropriate skepticism – and enthusiasm – that is needed to implement a governance framework around these revolutionary technologies.

AI is constantly evolving, and the governance frameworks for AI must continue to evolve with it. We expect this guidance to be updated with references to new interpretability and bias-related research as it becomes available. It is our hope that this guidance is both specific and flexible enough to easily incorporate these new methodologies.

Actuaries are particularly well suited for governing AI because the actuarial skillset has traditionally focused on both rigorous statistical models as well as the application of actuarial judgment to account for how these models may misbehave if applied directly in practice. The introduction of AI into actuarial practice may make this a more complicated endeavour, but not an unfamiliar one. We hope that this guidance can help actuaries translate their existing skillsets into the new world of AI governance.

DRAFT

Appendix A. Appendix

A.1. Glossary of Terms

[To be populated after the content is populated or in better shape]

A.2. Additional Resources and References

Case Studies

[To be populated]

DRAFT

Appendix B. Roles and Responsibilities

Responsible Party	Roles	Main Responsibilities
Data & Model Committee	This is a cross-functional executive body that has ultimate authority over Data Model Governance and is responsible for maintaining Data Model Governance across all functional areas and driving strategic initiatives to support its adoption across the enterprise. Could also include an ethics committee	• Provide high-level authority for the global policy and standards process • Review the development, deployment or procurement of AI systems, and to serve as an escalation point for the assessment of risks and dependencies. • Approve the strategic direction for data & model governance • Help the organization to communicate and promote the governance strategy • Authorize additional funding, if needed, for the operation of the data governance
Model Governance Committee	Executive body that has the ultimate authority for Model development, deployment and maintenance	• Responsible for establishing policies, standards and best practices. • High level oversight of model governance practices and their alignment with the organization's goals and regulatory requirements • Evaluate risks, set tolerance levels and review mitigation strategies • Approve key decisions such as deployment of high impact models or decommissioning of outdated ones • Periodically review governance processes to ensure compliance, security and effectiveness.
Model Owner	Responsible for adhering to the Data & Model Governance processes and regulations within their functional group. They own the data of their	• Manage models and approve their definitions and requirements • Review and approve specifications related to the model standard and revisions to ensure that any key changes are sufficiently understood

DRAFT

DRAFT

	functional groups and have the authority to approve the definitions of data and calculation requirements and models.	and their impact fully assessed and integrated • Ensure consistency and quality across different types of models and data (i.e. actuarial assumptions) • Ensure quality and consistency of models
Model Owner	Owns all aspects of the model, including development, performance monitoring, compliance, and maintenance. Ensures the model meets specific business objectives, aligns with strategic goals and business objectives, complies with regulations, remains effective in its purpose.and provides value. Responsible for the entire lifecycle of the model, from inception through decommissioning.	• Define the responsibilities for the model stakeholders including data scientists, IT, compliance, and business units. • Assign stewards responsible for maintaining, updating, and monitoring models throughout their lifecycle • Establish and/or approve workflows for key stages of model development, such as pre-deployment, post-deployment, and significant updates. • Set up accountability structures to ensure that the models are used responsibly and in alignment with organizational values • Assemble the right team (data scientists, compliance officers, IT security, and business leaders) to support governance and works closely with the team to ensure the model adheres to internal policies and external regulations • Regularly update stakeholders on model performance, risks, and any significant changes. • Decides when the model should be retrained, updated, or retired, based on performance and relevance.
Model Manager	Responsible for the data control system that ensures the efficiency of the consolidated Data & Model Governance processes. Produces monthly reports on the	• Manage the data and model quality assessment process: monitoring, management of any degradation, etc.

	evaluation of the data quality and model consistency or the identification of any problems	• Manage the Model control framework to ensure the efficiency of the Data & Model Governance processes • Contribute to improving or modifying the model control framework to ensure optimal operation • Ensure training and information for the main Data & Model Governance actors • Analyse control reports • Produce monthly reports on the assessment of the quality and consistency of the models and identification of problems • Manage the Data & Model indicators used in Governance activities and share them with all users
Model Manager	Oversees a portfolio of models and is responsible for managing the overall governance, performance, and risk across these models. Focuses on ensuring standardization, operational efficiency, and strategic alignment across the organization's models.	• Establishes governance standards and best practices for model development, deployment, and monitoring. • Assesses and mitigates risks associated with the model portfolio, such as bias, drift, or compliance issues. • Coordinates resources across models, optimizing for efficiency and effectiveness. • Tracks and reports on model performance metrics and governance compliance across the portfolio.
Data & Model Steward	Act on behalf of the model owners to confirm that the data is being used as intended and the processes and rules are implemented and respected. There may be multiple stewards for a single Model Owner if the amount of Model or the variety of topics is significant.	• Maintain data and models, propose new requirements and update the Data & Model Dictionary • Have an overall view of the entire data flow process as well as in-depth understanding of their specific areas • Perform manual quality checks • Monitor: Responsible for consolidating Consolidate the quality metrics (manual and automatic) in order to assess if a quality issue could occur

DRAFT

		• Coordinate the problem resolution process • Consolidate information to support and ensure Data & Model Quality • Resolve quality issues
Data Steward	Responsible for ensuring data quality, compliance, and appropriate governance, managing and overseeing the organization's data assets to ensure data quality, integrity, and accessibility,	• Monitor and enforce data quality standards. • Identify and rectify data quality issues, ensuring accuracy, completeness, and consistency across datasets. • Implement and support data governance policies and frameworks. • Collaborate with other data governance team members to outline best practices and data management processes. • Maintain documentation related to data definitions, data lineage, and metadata. • Ensure clarity and understanding of data standards, including source, structure, and intended use. • Oversee the data lifecycle, including data creation, storage, usage, and deletion. • Ensure compliance with data retention requirements and policies. • Define and manage access controls to sensitive data, ensuring that only authorized users can access or alter data. • Collaborate with IT and security teams to implement data protection measures. • Act as a liaison between technical teams, business users, and senior management regarding data-related issues and initiatives.

DRAFT

		• Provide training and guidance to staff on data management practices and tools. • Assess the impact of changes in data processes and systems on data quality and governance. • Help manage the implementation of new data systems or processes to minimize disruption and maintain data integrity. • Ensure that data handling practices comply with relevant laws and regulations (e.g., GDPR, HIPAA). • Stay updated on data protection regulations and help the organization adapt as necessary. • Establish metrics and key performance indicators (KPIs) to measure data quality and stewardship effectiveness. • Regularly report on data quality and governance status to stakeholders.
Data & Model Custodian	Responsible for information technology (e.g. ICT dpt.) which deals with the physical environment of the data and the management of the automation models or processes, as well as the tools necessary to facilitate the collection, processing, archiving and retrieval of data and calculation processes, as well as their access and use.	• Promote consistency of data management objectives, procedures, tools and techniques in a proactive manner • Perform impact analyses on changes to existing data sources, to the architecture of computerisation and calculation processes and to the operation of models • Implement adequate information security systems • Ensure that all system or process changes affecting the data have been notified to the entire Community• • Ensure automatic control of the Data Quality • Resolve problems with the data quality

DRAFT

Data Custodian	Responsible for enforcing the policies set by the Data Steward, handling the practical, day-to-day management and security of the data. Responsible for managing and securing the data lifecycle, ensuring that the models are built on reliable, high-quality, and compliant data, thereby supporting both the effectiveness of the AI and the organization's data integrity.	• Oversee the collection, storage, and maintenance of data. Protect sensitive data through encryption and access controls. • Ensure that the data are accessible, secure, and maintained in line with governance policies and regulatory requirements. • Ensure data integrity, accuracy, and quality. • Define data ownership and stewardship roles. • Ensure that data handling complies with regulatory requirements (e.g., GDPR, HIPAA) and organizational policies on data privacy and confidentiality. Work with legal and compliance teams to mitigate risks associated with data misuse. • Manage data from creation to disposal, ensuring appropriate retention and archival procedures. • Facilitate data access for authorized users while maintaining security and privacy. • Oversee who can access the data, ensuring that access aligns with governance and security policies. This includes setting permissions and maintaining access logs to track data usage. • Implement security protocols to protect sensitive data from unauthorized access, breaches, or leaks. Work with IT and security teams to enforce encryption, backup, and access restrictions. • Ensure that the data are accurate, clean, and well-maintained. • Maintain detailed documentation and metadata for datasets, which can include the data's origin, structure, and

DRAFT

		any transformations applied. Maintain thorough documentation of data sources, data flows, and data transformation processes. • Support the data steward.
Model User	Users reside in the business organization and are the owners of the processes that use the AI Model or the information produced by the models. They define the model requirements and model designs that are fit for purpose	• Own the processes that use the templates and define the requirements • Evaluate whether a potential issue (reported by the administrator) will impact the processes • During the design phase, they may also be involved in labelling datasets to be used in supervised learning. During the implementation phase, they should also provide feedback to the AI developers on any shortcomings of the AI systems or suggestions for improvement.
Model User	Responsible for ensuring that the models are leveraged effectively within the organization, providing a bridge between the model's technical functionality and its real-world impact.	• Interpret and act on the outputs provided by AI models to enhance decision-making, automate processes, or support specific objectives – must understand the model results and how they should influence actions. • Assess the model performance based on practical outcomes and provide feedback to model developers and data scientists, including reporting on the accuracy, usability, and relevance of the model's predictions in real-world scenarios. • Use the AI outputs ethically, particularly when dealing with sensitive data or decisions affecting individuals, being mindful of potential biases or limitations in the model and avoiding over-reliance on AI without human oversight. • Document how the model is used and ensure that the use aligns with the organization's policies and regulatory requirements.

DRAFT

		• Must understand changes in model behavior, adapt to new versions, and learn any updates in usage practices. • Bridge technical and non-technical teams by communicating AI-driven insights to stakeholders, helping to translate model outputs into actionable business terms.
IT security officer	Responsible for security (or team) should ensure that data model security is embedded in the information security management framework	• Appropriate technical measures must be taken to ensure the security of the data and the model • Assessed regularly throughout the application development and deployment lifecycle.
IT security officer	Responsible for safeguarding digital assets, network and data from cybersecurity threats.	• Create policies, standards and guidelines that align with industry best practice and any regulatory requirements • Prepare training material for use by others and/or conduct training for users • Conduct risk assessment and develop mitigation strategies • Continuously monitor threats and vulnerabilities • Develop and maintain a response plan for handling any incidents, lead or coordinate response effort and conduct post incident analysis to identify cause and improve mitigation strategies and response processes. • Implement continuous monitoring (logs, traffic and alerts) to detect unusual or suspicious activities, manage security tools and solutions, monitor and control access • Ensure compliance with regulations and industry standards • Assist with internal audits

DRAFT

		• Maintain documentation of security practices, policies, incidents, and audit results.
AI Officer	Ensure the consistency and coherence of the work performed by the business and technology teams and to act as a single point of contact for the organization.	Can be a new function/position or can be integrated into already existing functions (e.g. actuarial function, data protection officer, etc.) The AI officer would need to have sufficient expertise to oversee and advise on all functions and could coordinate the certification of activities. Primary focus can be on ensuring that each team is aware of the applicable policies and governance frameworks and that they are implemented for each AI use case. One potential disadvantage could be the establishment of parallel responsibilities.
	Consider adding Risk Management, Internal Audit, and Board roles	

DRAFT